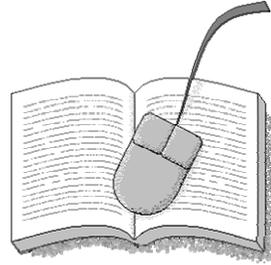


INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN
MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
Laboratorio de Lenguaje Natural y Procesamiento de Texto



**El formateo de entrada y la caracterización
morfológica de componentes de la base de datos
con colocaciones y relaciones semánticas
en español**

T E S I S

QUE PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA

ING. SABINO MIRANDA JIMÉNEZ

DIRECTOR:

DR. IGOR A. BOLSHAKOV

México, D. F., Junio de 2003

Dedicatoria

A mi padre
que fue un hombre ejemplar
y siempre estará en mi corazón.

Agradecimientos

A mi madre y hermanos
por el apoyo moral que me han dado
en los momentos más difíciles de mi vida.

Al Dr. Igor A. Bolshakov
por sus sabios consejos, paciencia y guía
en lo académico y personal.

A los Drs. Grigori Sidorov y
Alexander Gelbukh
por el apoyo que me dieron
durante mi estancia en el centro.

A todos mis compañeros
que de alguna manera han ayudado
a mi formación.

Al Centro de Investigación en Computación (CIC-IPN) y
al Programa Institucional de Formación de Investigadores (PIFI)
por el apoyo económico que me proporcionaron.

ÍNDICE GENERAL

RELACIÓN DE FIGURAS Y TABLAS	VIII
GLOSARIO DE TÉRMINOS	XI
CAPÍTULO 1. INTRODUCCIÓN	1
CAPÍTULO 2. ANTECEDENTES	44
CAPÍTULO 3. DESARROLLO DEL SISTEMA <i>CROSSLEXICA-ESP</i>	58
CAPÍTULO 4. CARACTERIZACIÓN MORFOLÓGICA DE LAS COLOCACIONES.....	96
CAPÍTULO 5. RESULTADOS DEL SISTEMA <i>CROSSLEXICA-ESP</i>	114
CAPÍTULO 6. CONCLUSIONES	149
REFERENCIAS	152
ANEXOS	155

ÍNDICE DETALLADO

RELACIÓN DE FIGURAS Y TABLAS	VIII
GLOSARIO DE TÉRMINOS	XI
CAPÍTULO 1. INTRODUCCIÓN	1
1.1 JUSTIFICACIÓN.....	2
1.2 OBJETIVOS.....	3
1.3 ORGANIZACIÓN DE LA TESIS	4
1.4 NOCIONES FUNDAMENTALES DEL LENGUAJE	5
1.4.1 <i>Niveles de descripción del lenguaje</i>	5
1.4.1.1 Fonética y Fonología	5
1.4.1.2 Morfología.....	6
1.4.1.3 Sintaxis	6
1.4.1.4 Semántica	9
1.4.1.5 Pragmática	9
1.4.2 <i>El español gramaticalmente</i>	10
1.4.2.1 Morfología.....	10
1.4.2.1.1 Flexión.....	11
1.4.2.1.2 Derivación	12
1.4.2.1.3 Composición.....	13
1.4.2.2 Sintaxis	14
1.4.2.2.1 El sustantivo y su papel.....	14
1.4.2.2.1.1 Categorías gramaticales del sustantivo.....	16
1.4.2.2.1.2 Sustantivos aumentativos, diminutivos y despectivos.....	18
1.4.2.2.1.3 Funciones sintácticas del sustantivo	19
1.4.2.2.2 El adjetivo y su papel	20
1.4.2.2.2.1 Posición del adjetivo respecto al sustantivo	21
1.4.2.2.2.2 Adjetivos determinativos.....	21
1.4.2.2.2.3 Adjetivos apocopados.....	22
1.4.2.2.2.4 Categorías gramaticales del adjetivo	22
1.4.2.2.2.5 Adjetivos aumentativos, diminutivos y despectivos	24
1.4.2.2.3 El verbo y su papel	24
1.4.2.2.3.1 Categorías gramaticales.....	24
1.4.2.2.3.1.1 Persona y número	25

1.4.2.2.3.1.2	Modo.....	25
1.4.2.2.3.1.3	Tiempo.....	26
1.4.2.2.3.2	Modelos de conjugación.....	26
1.4.2.2.3.3	Clases de verbos.....	27
1.4.2.2.3.4	Formas no personales del verbo.....	30
1.4.2.2.3.5	Funciones sintácticas del verbo.....	31
1.4.2.2.3.6	Uso del pronombre <i>se</i>	34
1.4.2.2.4	El adverbio y su papel.....	34
1.4.2.2.4.1	Adverbialización.....	35
1.4.2.2.4.2	Sufijación.....	35
1.4.2.2.4.3	Gradación.....	35
1.4.2.2.4.4	Clasificación de los adverbios.....	35
1.4.2.2.4.5	Locuciones adverbiales.....	36
1.4.2.2.5	El pronombre y su papel.....	36
1.4.2.2.6	El artículo y su papel.....	37
1.4.2.2.7	La preposición y su papel.....	39
1.4.2.3	La concordancia.....	40
CAPÍTULO 2.	ANTECEDENTES	44
2.1	COLOCACIONES EN GENERAL.....	45
2.2	CARACTERÍSTICAS DE LAS COLOCACIONES.....	46
2.3	CLASIFICACIÓN DE LAS COLOCACIONES.....	47
2.4	TRATAMIENTO DE LAS COLOCACIONES EN LOS DICCIONARIOS.....	47
2.4.1	<i>Las colocaciones en el Diccionario Explicativo Combinatorio (DEC)</i>	50
2.4.2	<i>Las funciones léxicas en el DEC</i>	53
2.5	APLICACIONES DE LAS COLOCACIONES.....	55
CAPÍTULO 3.	DESARROLLO DEL SISTEMA <i>CROSSLEXICA-ESP</i>	58
3.1	ESTRUCTURA DEL SISTEMA <i>CROSSLEXICA-ESP</i>	58
3.2	BASES PRIMARIAS DE COLOCACIONES EN <i>CROSSLEXICA-ESP</i>	61
3.2.1	<i>Archivo de modificadores</i>	63
3.2.2	<i>Archivo de valencias verbales</i>	64
3.2.3	<i>Archivo de valencias sustantivas</i>	65
3.2.4	<i>Archivo de hipónimos / hiperónimos</i>	66
3.3	ESTRUCTURA DE LAS BASES PRIMARIAS EN GENERAL.....	67
3.3.1	<i>Palabra clave</i>	67

3.3.1.1	Información semántica	67
3.3.1.2	Parámetros morfológicos	68
3.3.2	<i>Palabras relacionadas</i>	70
3.3.2.1	Idiomatización	70
3.3.2.2	Estilo de la combinación	71
3.4	ESTRUCTURA DE LAS BASES DE COLOCACIONES MODIFICADORAS	72
3.4.1	<i>Base primaria</i>	72
3.4.2	<i>Base secundaria</i>	75
3.4.2.1	Descripción de las columnas	75
3.5	ESTRUCTURA DE LAS BASES DE COLOCACIONES CON VALENCIAS VERBALES	79
3.5.1	<i>Base primaria</i>	79
3.5.1.1	Valencia omitida	81
3.5.1.2	Posición de la palabra clave	81
3.5.2	<i>Base secundaria</i>	82
3.5.2.1	Descripción de las columnas	83
3.6	ESTRUCTURA DE LAS BASES DE COLOCACIONES CON VALENCIAS SUSTANTIVALES	86
3.6.1	<i>Base primaria</i>	86
3.6.2	<i>Base secundaria</i>	88
3.6.2.1	Descripción de las columnas	89
3.7	ESTRUCTURA DE LAS BASES DE HIPÓNIMOS / HIPERÓNIMOS	90
3.7.1	<i>Base primaria</i>	90
3.7.1.1	Clasificación de hipónimos	92
3.7.2	<i>Base secundaria</i>	93
3.7.2.1	Descripción de las columnas	94
CAPÍTULO 4. CARACTERIZACIÓN MORFOLÓGICA DE LAS COLOCACIONES		96
4.1	DETERMINACIÓN DE LAS CATEGORÍAS GRAMATICALES	96
4.1.1	<i>Clase de palabra</i>	96
4.1.1.1	Sufijos característicos para sustantivos	96
4.1.1.2	Sufijos característicos para adjetivos	97
4.1.1.3	Sufijos característicos para verbos	97
4.1.1.4	Sufijo característico para adverbios	97
4.1.2	<i>Categoría de género</i>	98
4.1.3	<i>Categoría de número</i>	98
4.2	DETERMINACIÓN DE LA CLASE DE INCLINACIÓN	100

4.3	PROCEDIMIENTO PARA COLOCACIONES MODIFICADORAS	101
4.3.1	<i>Hipótesis para colocaciones modificadoras</i>	101
4.3.2	<i>Comprobación de la concordancia</i>	102
4.4	PROCEDIMIENTO PARA COLOCACIONES CON VALENCIAS VERBALES	102
4.4.1	<i>Hipótesis para colocaciones con valencias verbales</i>	103
4.4.2	<i>Comprobación de la concordancia</i>	103
4.5	PROCEDIMIENTO PARA COLOCACIONES CON VALENCIAS SUSTANTIVALES	104
4.5.1	<i>Hipótesis para colocaciones con valencias sustantivales</i>	104
4.6	PROCEDIMIENTO PARA HIPÓNIMOS / HIPERÓNIMOS	104
4.7	IMPLEMENTACIÓN DEL SISTEMA CROSSLEXICA-ESP	105
4.8	DIAGRAMAS DE CASOS DE USO	105
CAPÍTULO 5. RESULTADOS DEL SISTEMA CROSSLEXICA-ESP		114
5.1	RESULTADOS DE LA CARACTERIZACIÓN MORFOLÓGICA	114
5.1.1	<i>Descripción de los resultados de la base secundaria de colocaciones modificadoras</i>	114
5.1.1.1	Información semántica	117
5.1.1.2	Información estilística	118
5.1.2	<i>Descripción de los resultados de la base secundaria de colocaciones con valencias verbales</i>	118
5.1.3	<i>Descripción de los resultados de la base secundaria de colocaciones con valencias sustantivales</i>	122
5.1.4	<i>Descripción de los resultados de la base secundaria de hipónimos / hiperónimos</i>	125
5.1.5	<i>Protocolo de errores para la caracterización morfológica</i>	127
5.2	SISTEMA CROSSLEXICA-ESP VERSIÓN BETA 1.0	128
5.2.1	<i>Funcionamiento del sistema CrossLexica-Esp</i>	129
5.2.2	<i>Ejemplos de colocaciones modificadoras</i>	130
5.2.2.1	Combinaciones sustantivo → adjetivo	131
5.2.2.2	Combinaciones adjetivo → adverbio	133
5.2.2.3	Combinaciones verbo → adverbio	135
5.2.3	<i>Ejemplos de colocaciones con valencias verbales</i>	136
5.2.3.1	Combinaciones verbo → sustantivo	136
5.2.4	<i>Ejemplos de colocaciones con valencias sustantivales</i>	138
5.2.4.1	Combinaciones sustantivo → sustantivo	138
5.2.5	<i>Ejemplos de combinaciones de palabras con información semántica</i>	139

5.2.6	<i>Ejemplos de combinación de palabras con idiomatización y estilo</i>	142
5.2.7	<i>Ejemplos de relaciones semánticas</i>	144
5.2.8	<i>Inferencia de colocaciones</i>	145
CAPÍTULO 6. CONCLUSIONES		149
6.1	CONTRIBUCIONES	150
6.2	PUBLICACIONES	150
6.3	TRABAJO FUTURO	150
REFERENCIAS		152
ANEXOS		155
ANEXO A. BASES PRIMARIAS DE COLOCACIONES		155
	<i>Anexo A.1 Base primaria de colocaciones modificadoras</i>	155
	<i>Anexo A.2 Base primaria de colocaciones con valencias verbales</i>	162
	<i>Anexo A.3 Base primaria de colocaciones con valencias sustantivales</i>	167
	<i>Anexo A.4 Base primaria de hipónimos / hiperónimos</i>	172
ANEXO B. BASES SECUNDARIAS DE COLOCACIONES		179
	<i>Anexo B.1 Base secundaria de colocaciones modificadoras</i>	179
	<i>Anexo B.2 Base secundaria de colocaciones con valencias verbales</i>	185
	<i>Anexo B.3 Base secundaria de colocaciones con valencias sustantivales</i>	190
	<i>Anexo B.4 Base secundaria de hipónimos / hiperónimos</i>	195
ANEXO I. PREPOSICIONES		201
ANEXO II. ADVERBIOS		204
ANEXO III. ARCHIVOS AUXILIARES		207

Relación de figuras y tablas

Figura 1.1 Ejemplo de árbol de constituyentes.....	7
Figura 1.2 Ejemplo de árbol de dependencia	8
Figura 2.1 Ejemplo de colocaciones del BBI	48
Figura 2.2 Entrada de colocación del OCD.....	48
Figura 2.3 Ejemplo de colocación del sistema ARCS.....	49
Figura 2.4 Ejemplo de colocación del sistema CrossLexica versión rusa.....	50
Figura 2.5 Ejemplo de entrada del DEC (francés).....	53
Figura 3.1 Esquema de bloques del sistema CrossLexica-Esp.....	61
Figura 3.2 Ejemplo de artículo de una base primaria.....	62
Figura 3.3 Enlaces sintácticos para modificadores.....	63
Figura 3.4 Enlaces sintácticos para valencias verbales	64
Figura 3.5 Enlaces sintácticos para valencias sustantivales	66
Figura 3.6 Esquema de un artículo de colocaciones modificadoras.....	73
Figura 3.7 Esquema de un artículo de colocaciones con valencias verbales.....	79
Figura 3.8 Esquema de un artículo de colocaciones con valencias sustantivales.....	86
Figura 3.9 Esquema de un artículo de hipónimos / hiperónimos	91
Figura 4.1 Diagrama de casos de uso del sistema CrossLexica-Esp.....	106
Figura 4.2 Diagrama de secuencia del caso de uso <i>Formatear colocaciones</i>	108
Figura 4.3 Diagrama de secuencia del caso de uso <i>Compilar enlaces entre palabras</i>	108
Figura 4.4 Diagrama de secuencia del caso de uso <i>Consultar colocaciones</i>	109
Figura 4.5 Modelo entidad-relación del sistema CrossLexica-Esp.....	110
Figura 5.1 Ejemplo de entrada de colocaciones en el sistema CrossLexica-Esp.....	130
Figura 5.2 Concordancia entre un sustantivo y su adjetivo.....	131
Figura 5.3 Concordancia entre un adjetivo y varios sustantivos	132
Figura 5.4 Combinación de un adjetivo y sus modificadores adverbiales	133

Figura 5.5 Combinación de un adverbio y los verbos que modifica	134
Figura 5.6 Combinación de un verbo y sus adverbios modificadores.....	135
Figura 5.7 Combinación de un verbo con sus complementos y sujetos.....	136
Figura 5.8 Sustantivo como complemento verbal	137
Figura 5.9 Sustantivo como complemento sustantival	138
Figura 5.10 Información adicional para palabras homónimas (primer caso).....	139
Figura 5.11 Información adicional para palabras homónimas (segundo caso)	140
Figura 5.12 Información adicional para palabras homónimas (tercer caso).....	141
Figura 5.13 Combinaciones con información idiomática.....	142
Figura 5.14 Combinaciones con información estilística	143
Figura 5.15 Combinaciones con relaciones semánticas	144
Figura 5.16 Inferencia de colocaciones para <i>azucenas</i>	145
Figura 5.17 Inferencia de colocaciones para <i>tulipanes</i>	146
Figura 5.18 Relaciones sintácticas y semánticas para <i>bebidas</i>	147
Figura 5.19 Inferencia de colocaciones para <i>Coca-Cola</i>	148
Tabla 1.1 Modelo para verbos de primera conjugación: <i>amar</i>	26
Tabla 1.2 Modelo para verbos de segunda conjugación: <i>comer</i>	27
Tabla 1.3 Modelo para verbos de tercera conjugación: <i>vivir</i>	27
Tabla 3.1 Marcos para palabras clave	70
Tabla 3.2 Marcos para palabras relacionadas.....	71
Tabla 3.3 Formato de la base secundaria de modificadores.....	75
Tabla 3.4 Papel de la palabra corriente.....	75
Tabla 3.5 Valores de la parte de la oración	76
Tabla 3.6 Valores para género.....	76
Tabla 3.7 Valores para número	76
Tabla 3.8 Desinencias para la clase del sustantivo.....	77

Tabla 3.9 Desinencias para la clase del adjetivo	77
Tabla 3.10 Desinencias para la clase del verbo	78
Tabla 3.11 Posición del adjetivo.....	78
Tabla 3.12 Formato de la base secundaria de valencias verbales.....	82
Tabla 3.13 Desinencias para verbos en infinitivo y pronombre <i>se</i>	83
Tabla 3.14 Tiempos de conjugación del verbo.....	84
Tabla 3.15 Persona de conjugación	85
Tabla 3.16 Artículos	85
Tabla 3.17 Formato de la base secundaria de valencias sustantivales.....	88
Tabla 3.18 Formato de la base secundaria de hipónimos / hiperónimos.....	94
Tabla 3.19 Tipo de concepto	94
Tabla 3.20 Valor de la parte de la oración.....	94
Tabla 4.1 Clases de inclinación para sustantivos	100
Tabla 4.2 Clases de inclinación para adjetivos.....	100
Tabla 4.3 Clases de inclinación para verbos.....	100
Tabla 4.4 Diccionario de datos de la tabla <i>Diccionario</i>	111
Tabla 4.5 Diccionario de datos de la tabla modificadores: <i>ColMod</i>	111
Tabla 4.6 Diccionario de datos de la tabla de complementos sustantivales: <i>ColSus</i>	112
Tabla 4.7 Diccionario de datos de la tabla de complementos verbales: <i>ColVer</i>	112
Tabla 4.8 Diccionario de datos de l atabla de clases y subclases: <i>Conceptos</i>	113

Resumen

Existen pocos diccionarios o sistemas especializados que demuestran las combinaciones de palabras más frecuentes, son conocidos como diccionarios combinatorios y sólo existen para algunos idiomas: inglés, ruso, francés y algunos otros, pero para español no se conoce ninguno.

En estos diccionarios, para una palabra consultada, se muestran los rasgos morfológicos de la palabra y las palabras con que se combina frecuentemente. La presentación de las combinaciones de palabras de algunos diccionarios no es muy buena ni la explicación de las relaciones sintácticas entre las palabras. Estos diccionarios son útiles para el aprendizaje del idioma y para resolver varias tareas de lingüística computacional.

Para la presentación correcta de las combinaciones de palabras (colocaciones), es necesario obtener las características morfológicas de los componentes de la colocación, así como los papeles sintácticos y semánticos que los vinculan.

En esta tesis se desarrollan las herramientas necesarias para la obtención de las características morfológicas y semánticas de cada componente de la colocación; se demuestra la organización especial de las colocaciones en archivos que representan los papeles sintácticos y semánticos entre los componentes de la colocación.

Además, se desarrolla el sistema que demuestra las combinaciones de palabras y los diferentes papeles sintácticos y semánticos entre ellas. Una característica adicional del sistema es la capacidad de inferir colocaciones que no han sido explícitamente escritas en los archivos de entrada, pero que pueden ser generadas de alguna forma lógica.

Abstract

There are few dictionaries or specialized systems that demonstrate the combinations of more frequent words, exist they are known as combinatorial dictionaries and they only exist for some languages: English, Russian, French and some other ones, but for Spanish none is known.

In these dictionaries, for a consulted word, the morphological features of the word and the words combines frequently are shown. The presentation of the combinations of words of some dictionaries is not good neither the explanation of the syntactic relationships among the words. These dictionaries are useful even the learning of the language and to solve several tasks of computational linguistic.

For the correct presentation of the combinations of words (collocations), it is necessary to obtain the morphological characteristics of the components of collocation, as well as the syntactic and semantic roles that link them.

In this thesis the necessary tools are developed for the obtaining of the morphological and semantic characteristics of each component of collocation; the special organization of the collocations is demonstrated in files that represent the syntactic and semantic roles among the components of collocation.

Also, the system is developed that demonstrates the combinations of words and the different syntactic and semantic roles among them. An additional characteristic of the system is the capacity to infer collocations that have not been written explicitly in the entrance files, but that they can be generated in some logical way.

Capítulo 1. Introducción

En este capítulo describimos los alcances de esta tesis y la justificación de la misma. En la última sección del capítulo se presentan algunos aspectos fundamentales del lenguaje y algunas peculiaridades del idioma español.

La comunicación entre los seres humanos es fundamental para el desarrollo de la vida en sociedad y se realiza mediante el empleo del lenguaje o de distintos sistemas de comunicación. El lenguaje natural es uno de los aspectos fundamentales que ha influido en el comportamiento humano. Gracias al lenguaje el hombre ha podido formar sociedades cada vez más complejas [13].

Existen muchos tipos de lenguajes en general como la pintura, la música, la mímica, los lenguajes de programación que utilizamos para comunicarnos con las computadoras, etc., pero indudablemente, el más importante es el lenguaje nombrado lenguaje natural.

La lengua o idioma, se diferencia de otros sistemas de comunicación porque es mucho más eficaz y preciso, además de ser exclusivo de los seres humanos. Es un producto de una convención social y constituye una herencia cultural. Las lenguas son sistemas que le permiten a los individuos entender y producir un número ilimitado de oraciones y mensajes, a partir de un número bastante limitado de signos y reglas.

La lengua la aprendemos desde niños por imitación de nuestros padres y de las personas que nos rodean. Para aprender una segunda lengua no es suficiente conocer las palabras de forma aislada de dicha lengua, sino también conocer las relaciones sintácticas entre las combinaciones de palabras que se vinculan (colocaciones).

Las relaciones sintácticas entre las combinaciones de palabras son, por ejemplo, *país grande* (sustantivo – adjetivo), *muy grande* (adverbio – adjetivo), *trabajar fuertemente* (verbo – adverbio), *voceros del gobierno* (sustantivo – sustantivo), etc. Estas combinaciones de palabras pueden unirse directamente o por medio de otras palabras funcionales (preposiciones) como en *voceros del gobierno*.

Las relaciones sintácticas mencionadas se describen en manuales usuales, pero para cada tipo de relación los pares concretos aparecen raramente en los diccionarios. Además, estas relaciones no son claras para un niño nativo en el proceso de aprendizaje del lenguaje ni

para un extranjero sin entrenamiento especial. Por eso sólo se aprenden con el uso del lenguaje.

Entonces, el conocimiento de cómo combinar las palabras es de gran importancia para el aprendizaje del idioma y para realizar varias tareas de lingüística computacional: analizadores sintácticos, desambiguación de sentidos de palabras, detección y corrección de errores semánticos, traducción automática, etc.

Existen pocos diccionarios que contemplan este tipo de relaciones. En la literatura son conocidos como *Diccionarios Combinatorios*. En general, contienen combinaciones de palabras que se encuentran con alta frecuencia en los textos, aunque sólo existen para algunos idiomas.

Para inglés, el *BBI Dictionary* [1] en su forma impresa cuenta con algunos centenares de combinaciones, el *Oxford Collocations Dictionary* [27] que promete contar con más de un millón de combinaciones y el sistema *ARCS* [18], en su forma para Internet, ofrece varios miles de combinaciones. Para ruso, el sistema *CrossLexica* [2], en su forma electrónica, cuenta con varios cientos de miles de combinaciones. Pero para español no existe ninguno de este tipo.

Otro tipo de diccionarios son los *Explicativos Combinatorios* [4], los que representan de una forma funcional y con un alto grado de complejidad, las relaciones dentro del diccionario, pero en cantidad muy restringida (algunos centenares de palabras). Estos diccionarios sólo existen para algunos lenguajes (ruso, francés y algunos otros), pero no se conoce ninguno de este tipo para español.

1.1 Justificación

Para el aprendizaje informatizado de una segunda lengua no es suficiente conocer el vocabulario de palabras aisladas, sino también las relaciones sintácticas entre las combinaciones de palabras. Las formas más frecuentes en que se combinan las palabras se demuestran en los diccionarios combinatorios, existen para algunos idiomas, pero para español no se conoce ninguno.

Por lo anterior y al ser el español la tercera lengua más hablada después del chino e inglés, con aproximadamente 400 millones de hablantes [20], es de gran interés desarrollar un recurso lingüístico como el diccionario de colocaciones para resolver varias tareas de lingüística computacional.

Para llevar acabo esto, son necesarias las herramientas que preparen la información de las colocaciones, es decir, los programas que obtengan las características morfológicas y semánticas de cada componente de las colocaciones, esto es, para poder presentarlas adecuadamente al usuario.

Este trabajo es importante porque forma parte de un proyecto mayor, apoyado por CGEPI-IPN: *Un diccionario de colocaciones para español con enriquecimiento con la inferencia lógica*; con registro CGEPI-IPN 20010767 dirigido por el Dr. Igor A. Bolshakov.

Contar con este tipo recursos lingüísticos, donde se demuestren las relaciones sintácticas y los pares concretos de combinaciones, son útiles para varios proyectos del Laboratorio de Lenguaje Natural como [22 y 25]; también son útiles para la enseñanza de idiomas [24].

1.2 Objetivos

Objetivos generales

1. Desarrollar las herramientas utilitarias que transformen los archivos con colocaciones de una forma primaria a una forma secundaria con todos los rasgos morfológicos y semánticos para cada componente de la colocación.
2. Desarrollar la aplicación que demuestre los diferentes papeles sintácticos y semánticos entre los componentes de la colocación.

Objetivos específicos

1. Recopilar las bases primarias de colocaciones para los diferentes enlaces sintácticos (modificadores, valencias verbales y valencias sustantivales) y enlaces semánticos (hipónimos e hiperónimos).
2. Desarrollar las utilerías que transformen las bases primarias de colocaciones a una forma secundaria con todas las características morfológicas y semánticas de cada componente de la colocación.
3. Desarrollar la utilería que compile las bases secundarias de colocaciones a dos bases de datos. Una base, el diccionario general, que contiene las palabras únicas y la otra, la que contiene los enlaces entre palabras (las relaciones sintácticas y semánticas).
4. Desarrollar la interfaz que demuestre las colocaciones y los papeles sintácticos y semánticos entre sus componentes.

1.3 Organización de la tesis

Esta tesis esta dividida en seis capítulos. El primer capítulo consiste de la parte introductoria y tiene la finalidad de establecer los objetivos y la justificación de este trabajo, así como describir algunas nociones fundamentales del idioma español.

En el capítulo dos se describen los antecedentes acerca de las colocaciones en general, algunas definiciones de ellas y el tratamiento que se les ha dado en los diccionarios para algunas lenguas como inglés, ruso y francés.

En el capítulo tres se presenta la estructura del sistema CrossLexica para español y las especificaciones de los archivos para algunos tipos de colocaciones con relaciones sintácticas y semánticas.

En el capítulo cuatro se describe el método utilizado y el procedimiento para la obtención de los rasgos morfológicos de los componentes de las colocaciones. Se describen también las herramientas que se utilizaron en el desarrollo del sistema.

En el capítulo cinco se demuestran los resultados obtenidos del formateo de las colocaciones para cada tipo que maneja el sistema y se presenta la interfaz del sistema CrossLexica para español.

En el capítulo seis se muestran las conclusiones y las aportaciones de esta tesis. Adicionalmente, se mencionan algunas tareas futuras a ser consideradas.

En la última parte de la tesis se presentan los anexos, aquí se muestran las bases de colocaciones primarias y las bases formateadas con todos los rasgos morfológicos; se proporcionan también los archivos auxiliares (preposiciones, adverbios, excepciones de género, excepciones de número, etc.) que se requieren para el formateo adecuado de las colocaciones.

1.4 Nociones fundamentales del lenguaje

En los últimos años el procesamiento del lenguaje natural ha tenido gran interés. Hemos visto películas de ciencia ficción donde las computadoras pueden hablar, entendernos, entablar un diálogo e incluso expresar sentimientos [29]. Esto parece maravilloso, pero la realidad es otra: no se ha logrado todo esto con gran éxito. Existen muchos problemas en el lenguaje tales como la ambigüedad en las frases (decidir qué sentido de la frase es el más adecuado para tal o cual situación); generar combinaciones de palabras que sintácticamente son correctas, pero semánticamente incorrectas en un sentido pragmático; traducción automática, etc. Todos estos problemas son analizados por la ciencia denominada lingüística computacional, cuya principal tarea es desarrollar programas de computadora para el procesamiento de texto en lenguaje natural.

La lingüística computacional tiene sus bases en la lingüística general. La lingüística es la ciencia que estudia las lenguas en todos sus aspectos. Para su estudio, la lingüística se divide en diferentes niveles.

1.4.1 Niveles de descripción del lenguaje

Existen varios enfoques para estudiar y describir las lenguas y los cambios habidos en ellas. Típicamente, el lenguaje se divide en los siguientes niveles [3]:

- Fonología / Fonética
- Morfología
- Sintaxis
- Semántica
- Pragmática

Los niveles se diferencian únicamente por el enfoque que se le da al análisis del lenguaje.

1.4.1.1 Fonética y Fonología

Los hablantes de cualquier lengua articulan sonidos lingüísticos para formar las palabras. Los sonidos lingüísticos se producen a través de los órganos de fonación (labios, dientes, alvéolos, lengua, cuerdas vocales, etc.).

Un sonido puede presentar variaciones en el momento de articularse, debido a la influencia del sonido que le sigue o precede. Por ejemplo, la [n] (se representa un sonido entre corchetes, [X]) de la palabra *nube*, se pronuncia, generalmente, como [m] en las siguientes pala-

bras: *invierno* y *envase*. Estas distintas producciones de un mismo sonido se llaman alomorfos.

La fonética estudia la producción de los sonidos lingüísticos, así como las diferentes realizaciones de los alomorfos.

La fonología se ocupa del estudio de los sonidos en unidades que provocan cambio de significado. Estas unidades se denominan fonemas. La fonología determina cuáles son los fonemas de una lengua y los organiza dentro de un sistema a partir de sus diferencias fundamentales. Por ejemplo, los fonemas /n/, /s/, /t/ (un fonema se representa entre barras, /X/) son distintos, porque la presencia de uno u otro, en un mismo contexto, ocasiona cambio de significado: *pana*, *pasa*, *pata*.

1.4.1.2 Morfología

La morfología estudia la estructura interna de las palabras. Dentro de las tareas principales: delimita, define y clasifica las unidades del componente morfológico, las agrupa en sus respectivos paradigmas y explica la forma en que las unidades integrantes de la palabra se combinan para constituir su estructura interna. La morfología la estudiaremos a detalle en el punto *1.4.2.1 Morfología*.

1.4.1.3 Sintaxis

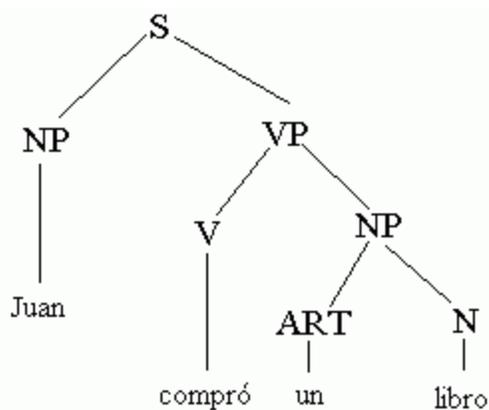
La sintaxis estudia la estructura de las frases y la forma en que se relacionan las palabras. Para la representación de las relaciones de una oración son conocidos dos enfoques: el enfoque de constituyentes [3] y el de dependencias [5].

Enfoque de constituyentes

En este enfoque la estructura sintáctica indica la forma en que las palabras están relacionadas unas con otras dentro de la oración, en el mismo la representación sintáctica del lenguaje se basa en la noción de una gramática libre de contexto (*context-free grammar*), la cual representa la oración en términos de componentes, frases que son subpartes de otras frases [30]. Estas representaciones son nombradas árboles de constituyentes. Por ejemplo, para la oración *Juan compró un libro*, usando la gramática 1, se genera el árbol de constituyentes que se muestra en la figura 1.1. Los símbolos NP (frase nominal), VP (frase verbal), ART (artículo), V(verbo), S(inicio de la frase), N(sustantivo) son las etiquetas de los nodos que aparentemente determinan las relaciones sintácticas entre los nodos.

Gramática 1

1. $S \rightarrow NP VP$	5. $NAME \rightarrow Juan$
2. $VP \rightarrow V NP$	6. $V \rightarrow compró$
3. $NP \rightarrow NAME$	7. $ART \rightarrow un$
4. $NP \rightarrow ART N$	8. $N \rightarrow libro$

**Figura 1.1 Ejemplo de árbol de constituyentes**

En este enfoque se requiere especificar las reglas que definen la gramática.

Enfoque de dependencias

El enfoque de dependencias encuentra una descripción más rigurosa en la Teoría Significado \Leftrightarrow Texto (Meaning \Leftrightarrow Text Theory, MTT) propuesta por I. Mel'čuk [5]. En este enfoque el conjunto de enlaces sintácticos entre las palabras representan la oración. Estos conjuntos de enlaces son llamados árboles de dependencia (*dependency tree*).

En los árboles de dependencia el orden de las palabras no es restricción alguna para la representación. Por el contrario, en los árboles de constituyentes es importante el orden en que aparecen las palabras para su análisis. Por ejemplo, el orden de las palabras en español tienen mayor grado de libertad que el inglés, es decir, para la oración, *el niño busca a su mamá* (sujeto (S), verbo (V), complemento (C)), puede describirse con diferente orden y aceptada sintácticamente, *Busca a su mamá el niño* (V, C, S), *A su mamá busca el niño* (C, V, S). Todas son correctas sintácticamente.

En los árboles de dependencia, las relaciones sintácticas mantenidas son de tipo binario. Estas relaciones son llamadas dependencias sintácticas y son representadas por flechas o

por arcos con flechas. Por ejemplo, $X \rightarrow Y$ (la flecha indica la dependencia sintáctica) Y depende de X o recíprocamente X gobierna a Y. En términos de Mel'čuk, X es el gobernador sintáctico y Y es el dependiente sintáctico [5].

Los árboles de dependencia ponen mayor énfasis en la especificación detallada de cualquier tipo de relación sintáctica entre dos elementos relacionados.

Por ejemplo, tenemos la siguiente oración y su árbol de dependencia.

El hombre es muy astuto

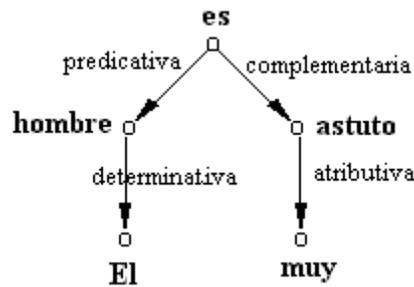


Figura 1.2 Ejemplo de árbol de dependencia

Las relaciones sintácticas para cada par de nodos se especifican como se muestra la figura 1.2.

El enfoque de dependencias se basa en la idea de que la sintaxis es casi completamente un componente de las capacidades de combinación y que las palabras cumplen ciertos requerimientos. El modelo para definir estos fenómenos es similar al de la química para formar las moléculas [31]. Las palabras tienen valencias, como los átomos de la química, que se usan para indicar con qué clase de palabras (sustantivo, verbo, etc) se pueden combinar y con cuántas a la vez. De esta manera se van formando estructuras más complejas.

Aunque las valencias se definen para todas las clases de palabras, son especialmente descritas para los verbos. Éstas tienen especial aplicación en la MTT y en especial en el Diccionario Explicativo Combinatorio (DEC) en lo que se llama *patrones de rección* [7] o *esquema de régimen* [6] (*government pattern*).

Los patrones de rección describen las valencias para cualquier clase de palabras, así como la correspondencia entre las valencias semántica y sintáctica. Los patrones de rección describen el conocimiento lingüístico que no se basa en la lógica y debe describir de forma

manual la colección completa de todos los posibles objetos de las palabras específicas [7] (verbos, sustantivos o adjetivos).

Los patrones de rección para verbos, se representan por medio de una tabla en la cual se especifican todas las posibles valencias. La tabla es precedida con una fórmula que expone la semántica de la situación reflejada por el verbo con todas sus valencias. Véase Capítulo 2, *Figura 2.5 Ejemplo de entrada del DEC (francés)*.

En este nivel del lenguaje se pretende describir todo el posible conocimiento lingüístico de la estructura de la oración, para facilitar el análisis del siguiente nivel.

1.4.1.4 Semántica

Estudia el significado de los signos lingüísticos, esto es, las palabras, las expresiones y las oraciones. La finalidad de la semántica es establecer el significado de los signos dentro del proceso que asigna tales significados. Para esto se debe estudiar qué signos existen y cuáles de ellos tienen significación dentro de una realidad.

Mediante la construcción de sistemas de reglas semánticas que sirvan para cualquier lenguaje, la semántica intenta responder a preguntas como: ¿A qué realidad se refiere cuando se habla de significado?, ¿Qué condiciones (además de las meramente sintácticas) debe reunir una palabra o una frase para ser significativa?, ¿Se puede atribuir algún significado a expresiones lingüísticas tales como *los números primos tienen alergia*, o *los pensamientos están resfriados*?

1.4.1.5 Pragmática

La pragmática ha sido definida como la disciplina que estudia el discurso, es decir, el habla, el acto de hablar o de escribir como un acto humano que se dirige a la producción de ciertos efectos. Su campo de interés es más amplio que el mero análisis del discurso. La pragmática trata el origen de los signos, de sus usos y de los efectos que producen en la situación en la cual aparecen.

Cuando una persona (el emisor) dice algo (emite un mensaje) a otra persona (el receptor), intenta siempre producir algún efecto en este último, se busca influir en él. Esta influencia a veces se realiza y a veces no, depende de que el mensaje llegue al destino deseado y el emisor comparta el mismo código lingüístico que el receptor.

Hay muchas formas de influir en el otro. El emisor puede tratar de informar, de preguntar, de pedir, de insultar, de persuadir, de dominar, de halagar, de ordenar, de despreciar, de engañar, de entretener.

Por ejemplo, si alguien dice a otra persona *¿me podría decir qué hora es?*. Esta expresión intenta influir en el receptor del mensaje a que actúe (o se abstenga de actuar) de cierta manera. El oyente puede indicar la hora, que es lo que se pide, o simplemente responder de manera afirmativa a la pregunta sin indicar la hora o ambas cosas.

Otro ejemplo, si queremos comunicar un sentimiento o emoción por algún hecho o acontecimiento. Por ejemplo, *¡por Dios!*, intenta expresar la sorpresa por algún acontecimiento dentro del discurso.

Todas las relaciones entre las oraciones o frases y el mundo externo es el estudio de la pragmática.

1.4.2 El español gramaticalmente

La gramática estudia el sistema de la lengua. Se define la gramática como la parte de la lingüística que estudia el conjunto de reglas de una lengua para formar las palabras y combinarlas para formar oraciones.

Las partes fundamentales de la gramática son la morfología y la sintaxis [11].

1.4.2.1 Morfología

Como se mencionó, la morfología estudia cómo se forman las palabras y qué modificaciones sufren para indicar los distintos accidentes gramaticales de género, número, tiempo, modo. Establece, además, las clases de palabras o categorías gramaticales: sustantivo, adjetivo, verbo, pronombre, artículo, adverbio, preposición y conjunción.

Las palabras están formadas por pequeñas unidades que tienen significado. Estas unidades se llaman morfemas, las que no necesariamente coinciden con las sílabas:

niñ-o
cas-a
roj-os
com-ió

Las palabras anteriores tienen dos morfemas:

- a) El morfema raíz, llamado lexema: **niñ-, cas-, roj-, com-**. Éste se mantiene invariable, generalmente, y porta el significado básico de la palabra.
- b) El morfema flexivo o derivativo, llamado también desinencia: **-o, -a, -os, -ió**. Éste siempre varía y agrega el significado de género, número, tiempo, etc.

Los procesos morfológicos más importantes que presentan las palabras son tres: flexión, derivación y composición [19].

1.4.2.1.1 Flexión

Es el proceso mediante el cual se agrega una determinada desinencia a un morfema raíz, para indicar las variaciones de género, número, tiempo y modo, además, para formar aumentativos, diminutivos, despectivos. Por ejemplo, la palabra *mesa* se le puede agregar la desinencia del plural *-s*: *mesas*. El verbo *comer* se puede flexionar para indicar el modo, tiempo, número y persona: *com-imos*; *comer* y *comimos* son formas verbales.

No todas las palabras tienen este tipo de variación. Las que lo presentan son:

- 1) Los sustantivos, adjetivos, artículos y pronombres pueden tener las desinencias de género y número. Por ejemplo.

Morfemas de género.

<i>escritora</i>	sustantivo femenino
<i>mexicano</i>	adjetivo masculino
<i>una</i>	artículo femenino
<i>ella</i>	pronombre femenino

Morfemas de número.

<i>lápices</i>	sustantivo plural
<i>verdes</i>	adjetivo plural
<i>Las</i>	artículo plural
<i>Ellos</i>	pronombre plural

Para el singular, en español, no se emplea ninguna desinencia: *lápiz, verde, etc.*

- 2) A los sustantivos y adjetivos, se les puede agregar morfemas para formar aumentativos, diminutivos o despectivos. Por ejemplo:

Aumentativos

<i>casa</i>	<i>casona</i>
<i>grande</i>	<i>grandote</i>

Diminutivos

<i>casa</i>	<i>casita</i>
<i>pequeño</i>	<i>pequeñito</i>

Despectivos

<i>casa</i>	<i>casucha</i>
<i>delgado</i>	<i>delgaducho</i>

3) Los adjetivos pueden expresar grado superlativo.

<i>fácil</i>	<i>facilísimo</i>
<i>dulce</i>	<i>dulcísimo</i>

4) Los verbos pueden tener morfemas para indicar modo, tiempo, persona y número (véase 1.4.2.2.3 *El verbo y su papel*).

<i>comprar</i>	<i>compraste</i>
<i>regresar</i>	<i>regresará</i>

En la primera palabra, la desinencia expresa el modo indicativo, tiempo pasado, segunda persona y número singular. En la segunda, modo indicativo, tiempo futuro, tercera persona y singular.

Las preposiciones, las conjunciones y los adverbios, son palabras invariables desde el punto de vista de la flexión. Sin embargo, en el lenguaje coloquial de algunas regiones del mundo, es posible encontrar flexión en algunos adverbios: *ahora* – *ahorita*, *cerca* – *cerquita*, *luego* – *lueguito*.

1.4.2.1.2 Derivación

Es el procedimiento que consiste en agregar un morfema derivativo a un morfema raíz para formar una nueva palabra. Es muy común que los procesos de derivación provoquen cambio en la categoría de las palabras y en su significado. Por ejemplo:

- 1) El morfema **-ción** se agrega a raíces verbales para formar sustantivos.

<i>traducir</i>	<i>traducción</i>
<i>producir</i>	<i>producción</i>
<i>retener</i>	<i>retención</i>

- 2) El morfema **-ble** se agrega a raíces verbales para formar adjetivos.

<i>lavar</i>	<i>lavable</i>
<i>comprar</i>	<i>comprable</i>
<i>recomendar</i>	<i>recomendable</i>

- 3) El morfema **-mente** se agrega a raíces adjetivas para formar adverbios.

<i>fácil</i>	<i>fácilmente</i>
<i>rápido</i>	<i>rápidamente</i>
<i>loco</i>	<i>locamente</i>

1.4.2.1.3 Composición

Es el procedimiento que consiste en unir dos o más palabras para formar una nueva. En la composición pueden participar casi todas las categorías gramaticales. Algunas de las combinaciones más comunes son las siguientes:

- 1) sustantivo + sustantivo.

<i>boca + calle</i>	<i>bocacalle</i>
<i>agua + miel</i>	<i>aguamiel</i>
<i>compra + venta</i>	<i>compraventa</i>

- 2) verbo + sustantivo.

<i>limpiar + botas</i>	<i>limpiabotas</i>
<i>sacar + puntas</i>	<i>sacapuntas</i>
<i>limpiar + parabrisas</i>	<i>limpiaparabrisas</i>

- 3) adjetivo + adjetivo.

<i>claro + oscuro</i>	<i>claroscuro</i>
<i>agrio + dulce</i>	<i>agridulce</i>

4) sustantivo + adjetivo.

<i>boca + abierto</i>	<i>boquiabierto</i>
<i>pelo + rojo</i>	<i>pelirrojo</i>

1.4.2.2 Sintaxis

Las palabras pueden agruparse en categorías gramaticales dependiendo de su estructura, dependiendo de la función que desempeñen dentro de la oración y de su significado. Las categorías que existen en el español son las siguientes: sustantivo, adjetivo, verbo, adverbio, pronombre, artículo, preposición y conjunción.

1.4.2.2.1 El sustantivo y su papel

Los sustantivos son palabras que nombran a seres tanto materiales como abstractos. Sirven para designar personas, animales o cosas que tienen existencia propia, ya sea por la realidad o por abstracción [8].

Clasificación de los sustantivos

Los Sustantivos se clasifican respecto a varios criterios:

a) Sustantivos comunes y propios

Sustantivos comunes. Identifican seres no únicos. Por ejemplo, *país, hombre, planeta, ciudad, lago, río*, etc.

Sustantivos propios. Señalan seres determinados. Su función es de manera identificativa. Son sustantivos propios los nombres de persona, los apellidos, los nombres de lugares (topónimos), etc. En general, cualquier realidad declarada única y bautizada por el hombre. Por ejemplo, *México, España, Rusia, Japón, Venus, Marte*, etc.

b) Sustantivos concretos y abstractos

Sustantivos concretos. Designan seres o cosas reales que son tangibles. Por ejemplo, *persona, animal, perro, gato, casa, manzana, agua*, etc.

Sustantivos abstractos. Son los que nombran a una cualidad sólo aprehensible por el intelecto. Por ejemplo, *belleza, habilidad, torpeza, secreto, ansia, sabiduría*, etc.

c) Sustantivos individuales y colectivos

Sustantivos individuales. Nombran una sola entidad cuando están en forma singular. Por ejemplo, *botella, libro, aparato, cenicero, mujer*, etc.

Sustantivos colectivos. Estando en singular, nombran un conjunto de seres. Por ejemplo, *enjambre, rebaño, ejército, jauría, vestimenta, gente, tropa*, etc.

d) Sustantivos animados e inanimados

Sustantivos animados. Nombran seres considerados vivientes. Por ejemplo, *perro, gato, hombre, persona, niño*, etc.

Sustantivos inanimados. Nombran seres inertes. Por ejemplo, *papel, cristal, piedra, cigarro, carta*, etc.

e) Sustantivos contables e incontables

Sustantivos contables. Señalan realidades que se pueden contar. Por ejemplo, *cinco niños, tres rocas, trece pesos*, etc.

Sustantivos incontables. Señalan realidades que no se pueden contar, salvo al referirse a clases o variedades distintas. Por ejemplo, *leche, humor, aire, humo, basura*, etc.

Los sustantivos contables se combinan con cuantificadores plurales sin alterarse semánticamente. En cambio, los incontables solamente pueden combinarse con cuantificadores en singular sin modificar su significado. Sólo admiten numerales cardinales o cuantificadores plurales cuando indican una clase o modalidad.

Por ejemplo:

tres vasos de leche

muchos vasos de leche

Los humos emitidos por los ácidos suelen ser nocivos

Existen casos en los que sustantivos incontables tienen distintos significados o matices en plural. Por ejemplo:

aires de grandeza

malos humos

aguas internacionales

1.4.2.2.1.1 Categorías gramaticales del sustantivo

Los sustantivos son palabras variables: presentan distintas desinencias para indicar las categorías de género y número; también para formar aumentativos, diminutivos y despectivos.

Género

En español los sustantivos sólo pueden ser masculinos o femeninos. Si se refieren a personas o algunas especies de animales, el género alude a la calidad de mujer o hembra y a la de varón o macho. Por ejemplo.

1. La leona cuida mucho a sus cachorros (sustantivo femenino)
2. El oso polar es de color blanco (sustantivo masculino)

Los sustantivos que se refieren a cosas también tienen género, masculino o femenino, aunque éste no corresponda a una distinción sexual. Se trata de las palabras que adoptaron un género desde los orígenes de la lengua española. Por ejemplo.

1. La luz de la televisión me molesta (sustantivo femenino)
2. El mes pasado fui a Cancún (sustantivo masculino)

Generalmente, la distinción de género se marca mediante el uso de la desinencia **-o**, para el masculino, y **-a** para el femenino, pero existen excepciones a esta regla como las siguientes:

- a) Palabras masculinas terminadas en **-a**: *problema, sistema, esquema, drama*, etc.
- b) Palabras femeninas terminadas en **-o**: *mano, soprano, foto*, etc.

Algunas desinencias comunes para identificar el género son las siguientes:

1. Se consideran palabras masculinas las terminadas en: **tor, sor, dor, or, ma, án, ín, és, ó, e, é, l, x, n, y, o**.

Por ejemplo, *matador, imán, invasor, acné, café, buró, ademán, peón, latex, buey, aerosol*, etc.

2. Se consideran palabras femeninas las terminadas en: **ión, tad, tud, dad, dura, isa, a, á, eza, ina, triz, ez, esis, osis, itis, sis, z, ed**.

Por ejemplo, *canción, bondad, armadura, actriz, avidez, celulitis, catequesis, dosis, red*, etc.

En español hay algunos sustantivos que no presentan variación para marcar el género, por lo que aceptan ser acompañados tanto por el artículo masculino como por el femenino. Por ejemplo.

<i>el mar</i>	<i>la mar</i>
<i>el artista</i>	<i>la artista</i>
<i>el mártir</i>	<i>la mártir</i>

Número

El número de los sustantivos puede ser singular o plural. El singular se refiere a una persona, animal o cosa y carece de desinencia específica. El plural se refiere a dos o más cosas y los morfemas característicos son **-s** y **-es**, de acuerdo con las siguientes reglas:

1) Se añade **-s** a los sustantivos:

- Terminados en vocal no acentuada.

Singular	Plural
<i>pera</i>	<i>peras</i>
<i>calle</i>	<i>calles</i>
<i>batalla</i>	<i>batallas</i>

- Terminados en **-e** tónica

Singular	Plural
<i>café</i>	<i>cafés</i>
<i>pie</i>	<i>pies</i>
<i>té</i>	<i>tés</i>

- Algunos sustantivos de origen extranjero

Singular	Plural
<i>complot</i>	<i>complots</i>
<i>coñac</i>	<i>coñacs</i>
<i>jet</i>	<i>jets</i>

2) Se añade **-es** a los sustantivos:

- Terminados en consonante.

Singular	Plural
<i>comedor</i>	<i>comedores</i>

Singular	Plural
<i>reloj</i>	<i>relojes</i>
<i>pez</i>	<i>peces</i>

Los sustantivos terminados en **-z** hacen el plural con la desinencia **-ces**.

- Terminados en vocal acentuada.

Singular	Plural
<i>rubí</i>	<i>rubíes</i>
<i>bambú</i>	<i>bambúes</i>
<i>tabú</i>	<i>tabúes</i>

Existen excepciones para la regla. Por ejemplo.

Singular	Plural
<i>Sofá</i>	<i>sofás</i>
<i>mamá</i>	<i>mamás</i>
<i>papá</i>	<i>papás</i>
<i>menú</i>	<i>menús</i>

Las excepciones más comunes en la formación del plural son:

- Los sustantivos cuya forma singular terminan en **-s**, no añaden ninguna desinencia para el plural, el cual se indica con el artículo.

Singular	Plural
<i>el lunes</i>	<i>los lunes</i>
<i>la tesis</i>	<i>las tesis</i>
<i>la crisis</i>	<i>las crisis</i>

- Los sustantivos que se emplean sólo en su forma plural.

Singular	Plural
-	<i>nupcias</i>
-	<i>creces</i>
-	<i>viveres</i>

1.4.2.2.1.2 Sustantivos aumentativos, diminutivos y despectivos

Los sustantivos pueden flexionarse para indicar aumento o disminución en el significado, así como para expresar burla o desprecio.

Aumentativos. Las desinencias más comunes son: **-on, -ona, -azo, -aza, -ote, -ota**.

Palabra	Aumentativo
<i>hombre</i>	<i>hombrón</i>
<i>casa</i>	<i>casona</i>
<i>libro</i>	<i>librote</i>

Diminutivos. Las desinencias más comunes son: **-ito, -ita, -illo, -illa, -ico, -ica, -in, -cito, -cita**.

Palabra	Diminutivo
<i>dibujo</i>	<i>dibujito</i>
<i>guitarra</i>	<i>guitarrita</i>
<i>mujer</i>	<i>mujercita</i>
<i>maceta</i>	<i>macetita</i>

Despectivos. Las desinencias más comunes son: **-uza, -aco, -zuelo, -zuela, -ucho, -ucha**; en ocasiones, **-illo, -illa**.

Palabra	Despectivo
<i>gente</i>	<i>gentuza</i>
<i>escritor</i>	<i>escritorzuelo</i>
<i>papel</i>	<i>papelucho</i>
<i>mujer</i>	<i>mujerzuela</i>

1.4.2.2.1.3 Funciones sintácticas del sustantivo

Las funciones sintácticas del sustantivo podemos catalogarlas en las siguientes [12]:

Como sujeto. Este es un uso esencialmente sustantivo, puesto que toda palabra, frase u oración que sirva de sujeto queda sustantivada por este hecho. A los conceptos sustantivos corresponde exclusivamente la función de sujeto.

Por ejemplo.

1. *El amor breve es complicado*
2. *Los niños sueñan despiertos*
3. *El fumar puede causar la muerte*

Como atributo. La significación de atributo es completamente adjetiva. El sustantivo que desempeña el papel de atributo se considera como el portador de un conjunto de cualidades,

y éste queda adjetivado. El atributo sustantivo clasifica al sujeto dentro de un grupo respecto a sus cualidades. Por ejemplo:

1. *Juan era soldado*
2. *El vecino es médico*

Los sustantivos *Juan* y *el vecino* adquieren las cualidades de los atributos *soldado* y *médico* respectivamente.

Como complemento del verbo. Este papel es completamente sustantivo, tanto si se trata del complemento directo, indirecto o circunstancial. Toda palabra, frase u oración que tenga el papel de complemento del verbo, está necesariamente sustantivada, es decir, el verbo sólo puede completar su significación en conceptos sustantivos.

1. *El hombre miró el reloj*
2. *Cambiaba frecuentemente las flores de los jarrones*
3. *Me corto el brazo si ese reloj se atrasa un minuto*

Como complemento de otro sustantivo. Un sustantivo puede determinar, aclarar, o precisar el significado de otro sustantivo, ya sea juntándose a manera de atributo calificativo, *El rey soldado*, o relacionándose por medio de una preposición, *puede de hierro*. La relación entre dos sustantivos se expresa más frecuentemente por medio de una preposición. Todas las preposiciones pueden usarse con los sustantivos complementarios. *Árbol sin fruto, café con leche, lucha por la libertad*, etc.

Como complemento de un adjetivo. La cualidad que expresa un adjetivo respecto al sustantivo que califica, puede limitarse mediante otro sustantivo precedido de una preposición. Por ejemplo, *el maestro es sabio*, en este caso se atribuye el adjetivo *sabio* de manera general, pero podía limitarse a *el maestro es sabio en matemáticas*. De esta manera los sustantivos con preposición determinan a los adjetivos: *amable con las damas, apto para el estudio, humilde ante los hombres*, etc.

1.4.2.2.2 El adjetivo y su papel

Un adjetivo necesita referirse a un sustantivo, el cual añade notas que lo determinan o precisan. La función del adjetivo es determinar o calificar al sustantivo, cualquiera que sea el oficio que éste desempeñe en la oración.

1.4.2.2.1 Posición del adjetivo respecto al sustantivo

El adjetivo en español, como es sabido, puede preceder o seguir al sustantivo a que se refiere, pero el valor expresivo no es el mismo en uno u otro caso. Bello [12] pone atención en el significado lógico de la anteposición y de la aposición. Bello advirtió que el adjetivo pospuesto determina o restringe la extensión del sustantivo. Si se dice *un edificio hermoso* se excluyen todos los edificios que no sean hermosos. En cambio, en *un hermoso edificio* se añade al sustantivo una cualidad, pero no la limita.

Anteposición del adjetivo

El adjetivo antepuesto supone por parte del hablante mayor atención hacia la cualidad que hacia el sustantivo: *verde prado*, *altas torres*, *buen persona*, etc.; matizan subjetivamente la expresión envolviendo al sustantivo que sigue.

Posposición del adjetivo

El adjetivo pospuesto se suma a la representación previa del objeto. Expresa una cualidad relativamente característica, pero no la realza. Por ejemplo, *prado verde*, *torres altas*, *persona buena*, etc. Se tiene un carácter objetivo.

La preferencia por una u otra posición, cuando las condiciones lógicas no exigen colocación fija, es sólo un elemento de caracterización de estilo.

1.4.2.2.2 Adjetivos determinativos

Los adjetivos determinativos se anteponen generalmente al sustantivo, por ejemplo, *veinte vacas*, *esta casa*, *muchos hombres*, etc.

Existen algunas excepciones:

- a) Los numerales cardinales, cuando se usan como ordinales, van siempre pospuestos, como *el siglo XXI*, *el día 13*, *Luis XVI*, etc.
- b) Los ordinales pueden anteponerse o posponerse, ‘*el primer día*’ o ‘*el día primero*’, ‘*la segunda columna*’ o ‘*la columna segunda*’, etc.
- c) El partitivo *medio* va antepuesto cuando no hay otro numeral y pospuesto cuando existe otro numeral, ‘*media hora*’ y ‘*dos horas y media*’, respectivamente.
- d) Los demostrativos y posesivos van comúnmente antepuestos, pero se posponen cuando están acompañados del artículo u otra palabra determinativa, ‘*la casa esa*’,

'*el día aquel*', etc. Los demostrativos *este* y *ese* se usan también de forma despectiva: '*el hombre ese*', '*la mujer esta*', etc.

Son adjetivos demostrativos: *este, esta, ese, esa, aquel, aquella*.

Son adjetivos posesivos: *mío, mía, tuyo, tuya, suyo, suya, nuestro, nuestra, vuestro, vuestra*.

Existen algunos adjetivos que cambian su significación según el lugar que ocupan con respecto al sustantivo y que por tanto tienen una colocación fija.

<i>cierta noticia</i> (noticia indeterminada)	<i>noticia cierta</i> (noticia verídica)
<i>pobre hombre</i> (un hombre desdichado)	<i>hombre pobre</i> (hombre con necesidades)
<i>simple soldado</i> (un soldado únicamente)	<i>soldado simple</i> (un soldado tonto)

1.4.2.2.3 Adjetivos apocopados

El fenómeno de apocopado consiste en suprimir uno o varios sonidos al final de ciertos adjetivos cuando se anteponen al sustantivo. Algunos de los adjetivos que tienen forma apocopada son los siguientes:

Forma completa	Apócope
<i>alguno</i>	<i>algún</i>
<i>bueno</i>	<i>buen</i>
<i>malo</i>	<i>mal</i>
<i>ciento</i>	<i>cien</i>
<i>cualquiera</i>	<i>cualquier</i>
<i>grande</i>	<i>gran</i>
<i>ninguno</i>	<i>ningún</i>
<i>primero</i>	<i>primer</i>
<i>santo</i>	<i>san</i>
<i>tercero</i>	<i>tercer</i>

1.4.2.2.4 Categorías gramaticales del adjetivo

Los adjetivos son palabras variables. Presentan distintas desinencias para indicar las categorías de género y número; también pueden formar aumentativos, diminutivos y despectivos.

Categoría de género

Los adjetivos pueden ser femeninos o masculinos y deben concordar con el género del sustantivo al que acompañan. Por ejemplo, *niño bonito* y *niña bonita*, *vecino atento* y *vecina atenta*.

- a) Los adjetivos masculinos, generalmente, tienen terminación **-o**. Por ejemplo, *interesado*, *feo*, *amarillo*, *absoluto*, etc.
- b) Los adjetivos femeninos, generalmente, tienen terminación **-a**. Por ejemplo, *interesada*, *fea*, *amarilla*, *absoluta*, etc.
- c) Otros adjetivos se consideran masculinos si terminan en: **-tor, -dor, -or, -án, -ín**. Por ejemplo, *protector*, *bebedor*, *holgazán*, *saltarán*. Para generar el correspondiente femenino se agrega la desinencia **-a**, *protectora*, *bebedora*, etc.
- d) Los adjetivos que no presentan variación, generalmente, presentan las terminaciones **-nte, -ble, -ista, -e, -ata, -ita, -al, -il, -az** y los que terminan en consonante, **-n, -s, -l**. Por ejemplo, *estudiante*, *amable*, *altruista*, *fundamental*, *israelita*, *audaz*, *sutil*, *joven*, etc.

Categoría de número

Los adjetivos sólo pueden estar en singular o plural. Siempre concuerdan con el sustantivo al que acompañan. Por ejemplo, *árbol seco* y *árboles secos*, *estrella luminosa* y *estrellas luminosas*, etc.

- a) Generalmente, al adjetivo se le agrega la desinencia **-s** para formar el plural. Por ejemplo, *rojo – rojos*, *sucia – sucias*, *brillante – brillantes*, etc.
- b) Cuando el adjetivo singular termina en consonante, se agrega la desinencia **-es** para el plural. Por ejemplo, *azul – azules*, *fácil – fáciles*, *útil – útiles*, etc.
- c) Los adjetivos terminados en **-z**, forman el plural con **-ces**. Por ejemplo, *feroz – feroces*, *audaz – audaces*, *capaz – capaces*, etc.
- d) Son escasos los adjetivos que únicamente tienen forma plural. Estos adjetivos no tienen variación. La misma forma se usa para el singular como para el plural. Por ejemplo, *isósceles*, *rubiales*, etc.

triángulo isósceles

triángulos isósceles

1.4.2.2.5 Adjetivos aumentativos, diminutivos y despectivos

Los adjetivos pueden flexionarse para indicar aumento o disminución en el significado, así como para expresar burla o desprecio, despectivos. Las desinencias que suelen usarse son las mismas que se emplean para el sustantivo. La siguiente tabla muestra algunos ejemplos.

Palabra	Aumentativo	Diminutivo	Despectivo
<i>flaco</i>	<i>flacote</i>	<i>flaquito</i>	<i>flacucho</i>
<i>soltero</i>	<i>solterote</i>	<i>solterito</i>	<i>solterillo o solterón</i>
<i>viejo</i>	<i>viejote</i>	<i>viejito</i>	<i>viejillo</i>

1.4.2.2.3 El verbo y su papel

El verbo es la clase de palabra que expresa acciones, actitudes, cambios, movimientos de seres o cosas. Siempre se refiere a las actividades que realizan o padecen las personas o animales, las situaciones o estados en que se encuentran, los cambios que sufren los objetos, las manifestaciones de diversos fenómenos de la naturaleza.

La manera que se emplea para enunciar los verbos es la forma infinitiva; ésta no expresa modo, tiempo, número ni persona. Los verbos se agrupan según su terminación y siguen el paradigma de conjugación según ésta. Tenemos tres formas *-ar*, *-er* y *-ir*. Por ejemplo, *amar*, *caminar*, etc. pertenecen a la primera forma; *comer*, *correr*, etc. pertenecen a la segunda forma; *vivir*, *salir*, etc. pertenecen a la tercera forma.

El verbo presenta variaciones en sus desinencias para indicar la persona que realiza la acción, el número de la persona, así como el modo y tiempo en que se realiza. Además, se distingue en el verbo un morfema invariable llamado raíz. Por ejemplo, *caminamos*, presenta las siguientes características:

Camin- Lexema, aporta el contenido semántico de la palabra.

-amos Indica, primera persona del plural, modo indicativo y tiempo presente.

1.4.2.2.3.1 Categorías gramaticales

En los verbos es posible identificar un morfema invariable llamado raíz y un morfema variable que expresa las diferentes categorías: persona, número, modo, tiempo.

1.4.2.2.3.1.1 Persona y número

Mediante una desinencia, los verbos marcan la persona gramatical que realiza la acción, sea singular o plural.

	Singular	Plural
Primera persona	(yo) <i>camin-o</i>	(nosotros) <i>camin-amos</i>
Segunda persona	(tú) <i>camin-as</i> (usted) <i>camin-a</i>	(vosotros/as) <i>camin-áis</i> (ustedes) <i>camin-an</i>
Tercera persona	(él, ella) <i>camin-a</i>	(ellos/as) <i>camin-an</i>

Las formas verbales de tercera persona singular y plural se emplean también con los pronombres de segunda persona, *usted* y *ustedes*, respectivamente.

1.4.2.2.3.1.2 Modo

Con el modo se expresa el punto de vista relativo ante la acción verbal que se enuncia. En español existen tres modos: indicativo, subjuntivo e imperativo [11].

El modo indicativo se usa, generalmente, para referir hechos reales, ya sea en el tiempo presente, pasado o futuro. Por ejemplo.

Los buitres planean como siempre sobre prometedoras agonías

El cantor va por el mundo sonriente o meditabundo

Descansaremos en las playas orientales

El modo subjuntivo se usa, generalmente, para expresar una acción posible, de deseo, de creencia, de duda. Por ejemplo.

Quiero que Jacinta cocine

Siempre temí que pasara esto

No sé si haya terminado el trabajo

El modo imperativo expresa súplica, mandato, o ruego. El imperativo sólo tiene las formas de segunda persona, singular y plural. Por ejemplo.

Apaga la luz

Escuchen ese ruido

Caminen aprisa

1.4.2.2.3.1.3 Tiempo

Señala el momento en que se realiza la acción: los tiempos básicos son presente, pasado y futuro.

Los tiempos verbales pueden ser simples o compuestos. Los primeros se forman a partir de la raíz del verbo, añadiendo una desinencia específica. Por ejemplo.

cant-o
cant-é
cant-aré

Para formar los tiempos compuestos se utiliza el verbo *haber* como auxiliar conjugado y el participio pasado del verbo de que se trate. Por ejemplo.

he cantado
hube cantado
habré cantado

Los verbos regulares son los que siguen modelos de conjugación. Los terminados en *-ar* siguen el modelo del verbo *amar* (Tabla 1.1), los terminados en *-er*, siguen el modelo del verbo *comer* (Tabla 1.2), y los terminados en *-ir*, siguen el modelo del verbo *vivir* (Tabla 1.3).

1.4.2.2.3.2 Modelos de conjugación

Se presenta a continuación los modelos de conjugación de los verbos regulares, en los tiempos simples del modo indicativo [14].

Tabla 1.1 Modelo para verbos de primera conjugación: *amar*

TIEMPOS SIMPLES					
Persona	Presente	Pasado	Futuro	Copretérito	Pospretérito
yo	<i>am-o</i>	<i>am-é</i>	<i>am-aré</i>	<i>am-aba</i>	<i>am-aría</i>
tú	<i>am-as</i>	<i>am-aste</i>	<i>am-arás</i>	<i>am-abas</i>	<i>am-arías</i>
usted	<i>am-a</i>	<i>am-ó</i>	<i>am-ará</i>	<i>am-aba</i>	<i>am-aría</i>
él / ella	<i>am-a</i>	<i>am-ó</i>	<i>am-ará</i>	<i>am-aba</i>	<i>am-aría</i>
nosotros	<i>am-amos</i>	<i>am-amos</i>	<i>am-aremos</i>	<i>am-ábamos</i>	<i>am-aríamos</i>
vosotros	<i>am-áis</i>	<i>am-asteis</i>	<i>am-areis</i>	<i>am-abais</i>	<i>am-aríais</i>
ustedes	<i>am-an</i>	<i>am-aron</i>	<i>am-arán</i>	<i>am-aban</i>	<i>am-arían</i>
ellos	<i>am-an</i>	<i>am-aron</i>	<i>am-arán</i>	<i>am-aban</i>	<i>am-arían</i>

Tabla 1.2 Modelo para verbos de segunda conjugación: *comer*

TIEMPOS SIMPLES					
Persona	Presente	Pasado	Futuro	Copretérito	Pospretérito
yo	<i>com-o</i>	<i>com-í</i>	<i>com-eré</i>	<i>com-ía</i>	<i>com-ería</i>
tú	<i>com-es</i>	<i>com-iste</i>	<i>com-erás</i>	<i>com-ías</i>	<i>com-erías</i>
usted	<i>com-e</i>	<i>com-ió</i>	<i>com-erá</i>	<i>com-ía</i>	<i>com-ería</i>
él / ella	<i>com-e</i>	<i>com-ió</i>	<i>com-erá</i>	<i>com-ía</i>	<i>com-ería</i>
nosotros	<i>com-emos</i>	<i>com-imos</i>	<i>com-eremos</i>	<i>com-íamos</i>	<i>com-eríamos</i>
vosotros	<i>com-éis</i>	<i>com-isteis</i>	<i>com-eréis</i>	<i>com-íais</i>	<i>com-eríais</i>
ustedes	<i>com-en</i>	<i>com-ieron</i>	<i>com-erán</i>	<i>com-ían</i>	<i>com-erían</i>
ellos	<i>com-en</i>	<i>com-ieron</i>	<i>com-erán</i>	<i>com-ían</i>	<i>com-erían</i>

Tabla 1.3 Modelo para verbos de tercera conjugación: *vivir*

TIEMPOS SIMPLES					
Persona	Presente	Pasado	Futuro	Copretérito	Pospretérito
yo	<i>viv-o</i>	<i>viv-í</i>	<i>viv-iré</i>	<i>viv-ía</i>	<i>viv-iría</i>
tú	<i>viv-es</i>	<i>viv-iste</i>	<i>viv-irás</i>	<i>viv-ías</i>	<i>viv-irías</i>
usted	<i>viv-e</i>	<i>viv-ió</i>	<i>viv-irá</i>	<i>viv-ía</i>	<i>viv-iría</i>
él / ella	<i>viv-e</i>	<i>viv-ió</i>	<i>viv-irá</i>	<i>viv-ía</i>	<i>viv-iría</i>
nosotros	<i>viv-imos</i>	<i>viv-imos</i>	<i>viv-iremos</i>	<i>viv-íamos</i>	<i>viv-iríamos</i>
vosotros	<i>viv-ís</i>	<i>viv-isteis</i>	<i>viv-iréis</i>	<i>viv-íais</i>	<i>viv-iríais</i>
ustedes	<i>viv-en</i>	<i>viv-ieron</i>	<i>viv-irán</i>	<i>viv-ían</i>	<i>viv-irían</i>
ellos	<i>viv-en</i>	<i>viv-ieron</i>	<i>viv-irán</i>	<i>viv-ían</i>	<i>viv-irían</i>

1.4.2.2.3.3 Clases de verbos

Los verbos que no siguen los modelos anteriores se consideran irregulares, pues presentan variaciones en su conjugación.

En general, no se consideran irregularidades los cambios de acentuación. Tampoco son irregularidades los cambios ortográficos que sufren algunos verbos [14]. Por ejemplo:

- 1) La letra **-c-**, con sonido fuerte, se escribe **-qu-** ante **-e-**.

indicar

indique

replicar

replique

- 2) La letra **-g-**, con sonido suave, se escribe **-gu-** ante **-e**.

<i>pagar</i>	<i>pague</i>
<i>regar</i>	<i>regué</i>

- 3) La letra **-z-**, se escribe **-c-** ante **-e**.

<i>rozar</i>	<i>roce</i>
<i>agilizar</i>	<i>agilice</i>

- 4) La letra **-c-** y **-g-**, se escribe **-z-** y **-j-**, respectivamente, ante **-a** y **-o**.

<i>ejercer</i>	<i>ejerza</i>
<i>zurcir</i>	<i>zurzo</i>
<i>recoger</i>	<i>recoja</i>
<i>fingir</i>	<i>finjo</i>

- 5) La letra **-u-** de los verbos terminados en **-guir-** se pierde ante **-a** y **-o**.

<i>distinguir</i>	<i>distinga</i>	<i>distingo</i>
<i>perseguir</i>	<i>persiga</i>	<i>persigo</i>

Verbos irregulares son aquellos que, al flexionarse, presentan alteraciones en su raíz o en su terminación, es decir, no siguen la conjugación del modelo al que pertenecerían por su desinencia de infinitivo, como *amar*, *comer*, o *vivir*.

Son pocos los verbos que presentan irregularidades tales como las de *ser* o *ir*, que tienen varias raíces, y por lo tanto tienen formas muy diversas como las siguientes:

ser	<i>soy</i>	<i>seré</i>	<i>es</i>	<i>era</i>	<i>fui</i>	<i>fuera</i>
ir	<i>vas</i>	<i>van</i>	<i>iba</i>	<i>ibas</i>	<i>fui</i>	<i>fuera</i>

Algunas de las irregularidades más comunes se enlistan a continuación:

- 1) Diptongación. En ocasiones, las vocales **-i-** y **-e-** diptongan en **-ie-** y las vocales **-o** y **-u** en **-ue-**. Por ejemplo, *acertar*, *calentar*, *fregar*, *comenzar*, *confesar*, *defender*, *encender*, *extender*, *gobernar*, *manifestar*, *merendar*, *almorzar*, *mostrar*, *mover*, *oler*, *probar*, *resolver*, *soñar*, *volar*, *volver*, etc.

<i>adquirir</i>	<i>adquiero</i>
<i>poder</i>	<i>puedo</i>
<i>pensar</i>	<i>pienso</i>
<i>jugar</i>	<i>juego</i>

- 2) Cambio de una vocal. Las vocales **-e** y **-o** cambian a **-i** y **-u**, respectivamente. Por ejemplo, *vestir, competir, concebir, elegir, freír, medir, pedir, reír, rendir, seguir, teñir*, etc.

<i>pedir</i>	<i>pido</i>
<i>poder</i>	<i>pude</i>
<i>concebir</i>	<i>concibo</i>
<i>morir</i>	<i>murió</i>

- 3) Sustitución de una letra por otra, **-c**, por **-g** o **-j**.

<i>hacer</i>	<i>haga</i>
<i>satisfacer</i>	<i>satisfaga</i>
<i>aducir</i>	<i>adujo</i>
<i>conducir</i>	<i>conduje</i>

- 4) Adición de una letra, **-d**, **-z** o **-g**.

<i>poner</i>	<i>pondré</i>
<i>tener</i>	<i>tendré</i>
<i>nacer</i>	<i>nazco</i>
<i>parecer</i>	<i>parezco</i>
<i>poner</i>	<i>pongo</i>

- 5) Modificación de dos o más letras.

<i>decir</i>	<i>diga</i>
<i>saber</i>	<i>sepa</i>
<i>caber</i>	<i>quepa</i>
<i>traer</i>	<i>traiga</i>

Verbos defectivos. Son los verbos que sólo se conjugan en algunas formas y carecen de otras. La mayoría de ellos sólo tienen la conjugación para la tercera persona, debido a su significado. Por ejemplo.

<i>atañer</i>	<i>atañe</i>	<i>atañen</i>
<i>acontecer</i>	<i>acontece</i>	<i>acontecen</i>
<i>concernir</i>	<i>concierne</i>	<i>conciernen</i>

El verbo *abolir* también es defectivo porque no se conjuga en todos los tiempos y personas.

Hay otros verbos defectivos de uso frecuente: *balbucir, balbucí; agredir, agredió; abolir, abolió; transgredir, transgredió; soler, suele, solía, solió.*

Verbos Impersonales. Son los verbos que sólo se conjugan en tercera persona del singular, en todos los tiempos, porque no tienen un sujeto determinado; se refieren a fenómenos meteorológicos. Por ejemplo, *llover, nevar, amanecer, anochecer, etc.*

Sin embargo, cuando el verbo se emplea en sentido figurado es posible atribuirles un sujeto, con lo que pierden el sentido de impersonalidad. Por ejemplo.

Llovieron piedras

Su vida anocheció demasiado pronto

1.4.2.2.3.4 Formas no personales del verbo

Las formas no personales del verbo no presentan variación para indicar persona, tiempo ni modo. Son el infinitivo, gerundio y participio.

1) Infinitivo

Es el nombre del verbo, es la expresión de la acción verbal. Las terminaciones son las ya mencionadas: *-ar, -er e -ir.* Por ejemplo, *ordenar, colgar, moler, asistir, etc.*

El infinitivo admite uno o dos pronombres enclíticos:

<i>golpear</i>	<i>golpear<u>te</u></i>	<i>golpear<u>nos</u></i>	<i>golpear<u>se</u></i>	<i>golpear<u>selos</u></i>
<i>sostener</i>	<i>sostener<u>me</u></i>	<i>sostener<u>la</u></i>	<i>sostener<u>se</u></i>	<i>sostener<u>selos</u></i>
<i>regir</i>	<i>regir<u>lo</u></i>	<i>regir<u>nos</u></i>	<i>regir<u>se</u></i>	<i>regir<u>selos</u></i>

2) Gerundio

Expresa una acción continuada, en progreso. Sus terminaciones son **-ando, -iendo.**

<i>golpear</i>	<i>golpeando</i>
<i>sostener</i>	<i>sosteniendo</i>
<i>asistir</i>	<i>asistiendo</i>

De la misma forma que el infinitivo, el gerundio acepta uno o dos pronombres enclíticos:

<i>dar</i>	<i>dá<u>ndole</u></i>	<i>dá<u>ndome</u></i>	<i>dá<u>ndonoslo</u></i>
<i>temer</i>	<i>temi<u>éndote</u></i>	<i>temi<u>éndose</u></i>	<i>temi<u>éndoselo</u></i>
<i>medir</i>	<i>midi<u>éndonos</u></i>	<i>midi<u>éndola</u></i>	<i>midi<u>éndoselos</u></i>

3) Participio

Expresa un acción ya realizada. Sus terminaciones son **-ado**, **-ido** y las irregulares **-to**, **-so**, **-cho**.

<i>calcular</i>	<i>calculado</i>
<i>salir</i>	<i>salido</i>
<i>escribir</i>	<i>escrito</i>
<i>decir</i>	<i>dicho</i>

Los participios, sí marcan género y número. Se emplean en la formación de perífrasis verbal; también es muy común usarlos como adjetivos.

<i>fueron <u>aprobados</u></i>	(perífrasis verbal)
<i>ha <u>caminado</u></i>	(perífrasis verbal)
<i>hombre <u>engreído</u></i>	(adjetivo)
<i>muchacha <u>alocada</u></i>	(adjetivo)

Los participios no admiten pronombres enclíticos.

1.4.2.2.3.5 Funciones sintácticas del verbo

La función propia del verbo es ser núcleo del predicado al que se refieren directa o indirectamente todos los complementos del sintagma [14].

Complementos del verbo

A veces los complementos son indispensables para el sentido del verbo. Sólo algunos verbos exigen casi siempre complemento por necesidad de su propia significación (*dar*, *comprar*, etc.). Los complementos del verbo, atendiendo a la función sintáctica pueden ser de tres clases: directos, indirectos y circunstanciales.

- 1) **En el complemento directo** (CD) recae inmediatamente la acción verbal, es decir, expresa la cosa hecha por el verbo. Por ejemplo, en *el perro comió la carne*, la cosa comida es *la carne* y el complemento directo de la oración. En *Juan escribe una carta*, la cosa escrita es *una carta*, que es el complemento directo.

- 2) **El complemento indirecto (CI)** expresa la persona o cosa que recibe daño o provecho de la acción del verbo. El complemento indirecto lleva siempre las preposiciones *a* o *para*. Por ejemplo:

*Los patinadores ofrecieron un gran espectáculo al público*¹

María explicó el tema a sus papás

Escribí un poema para mi novia

- 3) **El complemento circunstancial (CC)** expresa el lugar, modo, tiempo, medio, causa o instrumento de la acción verbal. Por ejemplo:

Cumpliré de buena gana el encargo de usted (modo)

La carta fue escrita con la pluma nueva (instrumento)

El director no se presentó por causas de fuerza mayor (causa)

Este complemento, generalmente, usa las preposiciones: *con, de, desde, en, hacia, hasta, por, sin, sobre, tras*, etc. y algunas veces *a* y *para* con significado distinto del uso en el CD.

Valencias del verbo

Las valencias de las palabras son similares a las valencias de los átomos de la química [31]. Las valencias de la palabra indican qué clase de palabras pueden combinarse con ella y cuántas palabras.

El uso de las valencias en los verbos es para describir sus valores sintácticos y semánticos. Por ejemplo, el verbo *dar* tiene la siguiente descripción semántica.

persona X da alguna cosa Y a la persona Z.

Ejemplo.

Juan (X) le dió una pelota (Y) a Luis (Z)

Este verbo acepta tres valencias (X, Y y Z).

Los verbos permiten cierto número de valencias. Algunos verbos no permiten valencias, otros una, dos, tres, cuatro, e incluso, cinco valencias, algunas obligatorias y otras no [7].

Mostraremos algunos ejemplos para algunos tipos de valencias [7].

¹ El CI es únicamente *el público*; *al* es la contracción de *a + el*.

a) Algunos verbos españoles intransitivos tienen únicamente la valencia que corresponde al sujeto, por ejemplo el verbo *cojear*:

cojear

persona o animal X camina cojamente

Ejemplos.

El hombre (X) cojea

El gato (X) cojea

Por definición, los verbos intransitivos no pueden tener un complemento directo. Sin embargo, la ausencia del complemento directo es una peculiaridad puramente sintáctica. Éstos pueden tener otras valencias representadas mediante diversos complementos indirectos.

perecer

X cesa de vivir debido a la causa Y

Ejemplo.

El hombre (X) perece de hambre (Y)

b) Los verbos transitivos, por definición, tienen una segunda valencia semántica denominada en el nivel sintáctico como complemento directo. Para estos verbos el complemento directo es obligatorio. Puede ser adyacente o a través de una preposición. Por ejemplo, **el niño quiere*, es incorrecta, porque el verbo requiere la indicación explícita de qué es lo que quiere.

querer

persona X experimenta sentimientos positivos hacia la persona Y

Ejemplo.

El niño(X) quiere a su mamá (Y, obligatoria)

c) Los verbos con tres valencias son considerados en la gramática tradicional como doble transitivos. La tercera valencia es la denominada complemento indirecto. En el español, los complementos indirectos siempre están unidos al verbo mediante preposiciones. Por ejemplo, el verbo *solicitar*.

solicitar

X solicita algo Y de Z

Ejemplo.

* Indica que es una construcción no gramatical del español.

El secretario (X) solicita una entrevista (Y, obligatoria) con el presidente (Z)

1.4.2.2.3.6 Uso del pronombre *se*

En ocasiones, el pronombre *se* (y los demás pronombres átonos correspondientes: *me, te, nos, os*) se convierte en componente inseparable del verbo y forma con él lo que se llama sintagma pronominal. Por ejemplo, *arrepentirse, quejarse, dormirse, irse, marcharse, arrodillarse, jactarse, enorgullecerse*, etc. Cuando el pronombre está unido al verbo se denomina enclítico.

Los verbos pueden expresarse en voz activa o en voz pasiva [11]. En la primera, el sujeto es el que realiza la acción. En la segunda, el sujeto es el que recibe la acción del verbo. Por ejemplo:

La profesora revisa los ejercicios (voz activa, sujeto)

Los ejercicios son revisados por la profesora (voz pasiva, sujeto)

El empleo de las oraciones pasivas es poco frecuente en español; en cambio aumenta el uso de las oraciones pasivas reflejas con el pronombre *se* y con el verbo en voz activa [12]. Por ejemplo, ‘*La paz fue firmada por los embajadores*’ equivale a ‘*Se firmó la paz por los embajadores*’, ‘*El acuerdo fue aceptado por el sindicato*’ equivale en pasiva a ‘*Se aceptó el acuerdo por el sindicato*’. Si desaparece el interés hacia el sujeto agente se dirá: ‘*Se firmó la paz*’ y ‘*Se aceptó el acuerdo*’.

1.4.2.2.4 El adverbio y su papel

Es la parte de la oración que sirve para modificar, precisar, matizar o ampliar el significado de un verbo, de un adjetivo o de otro adverbio [11]. Se caracteriza por su invariabilidad, con algunas excepciones, ya que muchos adverbios pueden admitir morfemas de sufijación. Por ejemplo, *ahora – ahorita, lejos – lejotes*, etc. Los adverbios carecen de morfemas de concordancia con otros elementos de la oración. Por ejemplo:

Luisa está mal

Los niños están mal educados

La función del adverbio puede desempeñarla una palabra simple, una compuesta, una frase o una oración:

Iré a verte mañana

Habla más despacio

Nos quedamos a oscuras

Te visitaremos cuando vayamos a México

Hay adverbios con formas simples: *hoy, mañana*, etc.; con formas compuestas: *anteayer, anteanoche*, etc.; con formas complejas como las locuciones adverbiales: *a lo grande, de vez en cuando, a ciegas*, etc.

1.4.2.2.4.1 Adverbialización

El adjetivo adverbializado es también un adverbio: *bajo, alto, oscuro, rápido*; también se adverbializa con el sufijo *mente*, que se forma como sigue. Si el adjetivo tiene dos terminaciones se utiliza la forma femenina: *buena (buenamente), loca (locamente)*, etc; si el adjetivo sólo tiene una terminación, se añade el sufijo *mente*: *atrozmente, felizmente*, etc.

1.4.2.2.4.2 Sufijación

Muchos adverbios tienen la posibilidad de admitir morfemas de sufijación para formar aumentativos, diminutivos o superlativos.

Aumentativos. *lejotes, arribota*, etc.

Diminutivos. *despacito, cerquita, ahorita, lueguito, poquito, prontito, tempranito*, etc.

Superlativos. *cerquísima, prontísimo, tardísimo, poquísimo, muchísimo, tempranísimo*, etc.

1.4.2.2.4.3 Gradación

Algunos adverbios admiten la gradación, característica de los adjetivos, por medio de la anteposición de cuantificadores (*más, menos, mucho, poco, muy*). Por ejemplo, *mucho mejor, poco antes, muy abajo, más cerca*, etc.

1.4.2.2.4.4 Clasificación de los adverbios

Los adverbios se pueden clasificar tradicionalmente de varias formas. En cuanto a su significado, existen adverbios de [8]:

- 1) **Tiempo.** *luego, ahora, antes, después, ayer, hoy, mañana, entonces, tarde.*
- 2) **Lugar.** *cerca, lejos, aquí, allí, arriba, abajo, fuera, alrededor, allá, ahí.*
- 3) **Modo.** *bien, mal, así, despacio, deprisa, aprisa, gratis.*

- 4) **Cantidad e intensidad.** *más, menos, poco, bastante, demasiado, muy, mucho, apenas, casi, medio, algo, nada.*
- 5) **Afirmación.** *sí, también, ciertamente, claro, desde luego, en efecto, asimismo.*
- 6) **Negación.** *no, nunca, jamás, tampoco.*
- 7) **Duda.** *acaso, quizá o quizás, tal vez, probablemente.*
- 8) **Demostrativos.** *aquí, entonces, ahora, así, luego, tal, tanto.*
- 9) **Relativos.** *donde, como, cuanto, cuando.*
- 10) **Interrogativos.** *cuándo, dónde, cómo, cuánto, qué.*

1.4.2.2.4.5 Locuciones adverbiales

Las locuciones o frases adverbiales son expresiones formadas por grupos de palabras que han quedado en el uso de la lengua como expresiones fijas y que adquieren la función y valor del adverbio. Por ejemplo, *a escondidas, a hurtadillas, a ciegas*, etc. Están compuestas por una preposición unida a sustantivos, adverbios o adjetivos: *enfrente, encima, despacio, debajo, acaso, adentro*, etc.

1.4.2.2.5 El pronombre y su papel

El pronombre es la parte de la oración que se emplea para sustituir un sustantivo y evitar, en ocasiones, su repetición. El pronombre carece de significado propio, lo adquiere en el contexto. La mayoría de los pronombres presentan desinencias para marcar el género y número. Los pronombres se clasifican como [11]:

- 1) **Personales.** Se refieren a las distintas personas gramaticales que intervienen en el diálogo.

Primera persona.

Singular: *yo, mí, me, conmigo.*

Plural: *nosotros, nosotras, nos.*

Segunda persona.

Singular: *tú, usted, ti, te, contigo.*

Plural: *ustedes, vosotros, vosotras, os.*

Tercera persona.

Singular: *él, ella, ello, sí, se, consigo, lo, la, le.*

Plural: *ellos, ellas, los, las, les, sí, se.*

2) **Demostrativos.** Señalan seres o cosas sin nombrarlos.

Singular.

Masculino: *éste, ése, aquél.*Femenino: *ésta, ésa, aquélla.*Neutro: *esto, eso, aquello.*

Plural.

Masculino: *éstos, éstos, aquéllos.*Femenino: *éstas, éstas, aquéllas.*3) **Poseivos.** Se refieren a seres, cosas o ideas poseídas por alguien.

Primera persona.

Singular: *mío, mía, nuestro, nuestra.*Plural: *míos, mías, nuestros, nuestras.*

Segunda persona.

Singular: *tuyo, tuya, vuestro, vuestra.*Plural: *tuyos, tuyas, vuestros, vuestras.*

Tercera persona.

Singular: *suyo, suya.*Plural: *suyos, suyas.*4) **Relativos.** Se refieren a alguien o algo, que se ha mencionado en el discurso.Singular: *que, quien, cual, cuanto, cuanta.*Plural: *quienes, cuales, cuantos, cuantas.*5) **Indefinidos.** Se refieren a seres o cosas cuya identidad o cantidad es imprecisa. Entre los más usuales, *alguien, nadie, algo, nada, cualquiera, alguno, todo, ninguno, muchos, pocos, varios, demasiados, otro, otra, bastantes, uno, etc.***1.4.2.2.6 El artículo y su papel**

Es la parte de la oración que precede al sustantivo para determinarlo y concuerda con él en género y número [11].

El cielo (singular - masculino)

La boca (singular - femenino)

Los cuadros (plural - masculino)

Las estrellas (plural - femenino)

Los artículos se clasifican en determinados o indeterminados.

	Artículos determinados			Artículos indeterminados	
	Masculino	Femenino	Neutro	Masculino	Femenino
Singular	<i>el</i>	<i>la</i>	<i>lo</i>	<i>un</i>	<i>una</i>
Plural	<i>los</i>	<i>las</i>		<i>unos</i>	<i>unas</i>

Los artículos determinados se refieren, generalmente, a seres o cosas previamente conocidos por los hablantes. Su presencia es casi siempre necesaria junto al sustantivo y se emplea para singularizar una entidad en particular o un conjunto genérico.

El cisne antes cantaba sólo para morir

Los mismos ruiseñores cantaban los mismos trinos

Los artículos indeterminados se refieren, generalmente, a seres o cosas no conocidos o imprecisos para el hablante.

Sólo una imagen lo vencía

Se escucharon unos disparos

El artículo neutro se usa para sustantivar un adjetivo. Se emplea en su forma masculina singular y adquiere un sentido abstracto.

lo raro

lo difícil

lo escandaloso

El artículo tiene la propiedad de sustantivar cualquier clase de palabra.

El vivir cómodamente era una obsesión (verbo sustantivado)

El ayer era para ellos un tiempo olvidado (adverbio sustantivado)

Un rojo pálido se veía en el horizonte (adjetivo sustantivado)

Los artículos *el* y *un* se emplean con sustantivos femeninos en singular que empiezan con **a** o **-ha** con acento ortográfico o prosódico

El vuelo del águila

El nivel del agua

El verdugo dejó caer con gran energía el hacha sobre el acusado

Cuando estos sustantivos están en plural, o cuando se interpone otra palabra entre el artículo y el sustantivo, se emplean los artículos femeninos.

Las águilas son aves

Las aguas negras alcanzaron la población
Juan compró unas hachas en el mercado

Existen sustantivos que nombran lo mismo con ambos géneros (sustantivos ambiguos), según ciertos usos o estilos.

el mar = la mar
el radio = la radio

Cuando el artículo *el* va precedido de las preposiciones *de* y *a*, se producen las formas *del* y *al* por contracción. (*de + el* y *a + el*). Si el artículo forma parte de un nombre propio no se lleva a cabo la contracción.

La casa del electricista
Voy a ir al parque
Voy a El Salvador
Acaba de regresar de El Cairo

1.4.2.2.7 La preposición y su papel

Las preposiciones son palabras invariables que sirven para relacionar términos [11].

El cantor va sobre la tierra
Fuertes cazadores armados de hierros
Su pecho tenía por coraza

Los términos relacionados pueden ser cualquier clase de palabras: sustantivo, adjetivo, verbo, adverbio.

Las preposiciones se pueden clasificar en simples y en locuciones prepositivas. Por ejemplo, simples: *a*, *ante*, *bajo*, *cabe*, *con*, *contra*, etc.

Las locuciones prepositivas permiten matizar o precisar lo que se enuncia. Éstas pueden formarse por:

1) Adverbio y preposición

antes de
después de
dentro de
cerca de
junto a
alrededor de

2) Preposición, un sustantivo y otra preposición

con arreglo a
con base en
de acuerdo con
por medio de
en virtud de
en nombre de

3) Dos o más preposiciones juntas

hasta por
por entre
en contra de
hasta en

Las preposiciones cumplen una función relacionante. El significado de las preposiciones se precisa en el contexto. En el Anexo I se proporciona de manera detallada las preposiciones para español.

1.4.2.3 La concordancia

Uno de los medios gramaticales de relación interna es la concordancia, es decir, la igualdad de género y número entre el sustantivo y el adjetivo, y la igualdad de número y persona entre un verbo y su sujeto.

En español, por el hecho de estar contenido el sujeto en la desinencia del verbo, la concordancia de verbo y sujeto rige únicamente para los casos de determinación y desarrollo del sujeto fuera del verbo que lo contiene.

Hay reglas que rigen la concordancia de modo constante, pero son expresiones de relaciones psíquicas a las cuales tratan de ajustarse. La relación entre el sustantivo y el adjetivo que lo califica, o entre el verbo y su sujeto, supone de parte del hablante un análisis del sujeto y sustantivo en cada caso. Pero la expresión gramatical producida puede no coincidir con la concordancia mental (*concordancia semántica*), es decir, a la relación que se genera en nuestra mente al relacionar los componentes de dicha expresión; puede obedecer a indecisiones en el análisis interno, o a deficiencia o indecisiones en la expresión o falta de pericia del hablante.

En otros casos la discordancia es deliberada para conseguir determinados efectos estilísticos. Lo normal es la concordancia semántica y la gramaticalmente expresada.

La rapidez improvisadora del habla coloquial favorece la aparición de discordancias, sobre todo en los niños y en personas poco instruidas.

Las reglas de la concordancia gramatical de forma general y los casos especiales en que se presentan son las siguientes [12].

I. Cuando el verbo se refiere a un solo sujeto, concuerda con él en número y persona, y cuando el adjetivo se refiere a un solo sustantivo, concuerda con él en género y número. Por ejemplo.

<i>El <u>niño</u> durmió largo rato</i>	(sustantivo – verbo)
<i>Los <u>niños</u> durmieron largo rato</i>	(sustantivo plural – verbo plural)
<i>El <u>caballo blanco</u> es bello</i>	(sustantivo – adjetivo)
<i>Los <u>caballos blancos</u> están atados</i>	(sustantivo plural – adjetivo plural)

Los casos especiales podemos agruparlos en los siguientes:

a) Sexo y género gramatical. Ocurre cuando hay discrepancia entre el sexo de las personas y el género gramatical o del sustantivo con el que se les designa. Estos son los títulos y tratamientos como *usted*, *Santidad*, *alteza*, *majestad*, etc.; van acordes con el adjetivo masculino o femenino según el sexo de la persona a la que se aplica. Por ejemplo, *usted es muy bondadoso*, *usted es muy bondadosa*. La aposición y las denominaciones familiares, cariñosas o irónicas como *vida mía*, *corazón de mis ojos*, etc., no impiden la concordancia con el sexo de la persona a quien se aplica.

b) Colectivos. Los colectivos como *gente*, *muchedumbre*, *número*, *multitud*, *pueblo*, *vecindario*, etc., debido a la pluralidad que encierran, cuando están en singular pueden ajustarse con un adjetivo o un verbo en plural, por ejemplo, *la multitud*, *al escuchar el disparo*, *se dispersaron*. Frecuentemente, se tiene indecisión entre la concordancia gramatical y la semántica. Pero hay circunstancias que favorecen a una u otra [12]. La indeterminación y heterogeneidad de los individuos que entran en la denominación colectiva favorece la idea de pluralidad (*gente*, *multitud*, *muchedumbre*), de ahí la concordancia semántica. Por ejemplo, *acudieron a la ciudad multitud de gente*. Por el contrario, la homogeneidad o la determinación de los individuos reafirma el carácter unitario del colectivo y la concordancia gramatical. De este modo, resulta fastidioso decir *El enjambre con la humareda se dispersan* (adecuado: *El enjambre con la humareda se dispersa*), o *El rebaño, con la sequía y la falta de pastos, perecían* (adecuado: *El rebaño, con la sequía y la falta de pastos, perecía*). Sin embargo, al sustituir los colectivos de estos ejemplos por *gente*, *muchedumbre*, etc., la concordancia en plural parecerá admisible. Por ejemplo, *La gente, con la sequía y la falta de alimentos, perecían*

Los sustantivos *mitad, tercio, parte, resto* y otros semejantes aplicados a un conjunto de individuos pueden ajustarse en singular o en plural. Por ejemplo, *el barco se hundió en altamar, la mitad se salvaron (o se salvó); guardaban (o guardaba) el paso una multitud de hombres armados*. Si el colectivo es acompañado por adjetivos o frases complementarias que refuercen su singularidad gramatical, la concordancia en plural es difícil o imposible. Por ejemplo, *Aquel grupo, entre todos los estudiantes, se había distinguido especialmente*.

La concordancia en plural de los colectivos singulares es favorecida por la distancia a que se encuentran el verbo del adjetivo con que deben concordar. Si las palabras que los dividen son muchas, la posibilidad de concordancia en plural aumenta. Por ejemplo, *El público, después de tan ruidosa propaganda en periódicos y carteles, se agolpaban en las taquillas del teatro*.

c) Discordancia deliberada. Algunas veces nos dirigimos a un sujeto singular con un verbo en plural para dar un efecto estilístico deliberado, o bien para hacerse participe en la actividad o estado del interlocutor. Por ejemplo, al preguntarle a un enfermo: *¿Cómo estamos?, ¿Qué tal vamos?*, etc.

II. Cuando el verbo se refiere a varios sujetos debe ir en plural. Si concurren varias personas verbales distintas, la segunda es preferida a la tercera y la primera a todas. Cuando el adjetivo se refiere a varios sustantivos, se expresa en plural. Si los sustantivos son de diferente género se prefiere el masculino.

Por ejemplo.

Juan, tú y yo viajaremos juntos

Ella y él venían muy cansados

Los casos especiales podemos agruparlos en los siguientes:

a) Pluralidad gramatical y sentido unitario. De la misma manera que un sustantivo colectivo puede ser analizado en sus elementos que lo componen, varios sustantivos asociados pueden ser analizados como un todo y tomarlo como un colectivo singular. Por ejemplo, *La entrada y salida de vapores ha sido aplazada*. Podemos disociar los sustantivos anteponiendo a cada uno el artículo o un demostrativo, para obtener la concordancia en plural casi siempre obligatoria, *la entrada y la salida de vapores han sido aplazadas*.

b) Los verbos en infinitivo, cuando desempeñan el papel de sustantivo, pueden reunirse como sujetos de un verbo en singular, en especial si no van acompañados de un artículo, de lo contrario provocaría la disociación y la concordancia en plural predominaría. Por ejemplo, *comer, beber y no hacer nada, le arruinó en poco tiempo (o le arruinaron)*.

Para dos o más pronombres demostrativos neutros, la concordancia se da en singular. Por ejemplo, *Todo esto y algo más motivó el desinterés*.

c) Posición del verbo respecto a los sujetos. Si el verbo es posterior a los sujetos, la pluralidad es enteramente visible y próxima y es raro que el verbo se ponga en singular. Por ejemplo, *la abuela y el nieto caminaban despacio*. Sin embargo, cuando el verbo precede a varios sujetos aumenta la posibilidad de que concuerde sólo con el primero, por ejemplo, *Le vendrá el señorío y la gracia como de molde*.

d) Posición de adjetivos respecto a los sustantivos. Cuando el adjetivo va después de dos o más sustantivos, concuerda con ellos en plural. Por ejemplo, *Claridad y erudición admirables*. Esta es la concordancia más general; sin embargo, aparecen escasamente casos con el adjetivo en singular. Por ejemplo, *audacia y valor indomable*, versus *indomables*. Otro caso depende del grado de cohesión de los sustantivos, por ejemplo, *Lengua y literatura españolas* versus *Lengua y literatura española*; en el primero presupone aislamiento de los dos sustantivos, y en el segundo se considera en su conjunto unitario que da pauta a ser calificado en singular.

Si el adjetivo precede a los sustantivos, concuerda generalmente con el más próximo. Por ejemplo, *El público lo recibió con entusiasta admiración y aplauso*. En este caso hay complicaciones semánticas si antepone o pospone un adjetivo al sustantivo, por ejemplo, *mujer pobre* versus *pobre mujer*, la semántica es distinta, en la primera expresa la pobreza de una mujer y en la segunda la desdicha de una mujer. Estos problemas de semántica sólo se presentan con algunos adjetivos.

Capítulo 2. Antecedentes

En este capítulo se tiene como propósito presentar un panorama general de las colocaciones y el tratamiento que se ha dado en los diccionarios. Además, se demostrarán las diferentes aplicaciones que tienen las colocaciones para resolver algunas tareas de lingüística computacional.

En los últimos años el procesamiento del lenguaje natural ha despertado gran interés. Se han desarrollado muchos diccionarios electrónicos monolingües, bilingües, terminológicos, etc. Pero sin duda, para el aprendizaje de una segunda lengua y para resolver algunas tareas de lingüística computacional, una de las mejores herramientas son los diccionarios denominados combinatorios. En estos diccionarios, se muestra cómo se combinan las palabras, generalmente llamadas colocaciones. Por ejemplo, *'té cargado'*, *'dar a luz'*, *'golpearse la cabeza'*, etc.

Las colocaciones son una característica de todos los lenguajes. Los lexicógrafos y lingüistas han estado interesados en las colocaciones. Las colocaciones muestran las formas frecuentes de cómo se usa una palabra [10] y son unidades multipalabra, las cuales tienen una existencia independiente y se piensa que probablemente deben formar parte de un diccionario [6].

Las colocaciones plantean un problema para las personas que tratan de aprender una lengua extranjera. El problema de las colocaciones no es tanto de comprensión sino de producción [33]; se dice que la mayoría de las colocaciones probablemente no son difíciles de entender. Expresiones como *'weak tea'*, *'ask a question'*, de hecho, no son difíciles de entender, *'té ligero'*, *'hacer una pregunta'*, pero es imposible crear la combinación correcta por pura deducción.

Mel'čuk afirma que la diferencia entre un nativo de una lengua y otro no nativo de la lengua es que el primero habla en frasemas (combinaciones de palabras con ciertos matices) y el otro no [6]. El valor expresivo en el diálogo y en la escritura se ve limitado por esta carencia. El conocimiento de las combinaciones de palabras adecuadas para una correcta expresión demuestra su importancia.

2.1 Colocaciones en general

En general, las colocaciones incluyen combinaciones de palabras como *sueño dorado*, *armas de destrucción masiva*, frases verbales como *dormir profundamente*, *dormir como un tronco* y algunas otras como *sano y salvo*, *principio y fin*. Normalmente, las combinaciones tienen un significado especializado o son frases idiomáticas como *caballos de fuerza*, *fuerza de poder* (frases técnicas) y *dar a luz*, *marchar sobre ruedas* (‘parir un bebe’ y ‘algo que se está desarrollando de una forma excelente’, ambas idiomáticas). Otras combinaciones no tienen nada de idiomático, sólo son adjetivos modificando a un sustantivo, adverbios modificando verbos o complementos de un verbo: *leer un libro*, *hombre bueno*, *vida sana*. Pero el uso muy frecuente de estas frases como expresiones fijas acompañadas de ciertas connotaciones justifica verlas como colocaciones.

Definimos una colocación como la coocurrencia de dos o más palabras con un espacio corto entre éstas y que aparecen con alta frecuencia en los textos. Estas palabras pueden ser adyacentes o unidas por una palabra funcional, generalmente preposición. Por ejemplo, *dar a luz*, *sueño dorado*, *marchar sobre ruedas*, etc.

Existen principalmente tres enfoques para el estudio de las colocaciones: lexicográfico, estadístico y lingüístico. En cada uno de éstos, el término colocación se usa diferentemente.

El enfoque tradicional ha sido lexicográfico. Aquí los diccionarios proporcionan información que es impredecible o idiomática (modismo). EL BBI [1] sintetiza los estudios de Hausmann sobre colocaciones. Definen las colocaciones como combinaciones de palabras recurrentes fijas o semi-fijas (‘recurrent semi-fixed combinations’), pero dentro del diccionario, además de las combinaciones lexicalizadas (*cellular phone*, *digital clock*), existen también combinaciones libres (*to nominate a candidate*, *to acces data*).

En términos de Hausmann [32] una colocación está compuesta por dos elementos: una **base** (‘base’) y un **colocado** (‘collocate’). La base es semánticamente autónoma, mientras que los colocados no pueden interpretarse de forma aislada, es decir, que el conjunto de colocados léxicos con los que se combina una base dada no pueden predecirse.

En los últimos años, se han aplicado métodos estadísticos para el estudio del lenguaje natural. Sinclair [26] expone que una palabra que ocurre con alta frecuencia próxima a la palabra de investigación es nombrada **colocado** de dicha palabra. Esta combinación de palabras es considerada como colocación.

Choueka [28] define una colocación como una secuencia de dos o más palabras que tienen características de unidad sintáctica y semántica, la que tiene un significado exacto o no puede ser derivado directamente de los significados o connotaciones de sus componentes.

El fenómeno de colocación, para Mel'čuk [4], lo define como una combinación léxica de dos elementos: un componente autónomo y otro dependiente, cada uno de los cuales son denominados como *palabra clave* ('key word') y *correlacionado léxico* ('lexical correlate'), respectivamente.

Una propuesta específica de cómo tratar las colocaciones en un modelo lingüístico fue desarrollado por Mel'čuk [4] con la introducción de las funciones léxicas.

2.2 Características de las colocaciones

Para el tratamiento lingüístico de las colocaciones se han considerado los siguientes criterios [9].

- **No composicionalidad de las colocaciones**

El significado de la colocación no es la composición de los significados de sus componentes, ni el significado es completamente diferente de la combinación libre (por ejemplo, *medias tintas*), o existe una connotación o un elemento agregado del significado que no puede ser predicho por sus partes. Por ejemplo, *fuerte como un toro*, *rojo como el jitomate* [23].

- **No sustitutividad de las colocaciones**

Los componentes de la colocación no pueden sustituirse por otras palabras como sus sinónimos cercanos. Por ejemplo, *vino blanco*, *blanco* no puede sustituirse por *claro* o *amarillo* intentando formar *vino claro*.

- **No modificabilidad de las colocaciones**

Muchas colocaciones no admiten modificaciones con unidades léxicas adicionales, o a través de transformaciones gramaticales, especialmente las expresiones idiomáticas. Por ejemplo, '*picar como los pájaros*' no podemos agregar *negros* ('*picar como los pájaros negros*'), aunque *negros* modifica a *pájaros*.

Generalmente, las colocaciones no se pueden traducir palabra por palabra de un idioma a otro (especialmente las frases idiomáticas), ni obtener su significado completo sumando los significados de sus componentes.

2.3 Clasificación de las colocaciones

Lexicógrafos y lingüistas presentan una amplia variedad de colocaciones individuales, intentan categorizarlas como parte de un esquema general. El BBI [1] distingue dos tipos de colocaciones, las colocaciones gramaticales y las colocaciones léxicas.

Las colocaciones gramaticales, anotan que son una combinación coocurrente que suele consistir de una palabra base (verbo, sustantivo, adjetivo) seguida de una palabra gramatical (colocado); generalmente, contienen una preposición, tales como *verb + preposition* (*come to, put on*), *adjective + preposition* (*afraid that, fond of*), y *noun + preposition* (*by accident, witness to*).

Las colocaciones léxicas son pares de palabras restringidas léxicamente, en las cuales solamente el subconjunto de sinónimos del colocador se pueden usar en el mismo contexto léxico. Por ejemplo, para colocaciones que consisten de verbos y sustantivo (*verb + noun / pronoun*), la mayoría de estas colocaciones consisten de un verbo que denota creación o activación (*come to agreement, make an impression, compose music, set a record*).

En trabajos computacionales son normalmente incluidos como colocaciones los nombres propios y las expresiones terminológicas que hacen referencia a conceptos y objetos en el dominio técnico.

2.4 Tratamiento de las colocaciones en los diccionarios

La información colocacional en diccionarios impresos no es fácil presentarla. Las colocaciones están compuestas por dos elementos, la base y el colocado. El BBI [1], organiza la estructura con las bases y lista los colocados en el cuerpo de la entrada. El BBI muestra varios cientos de combinaciones de palabras. Por ejemplo, algunas entradas del BBI se muestran en la Figura 2.1. Las relaciones sintácticas en estas colocaciones no se especifica explícitamente.

access *n.* 1. to gain, get ~ 2. to deny ~ 3. direct; easy, free, unlimited; limited ~ 4. (computer) random ~ 5. ~ to (we gained/got ~ to the files; ~ to a building)

affirm *v.* 1. to ~ categorically 2. (L) the ministry ~ed that the visit had been postponed

agitate *v.* 1. to ~ strongly 2. (D; intr.) to ~ against; for (they were ~ing for reform)

ago *adv.* 1. ~ that + clause (it was five years ~ that we met)

certain *adj.* 1. absolutely, completely, totally; almost, nearly; quite; very ~ 2. far from ~ 3.

for ~ 4. ~ about, of (we were ~ of his support) 5. ~ to + inf. (she is ~ to agree) 6. ~that + clause (it is ~ that they will sign the contract; make ~ that all doors are locked; are you ~ that you turned the gas off?)

Nota. La palabra en negritas es la *palabra base*. Ésta se combina con los colocados, o sustituyendo el símbolo (~) en la frase por la *palabra base*, así construyéndose la colocación

Figura 2.1 Ejemplo de colocaciones del BBI

Otro diccionario de inglés que trata las colocaciones es el Oxford Collocations Dictionary (OCD) [27]. Se dice que maneja más de 170,000 colocaciones. Una entrada del diccionario se presenta en la figura 2.2. De la misma manera que el BBI no se aprecian claramente las relaciones sintácticas entre la palabra base y sus colocados.

flow *verb*

- ADV. **easily, freely, smoothly** *We talked, and the conversation flowed freely.* ◊ *Wine and beer flowed freely.* | **fast, swiftly** *The river flows quite fast here.* | **slowly** | **naturally** *In a good production of the play, the action and the words flow naturally.* | **directly** *Some of these changes will flow directly from the legislation.* | **constantly, continuously** | **away, back, in, out, past**
- VERB + FLOW **seem to** | **begin to, start to** | **continue to** *Imported food aid continued to flow in.*
- PREP. **across, along, between, down** *a small stream that flowed down the hillside* **from** *Blood was still flowing from the wound.* **into** *One day seemed to flow into the next.* **out of, over, through** *Information flows continuously through the network.* **to** *to get blood flowing to the brain*

Figura 2.2 Entrada de colocación del OCD

El sistema ARCS[18], en su versión para Internet, ofrece 1,000,000 de colocaciones para inglés. Una de sus entradas se muestra en la figura 2.3.

<p>HEADWORD: general <i>adj.</i></p> <p>VERBS: be</p> <p>NOUNS: ability, acceptance, agreement, anaesthesia, analysis, appearance, appreciation, approach, aspects, assembly, assumption, attempt, authority, awareness, behaviour, belief, case, categories, changes, character, checking, chores, chronology, class of solutions, clause, code, complaints, concentration, concept, conceptions, conclusion, condition, conformity, consensus, considerations, conviction, cult, cultural knowledge, data, debility, decrease, defence, definition, description, desires, deterrent argument, development, difference, direction, discussion, dispositions, divide, doctrine, dollar, shortage, duty, election, elements, encouragement, endangerment, enjoyment, enquiry, enterprise, ethos, evolutionary level, exposition, facilitatory effects, failure, fashion, fate, fear, feature, feeling, forces, form, format, framework, grounds, guardian, guidelines, guide, habit, historical view, hospital, hypothesis, idea, immaterialism, immunity, importance, improvement, increase, integration, intention, interest, interpretation, irregularities, issue, itinerant, knowledge, lack of interest, ladder of offences, language, law, legacy, legal rules, lesson, limits, malaise, mechanisms, medical ward, medicine, method, moral principles, morality, muddle, name, nature;</p> <p>ADVERBS: algebraically, also, both, completely, essentially, however, more, most, only, sufficiently, too, totally, very;</p> <p>EXAMPLES: there is a general acceptance/ agreement/ assumption that it is ...; the same general analysis goes some way toward explaining why ...; the official responses to the riots took a general and specific form; the hope of an early peace had been too general; some further general aspects of cases involving ...; there was no general attempt to ...; in the more general case, however, they are complex; thus a general class of solutions can be obtained by considering ...; it is this more general class of solutions that is considered in this section; it is more important here to underline an important general characteristic which ...; the general clause has converted them into trusts; the Middle Ages developed no general concept of progress; general conformity brought the age of hypocrisy to an end; the general data presented by him are undeniable; the same applies to the general deterrent argument; the Conservative party was clearly losing three general election defeats in a row; there is a feeling that ...; the most general form of this view is that ...;</p> <p>SYNONYMS: common, accepted, prevailing, communal, universal, global, shared; inclusive, comprehensive, overall, unrestricted; ordinary, normal, customary, familiar, usual, everyday, unspecified; assorted, mixed, heterogeneous, diversified, extended, sweeping, catholic, composite, mongrel, combined, panoramic; vague, loose, ill-defined, generalized, undetailed, non-specific; approximate;</p> <p>GERMAN: allgemeine, generell, General-, Voll-, Haupt-, Ober-;</p>

Figura 2.3 Ejemplo de colocación del sistema ARCS

Aunque la información mostrada por el sistema ARCS es bastante, las relaciones sintácticas y el orden en que las palabras deben combinarse no se muestra de forma clara y precisa.

Otro diccionario combinatorio, para ruso, el sistema *CrossLexica* [2], en su forma electrónica, cuenta con varios cientos de miles de combinaciones. Su base de colocaciones es accesible para procesadores y para usuarios. Se manejan dos idiomas: el ruso y el inglés. En la figura 2.4 se muestra el sistema y el ejemplo de una entrada. Se presentan todas las pala-

bras con las cuales se puede combinar la palabra base y se clasifican por rubros los diferentes papeles sintácticos.

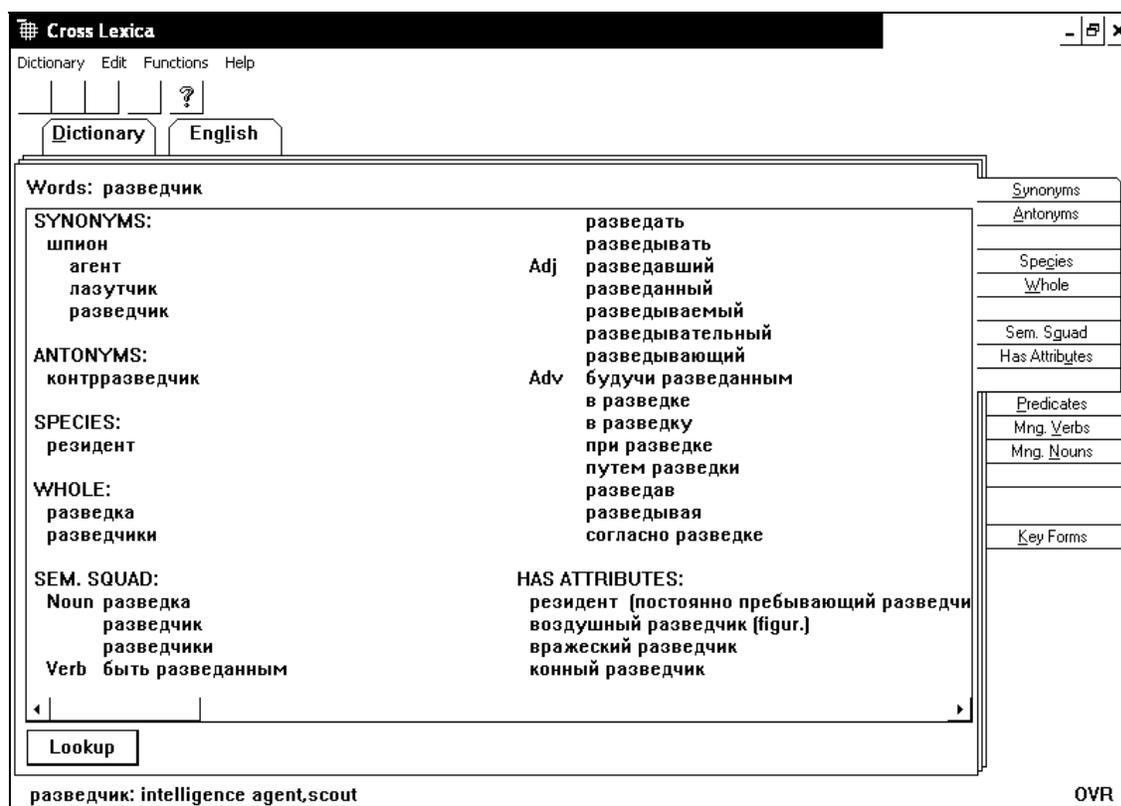


Figura 2.4 Ejemplo de colocación del sistema CrossLexica versión rusa

La manera en que presenta las combinaciones de palabras el sistema CrossLexica parece muy adecuada, además de presentar claramente los enlaces sintácticos y semánticos entre ellas.

2.4.1 Las colocaciones en el Diccionario Explicativo Combinatorio (DEC)

Mel'čuk afirma que un nativo de una lengua habla en frasemas (combinaciones de palabras) [6]. Esto apunta que no es suficiente el conocer el vocabulario de la lengua, ni las reglas gramaticales para hablar un idioma, sino que es necesario conocer también las expresiones fijas o frasemas de la lengua.

Mel'čuk cataloga los frasemas en cuatro grupos principales [6]:

1. **Frasemas pragmáticos o pragmatemas.**

Su forma y significado son claros, pero está fija en relación con una situación dada. Por ejemplo, en el discurso frases como, *¡es para ti!* , *¿qué tal?*.

2. **Frasemas completos o expresiones idiomáticas.**

Expresiones completamente fijas. El significado no es la suma de los significados de cada uno de sus constituyentes. Por ejemplo, *dar a luz* (parir la mujer).

3. **Semi-frasemas o colocación.**

Expresiones semi-fijas. El significado de la combinación adquiere el significado directo de uno de los constituyentes y el significado lateral del otro. Por ejemplo, *té cargado*, *poner atención*. De hecho, *cargado* es asociado primariamente con *peso*.

4. **Casi-frasemas o expresiones casi-idiomáticas.**

Expresiones casi-fijas. El significado de la combinación adquiere el significado de todos los constituyentes, pero que tiene un significado adicional. Por ejemplo, *dar el pecho [a X]*, ‘dar de mamar a un bebe X poniendo un pecho al alcance de su boca’.

Mel’čuk afirma que los pragmatemas, los frasemas completos y los casi-frasemas deben formar parte del diccionario de igual forma que las palabras habituales. Cada frasema debe tener su entrada léxica y tener la misma estructura que las demás entradas.

En el enfoque de la Teoría Significado \Leftrightarrow Texto (Meaning \Leftrightarrow Text Theory, MTT) el concepto de colocación se refiere a las construcciones semi-idiomáticas formadas por dos unidades léxicas L1 y L2, en donde L2 es seleccionada de manera parcialmente arbitraria por L1 para expresar un sentido dado y un papel sintáctico en función de L1. Sería el caso de combinaciones como *amor ardiente*, *promesa solemne*, *silencio absoluto*, *ruido infernal*, en donde no podríamos intercambiar los adjetivos que expresan el mismo sentido ‘intenso’ o ‘en alto grado’: ‘*promesa *infernal*’, ‘*silencio *ardiente*’.

Mel’čuk afirma que las colocaciones deben tener un tratamiento especial debido a su imprevisibilidad, es decir, que la idea que representa la combinación de las unidades léxicas L1 y L2, se determina solamente por la relación de los componentes, L2 no es de elección libre está restringida léxicamente por L1. Por ejemplo, *fuerte como un toro* y *valiente como un gallo*. No pueden intercambiarse: *un gallo* por *un toro*, y obtener una combinación nue-

* Indica que es una combinación no válida para un nativo de la lengua.

va como ‘*fuerte como un gallo’, la cual no expresaría el sentido esperado (para un nativo de la lengua) como en las primeras combinaciones.

Los elementos de una colocación no pueden ser invertidos; nada en la semántica y en la sintaxis de los lexemas obliga a hacer esta selección. Las expresiones son imprevisibles y deben aprenderse.

En la MTT, una colocación es, por tanto, un sintagma constituido por dos unidades léxicas, la base y el colocado, tales que la elección de la base está sólo controlada semánticamente, mientras que la elección del colocado está también controlada léxicamente por la base.

Mel’čuk [5] hace una propuesta específica para el tratamiento de las colocaciones en un modelo lingüístico, las que son descritas por medio de las funciones léxicas.

En esta teoría, se codifica el conocimiento léxico en una entrada del Diccionario Explicativo Combinatorio (DEC). Cada entrada se divide en tres zonas: la zona semántica (una red semántica que representa el significado de la entrada en términos de las palabras más primitivas); la zona sintáctica (las propiedades gramaticales de la entrada) y la zona combinatoria léxica (contiene los valores de las funciones léxicas). Las funciones léxicas son el estudio central de las colocaciones. Un ejemplo de entrada del Diccionario Explicativo Combinatorio para francés se muestra en la Figura 2.5 [6].

La organización de la información de un artículo del DEC gira en torno a la unidad léxica. Se considera lexema a aquella palabra precisada con un único sentido o acepción.

Como puede observarse en la Figura 2.5, el título introduce los rasgos morfológicos. Posteriormente, sigue la sección semántica o definición. Consta de una forma proposicional y de la definición propiamente dicha. La forma proposicional hace explícitos los actuantes semánticos de un lexema L, que se designan por medio de variables (X, Y, Z...).

Después sigue la zona sintáctica, incluye un *esquema de régimen* donde se especifican para cada uno de los actuantes semánticos del lema (X, Y, Z) sus actuantes correspondientes en el nivel sintáctico profundo (X=1, Y=2, Z=3).

Finalmente, la zona combinatoria léxica contiene las llamadas *funciones léxicas*, que son sentidos abstractos y generales (del tipo **Magn** (‘intensificador), **Oper** (‘realización’), etc.) que se añaden al lexema considerado dando como resultado una expresión parcialmente fija, llamada semifrasema o colocación, característica de una lengua concreta. Las funciones léxicas permiten así clasificar de manera sistemática todo un conjunto de colocaciones, diferentes de las llamadas expresiones idiomáticas.

REPROCHER 'reprochar', verbo, transitivo

la.

Definición.
 X reproche Y à Z = X étant convaincu que Y, dontal croit Z responsable, a eu ou a lieu, Y ou les conséquences de Y affectant négativement l'état émotionnel de X, ||
 X a présent dans sa conscience que X n'excuse pas Z pour Y, et aussi X croit que Z doit
2.1 être conscient du caractère relativement mauvais de Y.

Régimen (*government pattern*)

X=1	Y=2	Z=3
1. N	1. N 2. <i>de</i> V _{inf}	1. à N
	obligatorio	obligatorio

1) C₂₂ : M₁(M₂) = M₃
 ['M' significa actante sintáctico profundo: la fórmula quiere decir que el primer actante del infinitivo es el mismo que el tercer actante de **REPROCHER**]

2) C₁ + C₂ + C₃ : *je reproche à Marie sa nonchalance <le fait que je sois de mauvaise humeur = ma mauvaise humeur> Il nous reproche ses déboires; Marie lui reproche d'être indiscret; Il m'a longtemps reproché de ne pas lui avoir dit la vérité <de lui rendre la vie difficile>*

Funciones Léxicas

Syn_v : blâmer₁, désapprouver, réprocher; 'en vouloir₁, 'faire grief₁, 'tenir regueur₁, 'avoir une dent₁, 'en avoir gros sur le coeur₁, **fam** 'en avoir gros sur la patate₁; 'ne pas excuser₁; acuser₃,

Anti_v : excuser₃

Anti_v : apprécier, approuver; pardonner

S_o : reproche **la**

AntiAble_{v2} : irréprochable₁

PredAntiAble_{v2} : 'n'avoir rien à Pron_{v,en} reprocher₁, [*Sur ce plan, je n'avais rien à me reprocher = Sur ce plan, j'étais irréprochable*]

Magn['affecter'] : amèrement

Magn['ne pas excuser'] : beaucoup < fortement < énormément

Verb : 'à juste titre₁, avec raison

AntiVerb : 'à tort₁, injustement

Ejemplos

1. Ma mère m'a longtemps reproché de m'être marié
2. Marie lui reproche la noirceur de son actino
3. Marie reproche à Pierre d'avoir agi inconsciemment
4. Un récent sondage indiquait que les Montréalais reprochaient aux policiers leurs abus de pouvoir (76%), leur brutalité (64%) et leur racisme (53%)
5. Il nous reproche sa défaite
6. Sans jamais s'en ouvrir à Jacques, sa mère lui a toujours reproché amèrement ses échecs scolaires.

Figura 2.5 Ejemplo de entrada del DEC (francés)

2.4.2 Las funciones léxicas en el DEC

Las funciones léxicas se formalizan en el concepto de función matemática que se define de la siguiente forma [6]:

“Una función léxica (FL) denota una relación entre una unidad léxica de la lengua (L), llamada *argumento* de la función, y un conjunto de unidades léxicas (inclusive vacío) o expre

siones idiomáticas $\{v_1, v_2, \dots, v_n\}$, llamados *valores* de la FL aplicada a L". Se representa como:

$$FL(L) = \{v_1, v_2, \dots, v_n\}$$

Toda relación léxica puede ser examinada y modelada por la función léxica específica, tomada del conjunto de funciones léxicas estándares identificadas por la Teoría Significado \Leftrightarrow Texto.

Por ejemplo, algunas funciones léxicas se presentan a continuación [6]:

Syn, sinónimo exacto y **Syn \supset** , **Syn \subset** , **Syn \cap** casi-sinónimos (sentido más específico, menos específico e intersección de los significados).

Syn(avión) = aparato

Syn \supset (pueblo) = sociedad

Syn \subset (veneración) = respeto

Syn \cap (escritorio) = mesa

Anti, enlaza a su antónimo.

Anti(bajo) = alto

Anti(rápido) = lento

Magn, como intensificador.

Magn (réplica) = fulminante

Magn (memoria) = prodigiosa, excelente, asombrosa

Magn (ruido) = infernal, de todos los demonios

Magn (dormir) = profundamente, como un tronco, a pierna suelta.

Bon, bueno, expresión que se emplea como una alabanza

Bon(consejo) = valioso

Bon(eloio) = esmerado, bien formulado

Bon(tiempo) = buen

Oper, toma su palabra clave como complemento directo.

Oper (denuncia) = presentar [ART ~].

Oper (grito) = dar, pegar [ART ~].

Donde ART, indica que un determinante (artículo, adjetivo posesivo, o demostrativo, etc.) debe ser usado según las reglas de la gramática del español. La tilde ‘~’ indica repetición del argumento.

Para Mel’čuk el diccionario es un sistema de relaciones léxicas donde cada relación se ha de especificar. Mel’čuk define unas cincuenta funciones léxicas estándares [5].

De los diccionarios antes presentados, el sistema CrossLexica muestra de una forma adecuada las relaciones entre las combinaciones de palabras. Éste muestra ordenadamente los enlaces sintácticos y semánticos, en comparación con el OCD y el BBI que no muestran de forma explícita los enlaces entre las combinaciones de palabras ni de manera precisa las combinaciones. En contraste, el DEC asume que el destinatario del diccionario debe usar su lógica para realizar la interpretación adecuada. Además, el DEC está muy restringido en ejemplos de uso.

El presente trabajo pretende valerse de las ideas básicas del sistema CrossLexica versión rusa, y desarrollar las herramientas morfológicas para español [17] y el sistema que demuestra las colocaciones y sus papeles sintácticos y semánticos entre las palabras.

2.5 Aplicaciones de las colocaciones

Las colocaciones son muy importantes, principalmente cuando uno quiere aprender una segunda lengua [24]. Existen fórmulas en la lengua (las colocaciones) que se usan en diferentes contextos orales y escritos. Conociendo las colocaciones de la lengua, además del vocabulario y la gramática, uno podrá aprender a expresarse como un hablante nativo de la lengua.

No sólo las colocaciones son útiles para el aprendizaje de una segunda lengua, sino también para resolver algunas tareas de lingüística computacional. Las colocaciones pueden usarse para la desambiguación incluyendo la desambiguación de sentidos de palabras y la desambiguación estructural [25]. La tarea se basa en el principio: que una palabra tiende a coocurrir con un conjunto diferente para cada uno de sus sentidos. Por ejemplo, si tenemos la siguiente palabra con sus dos acepciones: *banco*₁ (‘asiento para una sola persona sin respaldo’) y *banco*₂ (‘organismo público de crédito’). En las siguientes combinaciones no se tiene la certeza a qué significado se refiere la combinación.

1. *banco de crédito*
2. *banco de madera*

En realidad, *de madera* es frecuente encontrarlo como modificador del sustantivo con el primer sentido (*banco*₁) y *de crédito* al segundo sentido (*banco*₂). Esto lo sabemos por medio del conjunto posible de palabras combinables.

Otra aplicación es en el análisis sintáctico, en la desambiguación de palabras. Por ejemplo, tenemos la siguiente combinación de palabras, *puerta de madera de la cocina*. En esta frase podemos tener las siguientes asociaciones:

1. [[*puerta de madera*] *de la cocina*]
2. [*puerta de* [*madera de la cocina*]]

Es más frecuente que se encuentre la primera asociación, con el sentido semántico: ‘una puerta que es de madera y que pertenece a la cocina’. En contraste, con la segunda asociación de sentido incoherente.

Otra tarea importante es la traducción automática, debido a que las colocaciones no se pueden caracterizar con base en las regularidades sintácticas y semánticas. Éstas no pueden ser traducidas palabra por palabra. Sin embargo, la lingüística computacional utiliza técnicas estadísticas en corpus para identificar las colocaciones y construir semi-automáticamente léxicos colocacionales bilingües. Estos pueden usarse como parte de un sistema de generación de lenguaje [9].

Otra de las tareas donde las colocaciones son de ayuda es en la detección y corrección de errores semánticos llamados malapropismos [22]. Por ejemplo.

1. *las orugas de la ropa* (en lugar de: *las arrugas de la ropa*)
2. *irrigación de la piel* (en lugar de: *irritación de la piel*)

Las combinaciones más frecuentes son *arrugas de la ropa* e *irritación de la piel*. Se puede dar la sugerencia para modificar los posibles errores.

Otras aplicaciones de las colocaciones son en la generación de lenguaje natural, recuperación de información [9] y coherencia en diálogos hombre-máquina [25].

Para la creación de las bases de colocaciones existen dos formas. Primera y la más confiable es recopilarlas de forma manual, lo que requiere de un enorme esfuerzo y tiempo. Con esta forma tenemos la certeza que las colocaciones reunidas son ciento por ciento confiables. La segunda es recopilarlas de forma automática, haciendo uso de los corpus existentes para español, uno de los más accesibles ‘Internet’ u otros como en [16 y 21].

Las colocaciones se pueden obtener por medios estadísticos. La forma más sencilla es la frecuencia de ocurrencia de las combinaciones de palabras. Se usan patrones de filtrado por

donde se pasan las frases candidatas a ser colocaciones [34]. Se pueden obtener bigramas (combinaciones de dos palabras) o n-gramas (combinaciones de más de dos palabras) según los patrones de filtrado. Otros métodos pueden implementarse, como el de la media y la varianza, prueba de chi-cuadrada de Pearson (X^2) e información mutua [9].

Las colocaciones compiladas en las bases primarias para nuestro sistema son bigramas, es decir, combinaciones de dos palabras, aunque a veces se encuentran unidas por una palabra funcional, generalmente preposición.

Capítulo 3. Desarrollo del sistema *CrossLexica-Esp*

En este capítulo se presentan el sistema *CrossLexica* ruso del cual se toman las ideas básicas para la construcción del sistema *CrossLexica* para español. Presentamos la estructura de las bases de colocaciones con enlaces sintácticos y semánticos para español, y la estructura de las bases formateadas, las que tienen los rasgos morfológicos y semánticos de cada componente de la colocación.

El sistema *CrossLexica* ruso es un sistema que maneja una base grande de combinaciones de palabras (colocaciones) y enlaces semánticos entre ellas (véase figura 2.4 del capítulo 2). Las colocaciones tienen dependencias gramaticales entre sus elementos. Los sinónimos, antónimos, subclases, superclases, etc. representan los enlaces semánticos en forma de tesoro. La estructura del sistema es universal y puede aplicarse a cualquier lenguaje (ruso, inglés, español, etc) [2].

El sistema *CrossLexica* para Español (*CrossLexica-Esp*) toma como base las ideas del sistema *CrossLexica* ruso. Por lo que el sistema *CrossLexica-Esp* es un diccionario especial de español que incluye enlaces sintácticos entre palabras como *país grande*, *ver al joven*, *muy bien*, *trabajar fuertemente*, etc. y semánticos como *bebida – refresco*, *país – México*, etc. Permite el acceso a usuarios (por medio de una interfaz especial) y a procesadores de texto (por medio de enlaces a combinaciones de palabras). Además, el sistema proporciona información idiomática entre las combinaciones de palabras, información para diferenciar homónimos e información acerca del estilo de uso de la combinación en el texto.

3.1 Estructura del sistema *CrossLexica-Esp*

El desarrollo de la aplicación *CrossLexica-Esp* consiste en dos tareas principales:

- Las utilerías para la preparación de la Base de Datos (conjunto de combinaciones de palabras en forma electrónica).
- La interfaz del usuario.

La interfaz del usuario para el sistema *CrossLexica-Esp* tiene la finalidad de mostrar la información sintáctica y semántica de la palabra clave introducida como petición. El sistema proporciona al usuario todos los enlaces entre la palabra clave y colocados (palabras rela-

cionadas) estructurados por rubros semánticos (hipónimos e hiperónimos) y sintácticos (modificadores que tiene una palabra, verbos que maneja una palabra y sustantivos que maneja una palabra).

La petición debe ser una palabra que forme parte del diccionario de *CrossLexica-Esp*, puede tener desinencia, pues la interfaz tiene la capacidad de transformar la petición a su forma estándar, para adjetivos la forma en singular-masculino y para verbos la forma infinitiva, y buscarla en la lista de palabras del diccionario.

Otra característica de la interfaz, además de mostrar los enlaces sintácticos y semánticos estructurados por rubros, es que muestra la información concordante entre las combinaciones de palabras y se proporcionan los artículos y preposiciones correctos que enlazan las combinaciones de palabras.

Una característica más del sistema es la inferencia de nuevas colocaciones con base a las contenidas en el sistema y haciendo uso de la inferencia lógica.

Herramientas utilitarias

El conjunto de las herramientas utilitarias del sistema *CrossLexica-Esp* es utilizado, al fin y al cabo, para la preparación de la información del diccionario que manejará el sistema. En sus etapas secuenciales, estas herramientas darán el formato y caracterización morfológica de las palabras que representan las colocaciones (combinación de palabras estables) y que son preparadas en la vista de *archivos primarios* (archivos especiales que contienen las colocaciones).

Cada archivo tiene una estructura especial que depende de los enlaces sintácticos que puede contener. Por ejemplo, la palabra *ataque* puede ser combinada con las palabras: *de nervios*, *falso*, *frontal*. Estas palabras modifican a *ataque*, se obtienen las combinaciones *ataque de nervios*, *ataque falso*, *ataque frontal*. La palabra principal la denominamos *palabra clave* y las complementarias *palabras relacionadas*. Tal como en las funciones léxicas propuestas por Mel'čuk [5], la *palabra clave* es el argumento y los posibles valores de la función son las *palabras relacionadas*.

A cada grupo de palabras, *palabra clave* y sus *palabras relacionadas*, le denominamos *artículo del diccionario*. Los artículos están contenidos en las bases primarias (archivos) de colocaciones (las que se describen posteriormente); así cada archivo es contenedor de un tipo de enlace sintáctico.

Las bases primarias de colocaciones se han reunido de forma manual. Existen otras formas de obtener las colocaciones por medios estadísticos [9], haciendo uso de corpus disponibles para español [16 y 21].

En la Figura 3.1 se muestra el esquema general del sistema *CrossLexica-Esp*. Las Herramientas Utilitarias (Figura 3.1, Parte I) dan formato y obtienen los rasgos morfológicos de las palabras que representan las colocaciones y que están preparadas en los archivos con forma original llamados *archivos primarios*. Todas las palabras clave son compiladas en un diccionario general y los enlaces (sintácticos y semánticos) entre palabras que representan las colocaciones son compilados para obtener una base de enlaces entre palabras.

Cada *archivo primario* (Subparte I.2) contiene un tipo especial de colocaciones, con restricciones específicas para cada uno. Estos archivos son legibles para la persona que prepara las colocaciones y las modifica.

Los archivos de forma codificada llamados *archivos secundarios* (Subparte I.4), contienen las colocaciones formateadas y la información morfológica de cada palabra constituyente. Estos son obtenidos por el proceso de formateo especial de colocaciones. De forma paralela, el proceso de formateo utiliza archivos auxiliares (Subparte I.1) y genera un protocolo de errores del formateo para el control del proceso (Subparte I.5).

Los archivos auxiliares contienen listas de excepciones que determinan casos raros, tales como adjetivos que únicamente se usan de forma plural, géneros de palabras que no siguen las reglas generales, adverbios, preposiciones, etc. Estos archivos consisten en información adicional para la caracterización adecuada de los componentes de las colocaciones.

Para fines de facilitar el proceso de formateo, se crea un protocolo de formateo. El protocolo informa al compilador los errores que ocurrieron durante el proceso y el tipo de errores producidos. Por medio de este protocolo podemos corregir las colocaciones caracterizadas erróneamente y compilarlas nuevamente. Este protocolo es el medio por el cual se garantiza que las colocaciones sean compiladas correctamente.

Para cada entrada del diccionario del sistema *CrossLexica-Esp*, se encuentra en su forma estándar (adjetivos y verbos); los sustantivos se distinguen entre la forma singular y la forma plural. En las combinaciones, en ocasiones, las palabras relacionadas sólo se usan con la forma singular y en ocasiones sólo con la forma plural del sustantivo. El sistema realiza la concordancia en género y número de forma automática en el momento de mostrar las colocaciones.

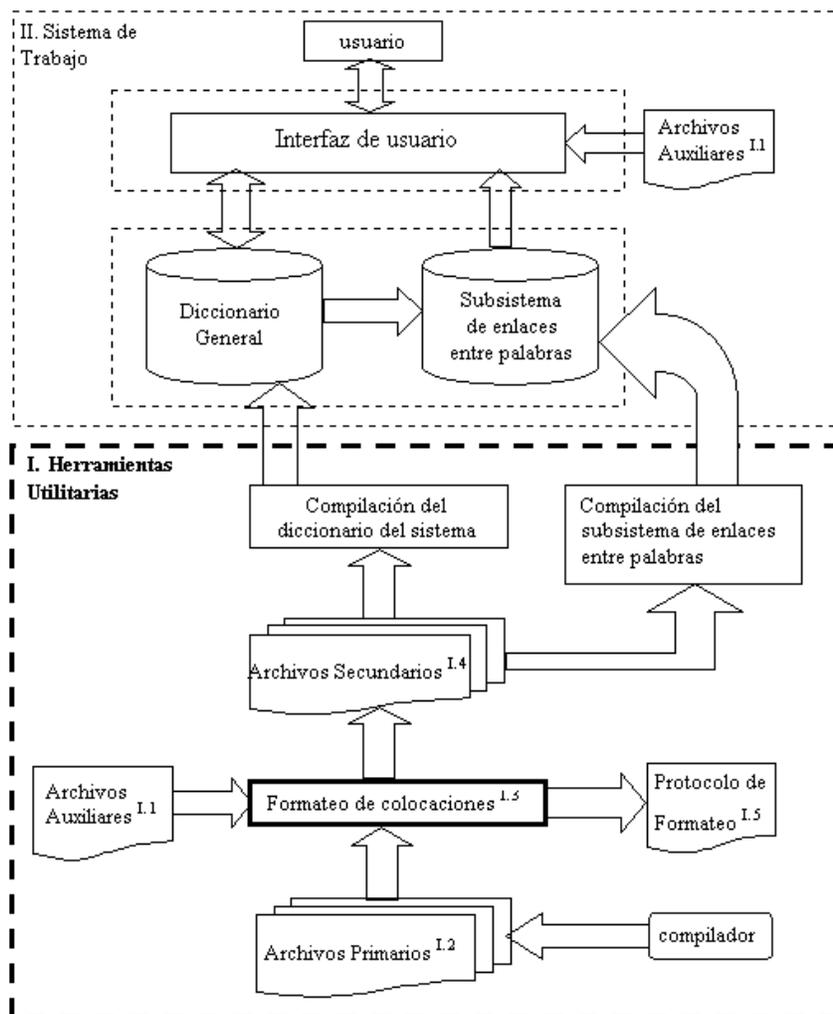


Figura 3.1 Esquema de bloques del sistema *CrossLexica-Esp*

En los siguientes puntos mostraremos la estructura de las bases primarias y secundarias de las colocaciones que conforman el sistema.

3.2 Bases primarias de colocaciones en *CrossLexica-Esp*

En esta etapa hay que precisar la noción de colocación. Como ya se mencionó, algunos autores la definen como una coocurrencia de palabras en un espacio corto entre ellas y que aparecen con alta frecuencia entre los textos. En correspondencia con el punto de vista de Mel'čuk [5], nosotros consideramos como colocaciones los siguientes grupos de palabras vinculadas sintácticamente:

- Expresiones idiomáticas o frases completos como *medias tintas* ‘con reservas’ o *campo santo* ‘cementerio’. El significado de la combinación tiene un sentido fijo, que no incluye ninguno de los sentidos de los constituyentes.
- Expresiones semi-idiomáticas o tales con funciones léxicas estándar [5] como *té cargado* o *poner atención*. El significado de la combinación adquiere el significado directo de uno de los constituyentes y el significado lateral del otro.
- Combinaciones libres como *acción buena* o *ver al joven*. El sentido de tal combinación es la suma del sentido de cada constituyente.

Entonces, nuestro término de colocación no corresponde a una mera coocurrencia de palabras dentro de un espacio corto de texto. Esto implica una dependencia sintáctica entre las combinaciones de palabras, de manera inmediata, o a través de una palabra funcional, usualmente preposición, por ejemplo, *ver* (\rightarrow *al*) \rightarrow *joven*, *agarrar* (\rightarrow *del*) \rightarrow *brazo*, etc.

Establecido lo anterior, las colocaciones se encuentran en las bases o archivos primarios. Cada uno de los archivos primarios contienen sólo un tipo de colocación y por lo tanto las dependencias sintácticas son predeterminadas para cada archivo.

Como se mencionó, los archivos están constituidos por *artículos del diccionario* (una *palabra clave* y múltiples *palabras relacionadas*). Las palabras pueden ser alguna de las partes de la oración (POS del Inglés, *Part Of Speech*).

La *palabra clave* es única en el archivo corriente y no es de elección libre, sino depende del archivo en el cual se trabaje. Las *palabras relacionadas* pueden ser múltiples. De forma recíproca las palabras relacionadas dependen del archivo en el cual se trabaje. Por ejemplo, en la figura 3.2 se muestra un artículo de una base primaria.

<p><i>material</i> <i>didáctico</i> <i>eléctrico</i> <i>gráfico</i> <i>incandescente</i></p>

Figura 3.2 Ejemplo de artículo de una base primaria

El la figura 3.2, *material*, se considera como la palabra clave y las demás como palabras relacionadas, es decir, las palabras con que se puede combinar *material*: *material didáctico*, *material eléctrico*, etc.

Las bases primarias de colocaciones con enlaces sintácticos se agrupan en los siguientes archivos:

- Modificadores
- Valencias verbales
- Valencias sustantivas

Las bases con enlaces semánticos se agrupan en:

- Hipónimos / Hiperónimos

3.2.1 Archivo de modificadores

Para este archivo, se considera como palabra clave al rector de la relación entre las partes de la oración (sustantivos, adjetivos, verbos y adverbios).

Este archivo tiene dos modos de uso, primero: los modificadores que puede tener la palabra clave proporcionada y segundo: qué palabras rectoras puede tener el modificador proporcionado.

Los enlaces sintácticos permitidos se muestran en la Figura 3.3; están definidos por las flechas [5].

El archivo de modificadores contiene los enlaces sintácticos: *tiene_modificadores* y su recíproco *es_modificador*. Por ejemplo, para el sustantivo *acción*, sus modificadores en la relación *tiene_modificadores* son: *buena*, *combinada*, *directa*, *prometida*. Generando las combinaciones: *acción* → *buena*, *acción* → *combinada*, *acción* → *directa*, *acción* → *prometida*. Para el adjetivo *agradecido* tiene modificadores *muy*, *poco*, *sumamente*. Obteniéndose las combinaciones *muy* ← *agradecido*, *poco* ← *agradecido*, *sumamente* ← *agradecido*.

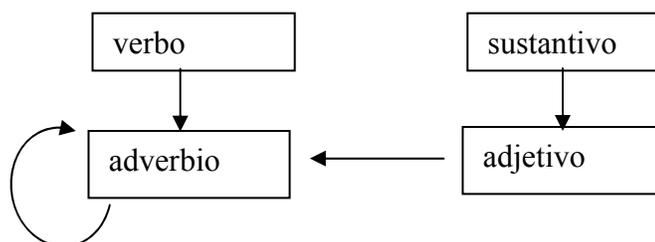


Figura 3.3 Enlaces sintácticos para modificadores

En el modo de uso inverso, se consulta la palabra: adjetivo o adverbio, se obtienen las combinaciones para las cuales éstos son modificadores. Por ejemplo, el adjetivo *bueno* modifica en la relación *es_modificador* varios sustantivos: *acción*, *apariencia*, *calidad*. Se obtienen las combinaciones *acción* → *bueno*, *apariencia* → *bueno*, *calidad* → *bueno*. Para español, la concordancia entre el sustantivo y el adjetivo, usualmente, el adjetivo concuerda con la forma morfológica de su sustantivo que lo acompaña.

La estructura para este archivo se detalla en el punto 3.4 *Estructura de las bases de colocaciones modificadoras*.

3.2.2 Archivo de valencias verbales

Para este archivo, el sustantivo dependiente se considera como la palabra clave. Tiene dos modos de uso, primero: qué verbos rigen la palabra clave, y segundo: qué complementos puede manejar el verbo. En la Figura 3.4 se observa el tipo de enlace sintáctico que admite este archivo.



Figura 3.4 Enlaces sintácticos para valencias verbales

Por ejemplo, para el sustantivo *atención* los verbos que lo rigen son: *atraer*, *dedicar*, *distraer*, *poner*. Con ello se obtienen las combinaciones: *atraer* → (*la* ←) *atención*, *dedicar* → *atención*, *distraer* → (*la* ←) *atención* y *poner* → *atención*.

Las relaciones sintácticas para este archivo son las siguientes: *tiene_complementos* , *tiene_sujetos* y sus recíprocos *es_complemento_de_verbo*, y *es_sujeto_para*.

Por ejemplo, para el verbo *abrir* y para la relación *tiene_complementos*. Se tienen las siguientes combinaciones.

abrir la boca
abrir una maleta
abrir una carta

Los complementos de *abrir* son *la boca*, *una maleta* y *una carta*.

En el otro modo, *la boca* es complemento para distintos verbos. Para la relación *es_complemento_de_verbo*.

abrir la boca
cerrar la boca
taparse la boca

Los verbos que manejan el complemento *la boca* son *abrir*, *cerrar* y *taparse*.

Para los verbos permitimos el uso del pronombre enclítico *se*, como en ‘*taparse la boca*’, ‘*abrirse paso*’, ‘*comerse una fruta*’, etc.

También se permite la construcción de oraciones pasivas reflejas con el pronombre *se* y con el verbo en voz activa. Véase el punto 1.4.2.2.3 Uso del pronombre *se*. Por ejemplo, ‘*se firmó la paz*’, ‘*se aceptó el acuerdo*’, etc.

Las oraciones pasivas reflejas hacen uso de las relaciones sintácticas *tiene_sujetos* y *es_sujeto_para*. Por ejemplo, para la oración *la maleta se abrió*, la relación *tiene_sujetos* se aplica al verbo *abrir* y el sujeto que tiene es *la maleta*. De forma inversa, la relación *es_sujeto_para* se aplica al sujeto *maleta* y el verbo para el cual es sujeto es *abrió* (*abrir*).

Para este tipo de oraciones con las relaciones sintácticas *tiene_sujetos* y *es_sujeto_para* se permite la conjugación de verbos. Los tiempos permitidos son los tiempos simples: presente, pasado y futuro, y las conjugaciones para la tercera persona del singular y plural.

La estructura de este archivo se detalla en el punto 3.5 *Estructura de las bases de colocaciones con valencias verbales*.

3.2.3 Archivo de valencias sustantivales

Para este archivo, el sustantivo dependiente (el segundo en el texto) se considera como la palabra clave. En la Figura 3.5 se observa el tipo de enlace sintáctico que admite este archivo.

Las relaciones sintácticas para este archivo son las siguientes: *tiene_complementos* y su recíproco *es_complemento_de_sustantivo*.

Por ejemplo, el sustantivo *problemas* y para la relación *tiene_complementos*. Se tienen las siguientes combinaciones:

problemas del país
problemas del cerebro
problemas de la maquinaria

Los sustantivos que complementan a *problema* son *país*, *cerebro* y *maquinaria*.

Para la relación *es_complemento_de_sustantivo*. Se tienen las siguientes combinaciones:

ciudades del país
problemas del país
ciudadanos del país

El sustantivo *país* es complemento para *ciudades*, *problemas* y *ciudadanos*.

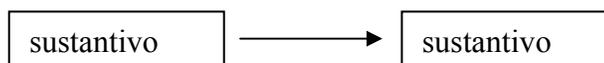


Figura 3.5 Enlaces sintácticos para valencias sustantivales

La estructura para este archivo se detalla en el punto 3.6 *Estructura de las bases de colocaciones con valencias sustantivales*.

3.2.4 Archivo de hipónimos / hiperónimos

Para este archivo que refleja las relaciones semánticas, se considera como la palabra clave los conceptos hiperónimos y como palabras relacionadas a sus hipónimos, es decir, la palabra clave es la superclase y las relacionadas son las subclases. Se clasifican las nociones de las palabras en forma jerárquica, los que son únicamente sustantivos.

Las relaciones semánticas para este archivo son las siguientes: **hiperónimo** y su recíproco **hipónimo**. Por ejemplo, la noción general de *anillos*, *aretes*, *collares* y *pendientes* es *adornos* que se representa de la forma siguiente:

adornos
anillos
aretes
collares
pendientes

Las relaciones entre *adornos* y *anillos* son: *adornos* es **hiperónimo** de *anillos* y *anillos* es **hipónimo** de *adornos*.

La estructura para este archivo se detalla en el punto 3.7 *Estructura de las bases de hipónimos / hiperónimos*.

3.3 Estructura de las bases primarias en general

Las bases primarias o archivos primarios son aquellos que contienen las colocaciones; cada archivo restringe el tipo de colocaciones que permite. Los siguientes rasgos morfológicos y semánticos se permiten para todos los tipos de colocaciones.

3.3.1 Palabra clave

Un artículo del diccionario está constituido por una palabra clave y al menos una o múltiples palabras relacionadas. La palabra clave define el artículo dentro del diccionario de colocaciones, por lo que no debe haber duplicidad de la palabra que se considera como palabra clave en el archivo.

La palabra clave puede contener los siguientes marcos morfológicos y semánticos, pero únicamente se puede usar uno a la vez que se coloca inmediatamente después de la palabra clave. Los valores para los marcos se muestran en la Tabla 3.1. Su uso se describe en los siguientes puntos.

3.3.1.1 Información semántica

Es un marco para indicar con palabras en una forma libre, pero concisa, información para distinguir las palabras clave que son homónimas o proporcionar la clase a la que pertenece la palabra corriente.

- **Información para homónimos**

En este modo de uso se proporciona información al usuario para distinguir las palabras homónimas. Este texto (de forma libre) debe ser escrito entre comillas ("*texto libre*") y de manera concisa. El marco comienza inmediatamente después de la palabra clave. El marco se define con el símbolo < (menor que) seguido de la información del homónimo. Véase Tabla 3.1. Por ejemplo, si se quiere distinguir las diferentes áreas en que se usa la palabra *tono*, se definiría como en el ejemplo siguiente.

```

tono1<"de música"
    alto
    bajo
    buen
tono2<"de afecto"
    cariñoso
    claro

```

poco seguro
tono3<"del cuerpo"
cardiaco
vital

Se procede de esta manera para incorporar la información de las palabras homónimas que serán de utilidad para ayuda del usuario. Además, para cada una de las acepciones, se etiqueta con un número de forma consecutiva.

- **Clase semántica**

En este modo de uso se define la clase superior de la palabra corriente. Se distingue de la *información para homónimos* por no llevar comillas, pero al igual que ésta es de forma libre y concisa. Normalmente es una palabra que generaliza el concepto de la subclase, la palabra corriente.

Por ejemplo, se tiene la palabra *agente* y se quiere especificar la clase a la que pertenece cada una de las acepciones, se debe realizar de la siguiente forma:

<i>agente1<hombre</i>	<i>agente2<sustancia</i>
<i>británico</i>	<i>dañina</i>
<i>de bolsa</i>	<i>de limpieza</i>

Con este marco se proporcionan dos nociones, primera: la clase a la que pertenece cada acepción (en este caso, a un *hombre* y a una *sustancia* respectivamente), segunda: hace referencia al humano y referencia a un tipo de líquido, según cada caso. Este marco debe usarse cuando existen varias acepciones para la palabra clave y el sentido de la palabra no es del todo claro, además de proporcionar la clasificación de la palabra corriente.

3.3.1.2 Parámetros morfológicos

Es un marco que se utiliza para eliminar la ambigüedad que se pueda presentar entre una palabra clave y alguna otra que se encuentre en el diccionario. Los marcos posibles son los siguientes:

- **Parte de la oración (POS)**. Es un marco que define la parte de la oración a la que corresponde (sustantivo, adjetivo, verbo o adverbio. Véase Tabla 3.1).

De forma general, los verbos se determinan por tener la terminación *ar*, *er* o *ir*. Por ejemplo, *poder* (facultad de hacer algo, verbo) y *poder* (tener la habilidad de hacer algo, sustantivo). Para este caso es necesario colocar el marco que lo distingue.

<i>poder1</i> < SUS	<i>poder2</i>
<i>económico</i>	<i>inmediatamente</i>
<i>estatal</i>	<i>rápidamente</i>
<i>gran</i>	
<i>Ejemplo1</i>	<i>Ejemplo2</i>

De igual manera sucede con palabras como *deber*, *escolar*, *mar*, *militar*, etc. Este marco debe usarse con las palabras que son sustantivos y tienen la terminación de verbo (*ar*, *er* o *ir*) o por alguna otra razón que cause conflicto al caracterizar la colocación.

- **Género.** Es un marco que define el género de la palabra (masculino o femenino. Véase Tabla 3.1). Este marco elimina la ambigüedad de palabras que tienen la misma forma para ambos géneros.

Por ejemplo, la palabra *analista*, en el ejemplo de abajo, la primera palabra es de género masculino y en la segunda femenino. Se necesita distinguir con el marco ambos géneros, para indicar que se usa la misma forma de palabra para ambos géneros.

<i>analista1</i> < MASC	<i>analista2</i> < FEM
<i>financiero</i>	<i>financiera</i>
<i>político</i>	<i>política</i>

Ejemplo3

- **Número.** Es un marco que define el número de la palabra (plural o singular. Véase Tabla 3.1). Este marco elimina la ambigüedad de palabras que tienen la misma forma para ambos números.

Por ejemplo, la palabra *déficit* en los ejemplos 4 y 5 presenta la misma forma para el singular y plural. Es necesario distinguir con el marco ambos números, los que usan la misma forma de la palabra.

<i>déficit1</i> < SING	<i>déficit2</i> < PLUR
<i>bancario</i>	<i>bancarios</i>
<i>hormonal</i>	<i>hormonales</i>
<i>Ejemplo4</i>	<i>Ejemplo5</i>

Tabla 3.1 Marcos para palabras clave

MARCO	Descripción	
<i>Parte de la oración (POS)</i>		
<SUS	Sustantivo	
<ADJ	Adjetivo	
<VER	Verbo	
<ADV	Adverbio	
<i>Número</i>		
<SING	Singular	
<PLUR	Plural	
<i>Género</i>		
<MASC	Masculino	
<FEM	Femenino	
<i>Información semántica</i>		
<"texto libre"	Información para homónimos	Texto libre
<Clase semántica		Superclase

3.3.2 Palabras relacionadas

Las palabras relacionadas son asociadas a la palabra clave. Se distinguen de la palabras clave por tener un símbolo de tabulación (→) al inicio de la palabra, el cual es obligatorio. Por ejemplo.

dedo

- *anular*
- *auricular*
- *delgado*
- *fino*

Adicionalmente, se cuenta con marcos opcionales de *idiomatización* y *estilo*. Éstos son exclusivos de las palabras relacionadas para indicar el uso de la combinación de palabras. Los marcos posibles son los siguientes.

3.3.2.1 Idiomatización

Este marco proporciona información acerca de la combinación corriente, es decir, si su uso es idiomático, semi-idiomático o neutral (véase Tabla 3.2).

<i>árbol</i>	<i>luz</i>	<i>cabeza</i>
<i>& de navidad</i>	<i>brillante</i>	<i>blanca</i>
<i>& genealógico</i>	<i>solar</i>	<i>& de chorlito</i>
	<i>@ verde</i>	
<i>Ejemplo6</i>	<i>Ejemplo7</i>	<i>Ejemplo8</i>

En los ejemplos anteriores, el símbolo (&) indica que la combinación es fija y tiene un significado idiomático (ejemplos 6 y 8). En el ejemplo 6, la combinación *árbol genealógico*, el sentido es ‘una tabla que indica la filiación de los miembros de una familia bajo forma de árbol’ y no tiene el sentido de ‘una planta’. Sin embargo, en el Ejemplo7, la combinación *luz verde* adquiere dos sentidos (denotada por el símbolo @), ‘tener autorización de proseguir con algún trabajo’ y ‘color de la luz’. Tal combinación se usa dependiendo del contexto, por ejemplo, primer sentido: ‘*El jefe dio luz verde al proyecto*’ y segundo sentido: ‘*la luz verde me cegó por un instante*’.

3.3.2.2 Estilo de la combinación

Este marco indica los contextos en que se usa normalmente la combinación. La Tabla 3.2 proporciona los valores para este marco.

Por ejemplo, el símbolo *S* indica que la combinación se usa en textos científicos y técnicos, el símbolo *C* que su uso es coloquial (comúnmente se usa en conversaciones amigables) y el símbolo *V* que su uso es popular o vulgar.

<i>caballos</i>	<i>andar</i>	<i>ramas</i>
<i>S de fuerza</i>	<i>despacio</i>	<i>apartar las ~</i>
<i>finos</i>	<i>&V como gallina clueca</i>	<i>&C andarse por las ~</i>
<i>puros</i>	<i>&V como perro sin mecate</i>	
<i>Ejemplo9</i>	<i>Ejemplo10</i>	<i>Ejemplo11</i>

La tilde ‘~’, indica la posición en la cual la palabra clave se debe repetir.

Tabla 3.2 Marcos para palabras relacionadas

MARCO	Descripción
Idiomatización	
&	Completamente idiomática, expresiones propias del idioma español
@	Se puede usar de forma idiomáticamente o de forma literal

MARCO	Descripción
Estilo de la combinación	
S	Combinación de uso científico, se usa en documentos científicos y técnicos.
C	Combinación de uso coloquial, se usa en conversaciones amigables, familiares.
V	Combinación de uso vulgar, se usa para matizar expresiones de uso popular
I	Combinación de uso indecente, son expresiones ofensivas.

La ausencia de marco indica que su uso es neutral.

Los marcos anteriores (para palabras clave y para palabras relacionadas) se aplican para todos los tipos de colocaciones del sistema CrossLexica-Esp.

3.4 Estructura de las bases de colocaciones modificadoras

Las *colocaciones modificadoras* se encuentran en la base primaria en una forma burda. Después del proceso de caracterización se obtiene una forma esquematizada, una base secundaria. En los siguientes puntos se describen ambas bases.

3.4.1 Base primaria

El archivo primario de *colocaciones modificadoras* contiene las colocaciones con las relaciones sintácticas del tipo *sustantivo* → *adjetivo* y [*adjetivo, verbo, adverbio*] → *adverbio*. En X → Y, la flecha indica la dependencia sintáctica como se mencionó [4]. Los adjetivos dependen del sustantivo y los adverbios dependen del adjetivo, verbo, o adverbio.

Las palabras clave permitidas para este archivo son sustantivos, adjetivos, verbos y adverbios y las palabras relacionadas son adjetivos, adverbios o términos multipalabra.

Los términos multipalabra, es decir, construcciones adjetivales (CONS_ADJ) y construcciones adverbiales (CONS_ADV) se consideran como una entrada del diccionario del sistema. Por ejemplo, las construcciones adjetivales que se tienen en las colocaciones *puerta de madera*, *aceite de almendras*, y las construcciones adverbiales en: *dormir a pierna suelta*, *saltar de alegría*, son entradas del diccionario general.

Una construcción adjetival puede tomar también el papel de construcción adverbial. Por ejemplo, en las combinaciones *chispa de alegría* y *saltar de alegría*. La construcción *de alegría* es a la vez CONS_ADJ y CONS_ADV relativamente. Es una forma de homonimia, tendrá dos entradas en el diccionario del sistema, una como construcción adjetiva y otra como construcción adverbial.

El esquema general de un artículo del archivo primario de modificadores es el de la figura 3.6.

PALABRA_CLAVE [< Información semántica] [<Parámetro morfológico]
[idiomatización][estilo] TAB <u>Palabra relacionada1</u> [~]
[idiomatización][estilo] TAB <u>Palabra relacionada2</u> [~]
[idiomatización][estilo] TAB <u>Palabra relacionada3</u> [~]
.
[idiomatización][estilo] TAB <u>Palabra relacionadaN</u>

Figura 3.6 Esquema de un artículo de colocaciones modificadoras

Las palabras en negritas son obligatorias para el artículo del diccionario.

La tilde ‘~’, indica la posición en la cual la palabra clave se ha de repetir.

Palabra clave. Es obligatoria para el artículo.

Palabra relacionada. Define el modificador de la palabra clave.

TAB. Es una tabulación al inicio de la palabra relacionada.

Los corchetes indican que su contenido es opcional. Los marcos de la palabra clave se usan de forma exclusiva. Para el uso de estos marcos véase 3.3.1.2 *Parámetros morfológicos*.

Para los adjetivos que se usan de forma antepuesta es necesario colocar la tilde (~) delante del adjetivo. Esta acción indica que la palabra clave debe sustituirse en esa posición logrando la anteposición al adjetivo. La ausencia de este símbolo indica posposición del adjetivo. Por ejemplo:

<i>noticia</i>	<i>soldado</i>
<i>cierta ~</i>	<i>simple ~</i>
<i>veraz</i>	<i>valiente</i>

En los ejemplos, las combinaciones se leen *cierta noticia*, *noticia veraz*.

En la tabla de abajo se muestran algunos artículos del archivo primario de colocaciones modificadoras.

<i>bien1</i> < <i>ADV</i>
<i>muy</i>
<i>bien2</i> < <i>SUS</i>
<i>social</i>
<i>bienes</i>
<i>inmuebles</i>
<i>muebles</i>
<i>bienestar</i> < <i>SUS</i>
<i>general</i>
<i>bloque</i>
<i>de concreto</i>
<i>temático</i>
<i>bloques</i>
<i>acabados</i>
<i>de piedra</i>
<i>blusa</i>
<i>blanca</i>
<i>de trabajo</i>
<i>rota</i>
<i>causa</i>
<i>común</i>
<i>justa</i>
<i>justificada</i>
<i>principal</i>
<i>comunidades</i>
<i>aisladas</i>
<i>autónomas</i>
<i>civilizadas</i>
<i>rurales</i>
<i>concentración</i>
<i>desproporcionada</i>

Una muestra del archivo de colocaciones modificadoras en su etapa inicial (archivo primario), se presenta en el *Anexo A.1 Base primaria de colocaciones modificadoras*.

3.4.2 Base secundaria

Cada renglón del archivo determina una palabra independiente. Los códigos son definidos del tamaño de un byte para los rasgos morfológicos y estilísticos (columnas 1 a la 8 de la Tabla 3.3) y posteriormente sigue la cadena de letras que representa la palabra seguida de su posible información semántica.

La posición de códigos en un renglón es la siguiente.

Tabla 3.3 Formato de la base secundaria de modificadores

Columna	Descripción
1	Papel que desempeña la palabra en la colocación (palabra clave versus relacionada)
2	Parte de la oración
3	Género
4	Número
5	Clase de inclinación
6	Posición del adjetivo
7	Marco de idiomatización
8	Marco de estilo
9	Palabra corriente con posibles marcos de homonimia y clase semántica

3.4.2.1 Descripción de las columnas

1. Papel que desempeña la palabra en el artículo del diccionario

Tabla 3.4 Papel de la palabra corriente

Valor	Descripción
+	Define la palabra clave de este artículo.
-	Define la palabra relacionada del artículo.

2. Parte de la oración

Tabla 3.5 Valores de la parte de la oración

Valor	Descripción
1	Sustantivo
2	Adjetivo
3	Verbo
4	Adverbio

3. Género

Sólo se aplica para sustantivos. Para adjetivos, verbos y adverbios el género tiene valor cero. El género del adjetivo no se refleja en el diccionario.

Tabla 3.6 Valores para género

Valor	Descripción
0	Invariable
1	Masculino
2	Femenino

4. Número

Sólo se aplica para sustantivos. Para verbos, adverbios y términos multipalabra el número tiene valor cero. El número del adjetivo no se refleja en el diccionario.

Tabla 3.7 Valores para número

Valor	Descripción
0	Invariable
1	Singular
2	Plural

5. Clase de inclinación para sustantivos, adjetivos y verbos

La clase de inclinación son las posibles desinencias para la palabra corriente. La clase de inclinación se determina de acuerdo con las desinencias de las tablas siguientes:

- **Sustantivos.** Para sustantivos se tienen las siguientes desinencias y se proporcionan algunos ejemplos.

Tabla 3.8 Desinencias para la clase del sustantivo

Clase	Desinencias		Ejemplos
	Singular	Plural	
0	0	0	<i>lunes<u>0</u> / lunes<u>0</u></i>
1	0	<i>s</i>	<i>acto<u>0</u> / acto<u>s</u></i>
2	0	<i>es</i>	<i>mineral<u>0</u> / mineral<u>es</u></i>
3	<i>z</i>	<i>ces</i>	<i>lapi<u>z</u> / lápice<u>s</u></i>

- **Adjetivos.** Para adjetivos se tienen las siguientes desinencias y se proporcionan algunos ejemplos.

Tabla 3.9 Desinencias para la clase del adjetivo

Clase	Desinencias				Ejemplos
	Singular Masc.	Singular Fem.	Plural Masc.	Plural Fem.	
0	0	0	0	0	<i>de violencia<u>0</u> *</i> <i>isósceles<u>0</u></i>
1	o	a	os	as	<i>bueno<u>o</u> / buena<u>a</u> / buenos<u>os</u> / buenas<u>as</u></i>
2	0	0	es	es	<i>elemental<u>0</u> / elemental<u>0</u> / elementales<u>s</u> / elementales<u>s</u></i>
3	0	0	s	s	<i>amable / amable / amables<u>s</u> / amables<u>s</u></i>
4	e	a	es	as	<i>regordete / regordeta / regordetes<u>s</u> / regordetas<u>s</u></i>
5	0	a	es	as	<i>invasor<u>0</u> / invasora<u>a</u> / invasores<u>s</u> / invasoras<u>as</u></i>
6	z	z	ces	ces	<i>locomotri<u>z</u> / locomotri<u>z</u> / locomotrices<u>es</u> / locomotrices<u>es</u></i>
7	1	a	os	as	<i>primero<u>e</u> / primera<u>a</u> / primeros<u>os</u> / primeras<u>as</u> **</i>
8	2	2	s	s	<i>grande<u>de</u> / grande<u>de</u> / grandes<u>s</u> / grandes<u>s</u> ***</i>

El cero indica la existencia de un sufijo vacío.

* Para la clase 0, el número cero en la desinencia indica que no se debe alterar la palabra (es invariable), tal como se indica en el ejemplo.

** Para la clase 7, el número uno en la desinencia *singular masculino* indica que se debe eliminar la última letra para este tipo de desinencia, tal como se indica en el ejemplo.

*** Para la clase 8, el número dos en la desinencia *singular masculino* y *singular femenino* indica que se deben eliminar las últimas dos letras para ambos tipos de desinencia, como

se indica en el ejemplo; esta clase es única para el apócope *gran*, el cual se transforma en el diccionario a *grande*, el cual indica que se debe colocar previo al sustantivo que lo maneje y eliminando la desinencia adecuada.

- **Verbos.** Para verbos se tienen las siguientes desinencias y se proporcionan algunos ejemplos.

Tabla 3.10 Desinencias para la clase del verbo

Tipo	Desinencias infinitivo	Ejemplos
1	ar	<i>entr<u>ar</u>, tomar<u>ar</u>, saltar<u>ar</u></i>
2	er	<i>corr<u>er</u>, comer<u>er</u>, ver<u>er</u></i>
3	ir	<i>abr<u>ir</u>, abat<u>ir</u>, salir<u>ir</u></i>
4	arse	<i>abraz<u>arse</u>, toc<u>arse</u></i>
5	erse	<i>acog<u>erse</u>, enfurec<u>erse</u></i>
6	irse	<i>deprim<u>irse</u>, salir<u>irse</u></i>

6. Posición del adjetivo

Define la posición del adjetivo en la combinación. Ésta se usa para matizar la combinación de palabras.

Tabla 3.11 Posición del adjetivo

Valor	Descripción
0	Posposición
1	Anteposición

7. Marco semántico

Define si la combinación de palabras se considera idiomática, semi-idiomática o neutral. La ausencia de este marco indica que es una combinación neutral. Los valores posibles para esta columna se presentan en la Tabla 3.2.

8. Marco estilístico

Define si la combinación de palabras (*palabra clave* y *palabra relacionada*) se considera especial dependiendo del tipo utilizado. La ausencia de este marco indica que es una combinación normal. Los valores posibles para esta columna se presentan en la Tabla 3.2.

9. Palabra corriente

En esta posición, inicia la palabra que se procesa actualmente seguida de sus posibles marcos de información homónima. Véase *Tabla 3.1 Marcos para palabra clave*.

Una muestra de la base secundaria de colocaciones modificadoras se presenta en el *Anexo B.1 Base secundaria de colocaciones modificadoras*.

3.5 Estructura de las bases de colocaciones con valencias verbales

Las *colocaciones con valencias verbales* se encuentran en la base primaria en una forma burda. Después del proceso de formateo se obtiene una forma esquematizada, una base secundaria. En los siguientes puntos se describen ambas bases.

3.5.1 Base primaria

El archivo primario de valencias verbales contiene las colocaciones con la relación sintáctica del tipo *verbo* → *sustantivo*. Por ejemplo, *llover* (→ *a*) → *cántaros*, *sentir* → *alegría*, *poner* → *atención*, etc. Para este archivo, tomamos como palabras clave los sustantivos dependientes y las palabras relacionadas son los verbos gobernantes.

El esquema general de un artículo del archivo primario de valencias verbales es el de la figura 3.7.

<p>PALABRA_CLAVE [< Información semántica] [<Parámetro morfológico]</p> <p>[idiomatización][estilo]TAB <u>Palabra relacionada1</u> ~ [...]</p> <p>[idiomatización][estilo]TAB <u>Palabra relacionada2</u> [...] ~</p> <p>[idiomatización][estilo]TAB ~ <u>Palabra relacionada3</u></p> <p>.</p> <p>[idiomatización][estilo]TAB <u>Palabra relacionadaN</u> ~</p>

Figura 3.7 Esquema de un artículo de colocaciones con valencias verbales

Las palabras en negritas son obligatorias para el artículo del diccionario.

Palabra clave. Es obligatoria para el artículo y es exclusivamente sustantivo.

Palabra relacionada. Define al verbo rector.

TAB. Es una tabulación al inicio de la palabra relacionada.

La tilde ‘~’, indica la posición en la cual la palabra clave se debe repetir.

Los puntos suspensivos (...), indican omisión de una valencia, la necesaria para dar sentido a la combinación completa.

Los corchetes, indican que es opcional su contenido. Los marcos de la palabra clave se usan de forma exclusiva. Para el uso de estos marcos véase *3.3.1.2 Parámetros morfológicos*.

En la tabla siguiente se muestran algunos artículos del archivo primario de valencias verbales.

<i>agua</i>	<i>diluir ... con el ~</i>
	<i>meterse en el ~</i>
	<i>saltar al ~</i>
<i>aguas</i>	<i>curarse con ~</i>
	<i>tomar las ~</i>
<i>alcohol</i>	<i>diluir ... con ~</i>
<i>alegría</i>	<i>sentir ~</i>
<i>atención</i>	<i>atraer la ~</i>
	<i>dedicar ~</i>
	<i>distraer la ~</i>
	<i>llamar la ~</i>
<i>&</i>	<i>poner ~</i>
<i>&</i>	<i>prestar ~</i>
<i>cántaros</i>	<i>& llover a ~</i>
<i>mesa</i>	<i>& poner la ~</i>
<i>muerto</i>	<i>& cargar con el ~</i>
	<i>sepultar al ~</i>
<i>pestañas</i>	<i>& quemarse las ~</i>

Una muestra de la base primaria de colocaciones con valencias verbales, en su etapa inicial (archivo primario), se presenta en el *Anexo A.2 Base primaria de colocaciones con valencias verbales*.

3.5.1.1 Valencia omitida

La valencia omitida se define por los puntos suspensivos (...). Ésta indica que es necesario un actuante para dar sentido completo a la colocación.

Por ejemplo.

<i>polvo</i>		<i>capa</i>
~ <i>penetra</i> ... (1)		<i>cubrir</i> ... con una ~
<i>reducir</i> ... a ~ (2)		<i>tapar</i> con la ~
<i>Ejemplo1</i>		<i>Ejemplo2</i>

En la combinación, *polvo penetra* ..., el sentido es ‘que el polvo penetra hacia algún sitio’, por tal motivo se requiere de una palabra auxiliar (definida por los puntos suspensivos, ...), para hacer la construcción con una semántica adecuada.

La posición de la valencia omitida se contabiliza de acuerdo al número de palabras anteriores a ésta. Entonces, para la primera combinación el valor de esta columna es tres.

polvo⁽¹⁾ *penetra*⁽²⁾ ...⁽³⁾

En la segunda combinación, *reducir ... a polvo*, el sentido es el siguiente ‘que se reduce alguna cosa a polvo’; por lo expuesto anteriormente, se requiere de una palabra para construir la combinación adecuada. Para esta combinación, el valor de esta columna será dos.

reducir⁽¹⁾ ...⁽²⁾ *a*⁽³⁾ *polvo*⁽⁴⁾.

La valencia omitida se intercala en la combinación según sea requerida.

3.5.1.2 Posición de la palabra clave

La tilde (~) representa la repetición de la palabra clave. Al igual que la valencia omitida, se puede intercalar en la frase según sea requerida. Esta columna define la posición de la palabra clave en la combinación.

Se enumera de acuerdo a la posición en la cual aparece dentro de la colocación.

<i>brazos</i>	<i>corazón</i>	<i>maleta</i>
<i>echarse</i> ... en los ~	~ <i>brincó</i>	~ <i>se abrió</i>
<i>abrir</i> los ~	~ <i>saltó</i>	<i>llevar</i> la ~
<i>cruzar</i> los ~	<i>romper</i> el ~	
<i>Ejemplo3</i>	<i>Ejemplo4</i>	<i>Ejemplo5</i>

Para la combinación, ‘*echarse ... en los brazos*’, Ejemplo3, se lee ‘*echarse algo o alguien en los brazos*’. La primera combinación del Ejemplo4, al sustituir la tilde por la palabra clave *corazón* se encuentra el sujeto de la oración por estar conjugado el verbo: (*el*) *corazón brincó*.

En el Ejemplo5 tenemos la construcción de oraciones de pasiva refleja, (*la*) *maleta se abrió*, se permite solamente el pronombre *se*, los demás pronombres no se permiten en la construcción de las combinaciones.

La posición se enumera de la misma forma que la valencia omitida, obteniéndose el valor de **uno** para la primera combinación del Ejemplo4 y el valor de **cinco** para el Ejemplo3.

corazón⁽¹⁾ *brincó*⁽²⁾
echarse⁽¹⁾ ... ⁽²⁾ *en*⁽³⁾ *los*⁽⁴⁾ *brazos*⁽⁵⁾

3.5.2 Base secundaria

Cada renglón del archivo determina una palabra independiente. Los códigos son definidos del tamaño de un byte para los rasgos morfológicos y estilísticos (columnas 1 a la 13 de la Tabla 3.12) y posteriormente sigue la cadena de letras que representa la palabra.

La posición de códigos en un renglón es la siguiente.

Tabla 3.12 Formato de la base secundaria de valencias verbales

Columna	Descripción
1	Papel que desempeña la palabra en la colocación (palabra clave versus relacionada)
2	Parte de la oración
3	Género
4	Número
5	Clase de inclinación
6	Idiomatización
7	Estilo
8	Tiempo
9	Persona
10	Número de preposición (codificada)
11	Artículo

Columna	Descripción
12	Valencia omitida
13	Posición de la palabra clave en la colocación
14	Palabra corriente

3.5.2.1 Descripción de las columnas

1. Papel que desempeña la palabra en el artículo del diccionario

Los valores posibles para esta columna son los de la Tabla 3.4.

2. Parte de la oración

Los valores posibles para esta columna son los de la Tabla 3.5. Los valores para adjetivo y adverbio no se aplican para este archivo de colocaciones.

3. Género

Sólo se aplica para sustantivos. Para verbos el género es invariable. Los valores posibles para esta columna son los de la Tabla 3.6.

4. Número

Sólo se aplica para sustantivos. Para verbos el número es invariable. Los valores posibles para esta columna son los de la Tabla 3.7.

5. Clase de inclinación para sustantivos y verbos

La clase de inclinación indica las posibles desinencias de la palabra corriente. La clase de inclinación se determina con base a las desinencias de las tablas siguientes:

- **Sustantivos.** Para determinar la clase de inclinación del sustantivo se usa la Tabla 3.8.
- **Verbos.** Para verbos en infinitivo se tienen las siguientes desinencias y se proporcionan algunos ejemplos.

Tabla 3.13 Desinencias para verbos en infinitivo y pronombre *se*

Tipo	Desinencias infinitivo	Ejemplos
1	ar	<i>entr<u>ar</u>, tomar, salt<u>ar</u></i>
2	er	<i>corr<u>er</u>, comer, ver</i>
3	ir	<i>abr<u>ir</u>, abat<u>ir</u>, sal<u>ir</u></i>

Tipo	Desinencias infinitivo	Ejemplos
4	arse	<i>abrazarse, tocarse</i>
5	erse	<i>acogerse, enfurecerse</i>
6	irse	<i>deprimirse, salirse</i>
7	Proclítico <i>se</i> infinitivo -ar	<i>la paz se firmó</i>
8	Proclítico <i>se</i> infinitivo -er	<i>los hombres se enfurecieron</i>
9	Proclítico <i>se</i> infinitivo -ir	<i>la maleta se abrió</i>

6. Marco semántico

Define si la combinación de palabras se considera idiomática, semi-idiomática o neutral. La ausencia de este marco indica que es una combinación normal. Los valores posibles para esta columna se presentan en la Tabla 3.2.

7. Marco estilístico

Define el contexto de posibles usos de la combinación de palabras. La ausencia de este marco indica que es una combinación normal. Los valores posibles para esta columna se presentan en la Tabla 3.2.

8. Tiempo del verbo

Los verbos admitidos para este archivo, deben estar en su forma infinitiva o ser conjugados en tercera persona del singular o tercera persona del plural. Los tiempos considerados sólo son el presente, pasado perfecto y futuro. Esta columna se usa para las combinaciones de palabras en oraciones pasivas reflejas.

Tabla 3.14 Tiempos de conjugación del verbo

Valor	Descripción
0	No existe tiempo de conjugación
1	Tiempo presente
2	Tiempo pasado
3	Tiempo futuro

9. Persona

Indica el número de la persona gramatical que entra en la conjugación del verbo, se permite solamente la tercera persona del singular o tercera persona del plural. Esta columna se usa

para la conjugación de las combinaciones de palabras que forman oraciones pasivas reflexas.

Tabla 3.15 Persona de conjugación

Valor	Descripción
0	No hay persona
1	1ra. Persona singular *
2	2da. Persona singular *
3	3ra. Persona singular
4	1ra. Persona plural *
5	2da. Persona plural *
6	3ra. Persona plural

* No se permiten estas conjugaciones.

10. Preposición

Define el número de preposición que contiene la colocación según la lista de preposiciones del *Anexo I. Preposiciones*.

11. Artículo

Define el número del artículo correspondiente que contiene la palabra relacionada de acuerdo a la siguiente tabla de artículos:

Tabla 3.16 Artículos

Valor	Descripción	Artículos
0	No existe artículo	-
1	Artículo definido	<i>el, la, las, los</i>
2	Artículo indefinido	<i>un, uno, una, unos, unas</i>

12. Valencia Omitida

Define la posición de la valencia omitida dentro de la colocación. El valor cero indica que no existe valencia. Para calcular el valor de esta columna véase el punto 3.5.1.1

13. Posición de la palabra clave

Define la posición de la palabra clave dentro de la colocación. Para calcular esta columna véase el punto 3.5.1.2

14. Palabra corriente

A partir de esta columna comienza la secuencia de caracteres que constituye la palabra. Si existe información semántica, se coloca después de la palabra con un marcador especial. Véase el punto 3.3.1.1.

Una muestra de la base secundaria de colocaciones con valencias verbales se presenta en el *Anexo B.2 Base secundaria de colocaciones con valencias verbales*.

3.6 Estructura de las bases de colocaciones con valencias sustantivales

Las colocaciones con valencias sustantivales se encuentran en la base primaria en una forma burda. Después del proceso de caracterización se obtiene una forma esquematizada, una base secundaria. En los puntos siguientes se describen ambas bases.

3.6.1 Base primaria

El archivo primario de *valencias sustantivales* contiene las colocaciones con la relación sintáctica del tipo *sustantivo* → *sustantivo*. Por ejemplo, *voceros* (→ *del*) → *gobierno*, *nivel* (→ *del*) → *agua*, *resultado* (→ *de*) → (*los* ←) *análisis*. Para este archivo se toma como palabras clave los sustantivos dependientes, es decir, el segundo sustantivo en el texto.

El esquema general de un artículo del archivo primario de valencias sustantivales es el siguiente.

<p>PALABRA_CLAVE [< Información semántica] [<Parámetro morfológico]</p> <p>[idiomatización][estilo]TAB <u>Palabra relacionada1</u></p> <p>[idiomatización][estilo]TAB <u>Palabra relacionada2</u></p> <p>[idiomatización][estilo]TAB <u>Palabra relacionada3</u></p> <p>.</p> <p>[idiomatización][estilo]TAB <u>Palabra relacionadaN</u></p>

Figura 3.8 Esquema de un artículo de colocaciones con valencias sustantivales

Las palabras en negritas son obligatorias para el artículo del diccionario.

Palabra clave. Es obligatoria para el artículo y define al sustantivo dependiente.

Palabra relacionada. Define al sustantivo rector de la palabra clave.

TAB. Es una tabulación al inicio de la palabra relacionada.

Los corchetes indican que su contenido es opcional. Los marcos de la palabra clave se usan de forma exclusiva. Para el uso de estos marcos véase 3.3.1.2 *Parámetros morfológicos*.

En la tabla siguiente se muestran algunos artículos del archivo primario de valencias sustantivales.

<i>ciudadano</i>	<i>deberes del</i>
	<i>garantías del</i>
	<i>obligación del</i>
<i>cliente</i>	<i>obligaciones del</i>
	<i>derechos del</i>
	<i>petición del</i>
<i>minerales</i>	<i>características de los</i>
	<i>detección de los</i>
	<i>descripción de los</i>
<i>parque</i>	<i>árboles del</i>
	<i>flores del</i>
	<i>palomas del</i>
<i>pueblo</i>	<i>canto del</i>
	<i>correo del</i>
	<i>defensor del</i>
	<i>familias del</i>
	<i>habitantes del</i>
	<i>héroes del</i>
	<i>teatro del</i>
	<i>partido del</i>
	<i>voz del</i>
<i>puerta</i>	<i>diseño de la</i>
	<i>manija de la</i>
	<i>marco de la</i>
<i>sustancias</i>	<i>estudio de las</i>
	<i>manejo de las</i>
	<i>nomenclatura de las</i>

<i>paso de las</i>
<i>texto</i>
<i>lectura del</i>
<i>resto del</i>
<i>resumen del</i>
<i>vida</i>
<i>reorden de la</i>
<i>momento de la</i>
<i>plano de la</i>
<i>resto de la</i>

Una muestra de la base primaria de colocaciones con valencias sustantivales en su etapa inicial (archivo primario) se presenta en el *Anexo A.3 Base primaria de valencias sustantivales*.

3.6.2 Base secundaria

Cada renglón del archivo determina una palabra independiente. Los códigos son definidos del tamaño de un byte para los rasgos morfológicos y estilísticos (columnas 1 a la 9 de la Tabla 3.17) y posteriormente sigue la cadena de letras que representa a la palabra.

La posición de códigos en un renglón es la siguiente.

Tabla 3.17 Formato de la base secundaria de valencias sustantivales

Columna	Descripción
1	Papel que desempeña la palabra en la colocación (palabra clave versus relacionada)
2	Parte de la oración
3	Género
4	Número
5	Clase de inclinación
6	Idiomatización
7	Estilo
8	Número de preposición
9	Artículo
10	Palabra

3.6.2.1 Descripción de las columnas

1. Papel que desempeña la palabra en el artículo del diccionario

Los valores posibles para esta columna son los de la Tabla 3.4.

2. Parte de la oración

Los valores posibles para esta columna son los de la Tabla 3.5. Los valores para adjetivo, verbo y adverbio no se aplican para este archivo de colocaciones.

3. Género

Sólo se aplica para sustantivos. Los valores posibles para esta columna son los de la Tabla 3.6.

4. Número

Sólo se aplica para sustantivos. Los valores posibles para esta columna son los de la Tabla 3.7.

5. Clase de inclinación para sustantivos

La clase de inclinación indica las posibles desinencias para la palabra corriente. La clase de inclinación se determina de acuerdo con las desinencias para sustantivos de la Tabla 3.8.

6. Marco semántico

Define si la combinación de palabras se considera idiomática, semi-idiomática o neutral. La ausencia de este marco indica que es una combinación normal, véase el punto 3.3.2.

7. Marco estilístico

Define el contexto de posibles usos de la combinación de palabras. La ausencia de este marco indica que es una combinación normal, véase el punto 3.3.2.

8. Preposición

Define el número de preposición que contiene la colocación de acuerdo con la lista de preposiciones del *Anexo I. Preposiciones*.

9. Artículo

Define el número del artículo correspondiente que contiene la palabra relacionada de acuerdo con la Tabla 3.16

El artículo adecuado para la colocación se elige de acuerdo con la concordancia de su palabra clave (sustantivo) en el momento de mostrar la combinación.

10. Palabra corriente

A partir de esta columna comienza la secuencia de caracteres que constituye la palabra. Si existe información semántica, se coloca después de la palabra con un marcador especial. Véase el punto 3.3.1.1.

Una muestra de la base primaria de colocaciones con valencias sustantivales se muestra en el *Anexo B.3 Base secundaria de colocaciones con valencias sustantivales*.

3.7 Estructura de las bases de hipónimos / hiperónimos

La base primaria de hipónimos e hiperónimos se encuentra en una forma burda. Después del proceso de caracterización se obtiene una forma esquematizada, una base secundaria. A continuación, se describen ambas bases.

3.7.1 Base primaria

El archivo primario de *hipónimos / hiperónimos* contiene la relación semántica **hipónimo** y su recíproco **hiperónimo**. Por ejemplo, *flores* — *rosas*, *país* — *México*, *hombre* — *agente*, *líquido* — *agua*, etc. Para esta relación semántica se refleja la relación de conceptos específicos y conceptos generales: *rosas* ES_UN tipo de *flores* y *México* ES_UN tipo de *país*. La relación que se describe es semántica y se realiza entre conceptos generales y específicos.

Para este archivo, se toman como palabras clave los sustantivos que representan a los hiperónimos o superclase y como relacionadas a los posibles hipónimos o subclases. Los conceptos se encuentran jerarquizados.

El esquema general de un artículo del archivo primario hipónimos / hiperónimos es el siguiente.

PALABRA_CLAVE [< Información semántica] [<Parámetro morfológico]
TAB <u>Palabra hipónima1</u>
TAB <u>Palabra hipónima2</u>
TAB <u>Palabra hipónima3</u>
.
TAB <u>Palabra hipónimaN</u>

Figura 3.9 Esquema de un artículo de hipónimos / hiperónimos

Las palabras en negritas son obligatorias para el artículo del diccionario.

Palabra clave. Es obligatoria para el artículo y representa el concepto general.

Palabra hipónima. Define a la subclase de la palabra clave.

TAB. Es una tabulación al inicio de la palabra relacionada.

Los corchetes, indican que su contenido es opcional. Los marcos de la palabra clave se usan de forma exclusiva. Para el uso de estos marcos véase 3.3.1.2 *Parámetros morfológicos*.

En la siguiente tabla se muestran algunos ejemplos de artículos del archivo primario de hipónimos / hiperónimos.

<i>adornos</i>
<i>anillos</i>
<i>aretes</i>
<i>collares</i>
<i>pendientes</i>
<i>animales</i>
<i>animales domésticos</i>
<i>felinos</i>
<i>rapaces</i>
<i>reptiles</i>
<i>roedores</i>
<i>animales domésticos</i>
<i>cerdo</i>
<i>ganado</i>
<i>gato</i>

<i>pájaro</i>
<i>perro</i>
<i>aparato1</i>
<i>aspiradora</i>
<i>iluminador</i>
<i>lavadora</i>
<i>refrigerador</i>
<i>tostador</i>
<i>ventilador</i>
<i>bebidas</i>
<i>bebida caliente</i>
<i>bebida fuerte</i>
<i>refrescos</i>
<i>bebida caliente</i>
<i>cacao</i>
<i>café</i>
<i>chocolate</i>
<i>té</i>
<i>bebida fuerte</i>
<i>champán</i>
<i>licor</i>
<i>ron</i>
<i>vodka</i>
<i>whisky</i>
<i>refrescos</i>
<i>Coca-cola</i>
<i>Fanta</i>
<i>Mirinda</i>
<i>Pepsi Cola</i>
<i>carne</i>
<i>bisté</i>
<i>chuleta</i>
<i>liebre</i>
<i>pollo</i>
<i>salchichas</i>

Una muestra de la base primaria de hipónimos / hiperónimos se presenta en el *Anexo A.4 Base primaria de hipónimos / hiperónimos*.

3.7.1.1 Clasificación de hipónimos

Los hipónimos, como mencionamos, se agrupan respecto a su concepto más general. Sin embargo, se puede hacer una clasificación dentro de los hipónimos. Se considera como

clasificación a un hipónimo que conste de más de una palabra, la cual debe ser palabra clave para los hipónimos que engloba.

Por ejemplo, para el concepto *animales*, tenemos una clasificación que se encuentra como hipónimo nombrada *animales domésticos*. La clasificación debe ser palabra clave para los hipónimos que subordina como se muestra en el ejemplo.

<i>animales</i>
<i>animales domésticos</i>
<i>felinos</i>
<i>reptiles</i>
<i>animales domésticos</i>
<i>cerdo</i>
<i>ganado</i>
<i>gato</i>
<i>pájaro</i>

Para los nombres propios (de una palabra y más de una palabra) se inicia con mayúscula, esto diferencia de los hipónimos considerados como clasificaciones. Por ejemplo.

bahía

Bahía de Vizcaya

Bahía de Los Ángeles

golfo

Golfo de México

Golfo Pérsico

Esta información es útil para el proceso de inferencia, donde no se intentará generar combinaciones con las etiquetas de clasificación.

3.7.2 Base secundaria

Cada renglón del archivo determina una palabra independiente. Los códigos son definidos del tamaño de un byte para los rasgos morfológicos (columnas 1 a la 6 de la Tabla 3.18) y posteriormente sigue la cadena de letras que representa la palabra.

La posición de códigos en un renglón es la siguiente:

Tabla 3.18 Formato de la base secundaria de hipónimos / hiperónimos

Columna	Descripción
1	Papel que desempeña la palabra (palabra clave versus relacionada)
2	Parte de la oración
3	Género
4	Número
5	Clase de inclinación
6	Palabra corriente

3.7.2.1 Descripción de las columnas

1. Papel que desempeña la palabra en el artículo del diccionario

Tabla 3.19 Tipo de concepto

Valor	Descripción
+	Define la superclase del artículo.
-	Define la subclase.

2. Parte de la oración

Tabla 3.20 Valor de la parte de la oración

Valor	Descripción
0	Clasificación
1	Sustantivo
2	Adjetivo (n/a)
3	Verbo (n/a)
4	Adverbio (n/a)

Las partes de la oración con la notación (n/a), no se aplican a este archivo.

El valor cero es únicamente de clasificación y no se deben realizar inferencias en el momento del enriquecimiento de las colocaciones.

3. Género

Sólo se aplica para sustantivos. Los valores posibles para esta columna son los de la Tabla 3.6.

4. Número

Sólo se aplica para sustantivos. Los valores posibles para esta columna son los de la Tabla 3.7.

5. Clase de inclinación para sustantivos

La clase de inclinación indica las posibles desinencias para la palabra corriente. La clase de inclinación se determina con base a las desinencias de la Tabla 3.8.

6. Palabra corriente

A partir de esta columna comienza la secuencia de caracteres que constituye la palabra. Si existe información semántica, se coloca después de la palabra clave con un marcador especial. Véase el punto 3.3.1.1.

Una muestra de la base secundaria de hipónimos / hiperónimos se presenta en el *Anexo B.4 Base secundaria de hipónimos / hiperónimos*.

Capítulo 4. Caracterización morfológica de las colocaciones

En este capítulo se presenta el procedimiento para la transformación de las bases primarias de colocaciones a las bases secundarias. También, se dan las tablas necesarias para la obtención de las características morfológicas de los componentes de las colocaciones. En la última sección de este capítulo se presentan los diagramas y las herramientas computacionales utilizadas para el desarrollo del sistema *CrossLexica-Esp*.

La metodología usada para identificar los rasgos morfológicos de cada uno de los componentes de las colocaciones, se hace con base a los sufijos característicos de español para cada categoría gramatical y a las hipótesis propuestas para cada archivo que contiene un tipo especial de colocaciones. Esto se detallará en el momento que sea analizado cada archivo que contiene las colocaciones.

4.1 Determinación de las categorías gramaticales

Las categorías gramaticales se determinan con base a los sufijos característicos para cada clase de palabra, a los sufijos característicos para género (masculino y femenino) y para número. Además, se consideraron listas de excepciones de las reglas de determinación de las categorías gramaticales. En los siguientes puntos se dan los sufijos para la determinación de cada categoría gramatical.

4.1.1 Clase de palabra

Agrupamos los sufijos característicos para identificar cada una de las partes de la oración (POS): sustantivos, verbos, adjetivos y adverbios. Las agrupaciones son las siguientes.

4.1.1.1 Sufijos característicos para sustantivos

Para identificar las palabras sustantivas se usan los siguientes sufijos.

ado, aje, ajo, al, ato, ario, astro, avo, azgo, azo, cracia, dad, dero, do, dor, dura, edad, edo, ería, ero, erío, ez, eza, esa, filia, fobia, fono, ia, ía, ío, icio, idad, ín, ión, isa, ismo, ísmo,

ista, itis, itud, és, ez, eza, iz, logía, logo, maquia, mento, metría, miento, ncia, nza, or, pata, patía, sión, sis, sor, tad, tor, tud, ura, zón, is, ma, ta, pa, ora, esa.

4.1.1.2 Sufijos característicos para adjetivos

Para identificar las palabras adjetivas se usan los siguientes sufijos.

ablo, áceo, aco, ado, al, án, anco, ante, ar, ardo, ario, arra, asco, ata, átil, ato, avo, az, ble, bundo, cio, cundo, dero, díó, dizo, do, eco, ego, ojo, el, enario, enco, endo, engo, eno, ense, eño, eo, erno, ero, és, esco, ésimo, este, estre, estre, eta, eyo, í, iaco, íaco, ial, iano, ibundo, ica, icida, icio, icola, ícola, icundo, ida, ido, iego, iego, ién, iento, ífero, ífico, iforme, ífugo, ígneo, igo, ígrado, ijo, il, ílocuo, imo, ín, ína, indo, ino, íntimo, io, ío, iondo, ípeto, isco, ista, ístico, ita, ito, ivo, ívoro, izante, izo, lento, ndero, ndino, ndo, neo, no, ntío, ófago, ófílo, ógeno, oide, ol, ólatra, ómano, ón, orro, oso, ota, ote, ple, plo, tario, ticio, tico, til, tivo-la, to, torio, uco, udo, uence, ueño, ùeño, ujo, ulo, uncho, uno, úpeto, uple, uplo, urno, us-co.

4.1.1.3 Sufijos característicos para verbos

Utilizamos las terminaciones para las tres clases de verbos, además de usar los enclíticos con el pronombre *se* y los verbos conjugados para las oraciones pasivas reflejas, (los verbos conjugados para tercera persona del singular y plural). Las desinencias son las siguientes:

ar, er, ir, arse, erse, irse.

Las desinencias para tercera persona del singular y plural, así como para los tiempos simples (presente, pasado y futuro) son las siguientes:

-a, -an, -ó, -aron, -ará, -arán, -e, -en, -ió, -ieron, -erá, -erán, -irá, -irán.

4.1.1.4 Sufijo característico para adverbios

Para catalogar los adverbios se usa el sufijo estándar *-mente*, además de una lista que contiene los adverbios que no siguen la norma como *cerca, ahora, muy, etc.* Véase *Anexo II. Adverbios*. Las locuciones adverbiales se determinan por la combinación de una preposición y otra parte de la oración y se consideran construcciones adverbiales. Véase el punto 3.4.1.

4.1.2 Categoría de género

Se identifica el género según la terminación de la palabra.

Si la POS es sustantivo.

- Se considera masculino, si la palabra tiene alguna de las siguientes terminaciones:

ente, ante, ar, er, ma, mas, tor, dor, or, án, ín, és, o, ó, e, é, l, x, n, y.

- Se considera femenino, si la palabra tiene alguna de las siguientes terminaciones:

tora, ora, ión, tad, tud, dad, dura, isa, eza, ina, triz, ez, itis, sis, esa, ía, ia, ta, pa, a, z, ed, lis, mis, xis.

- Se considera invariable o neutro, si es un término multipalabra o si la palabra se encuentra en la lista de excepciones.

Si la POS es adjetivo

- Se considera masculino, si la palabra tiene alguna de las siguientes terminaciones:

tor, dor, or, ma, án, ín, o, e, x, ez, oz.

- Se considera femenino, si la palabra tiene alguna de las siguientes terminaciones:

tora, ora, ión, tad, tud, dad, dura, isa, eza, ina, triz, itis, ía, ia, ta, pa, a.

- Se considera invariable o neutro, si la palabra se encuentra en la lista de excepciones para adjetivos o si la palabra tiene alguna de las siguientes terminaciones:

ista, ble, crata, ente, ante, i, í, ú, n, l, r, s, z, ar, er, te, tre.

4.1.3 Categoría de número

Para identificar el número de la palabra se toma en cuenta lo siguiente:

Si la POS es sustantivo o adjetivo.

- Se considera singular, si la palabra tiene alguna de las siguientes terminaciones:

-(s), és.

- Se considera plural, si la palabra tiene alguna de las siguientes terminaciones:
as, es, os, ses, s.
- Se considera invariable o neutro, si la palabra se encuentra en la lista de excepciones de número invariable o si la palabra tiene alguna de las siguientes terminaciones:
sis, tis, lis, nis, mis, cis, dis, xis, quis, us, x.

El símbolo \neg , indica que es cualquier terminación excepto los de la lista.

Se cuenta con archivos de excepciones de género, número, adverbios, etc. para la adecuada caracterización. Véase *Anexo III. Archivos auxiliares*.

Definiremos las siguientes funciones para determinar las clases de palabras y las categorías gramaticales de género y número. Estas funciones serán de utilidad para la definición de las hipótesis y la concordancia del archivo de colocaciones en cuestión.

F1. Sea **P** la función que determina el valor de la parte de la oración para la palabra r , definida como sigue:

$$\mathbf{P}(r) = v \mid v \in \{\text{SUS, ADJ, ADV, VER, CONS_ADJ, CONS_ADV}\}$$

F2. Sea **G** la función que determina el valor de género para la palabra r , definida como sigue:

$$\mathbf{G}(r) = v \mid v \in \{\text{NEUTRO, MASC, FEM}\}$$

y sea $\mathbf{G}(0) = \text{NEUTRO}$

F3. Sea **N** la función que determina el valor de número para la palabra r , definida como sigue:

$$\mathbf{N}(r) = v \mid v \in \{\text{NEUTRO, SING, PLUR}\}$$

y sea $\mathbf{N}(0) = \text{NEUTRO}$

4.2 Determinación de la clase de inclinación

Para encontrar la clase de inclinación, es decir, las posibles desinencias para la palabra, se proponen las siguientes tablas para cada parte de la oración con base al grupo de sufijos. Cada conjunto denota una clase de inclinación.

Tabla 4.1 Clases de inclinación para sustantivos

Clase	Sufijos
0	<i>itis, lis, nis, mis, cis, dis, xis, quis,</i> + lista de excepciones
1	<i>a, e, i, o, u, é, ó</i>
2	<i>d, r, n, l, j, y, á, í, ú, és, b, m</i>
3	<i>z, aces, eces, ices, oces, uces</i>

Tabla 4.2 Clases de inclinación para adjetivos

Clase	Sufijos
0	Lista (ver anexos)
1	<i>o, a</i>
2	<i>á, n, l, r, í, ú</i>
3	<i>ble, nte, ple, te, tre, ense, be, e</i>
4	<i>ota, ote, ete, eta</i>
5	<i>sor, sora, dor, dora, tor, tora, án, in, és, esa</i>
6	<i>z, aces, eces, ices, oces, uces</i>
7	Apócope
8	Apócope <i>grande</i>

Tabla 4.3 Clases de inclinación para verbos

Clase	Sufijos
1	<i>ar</i>
2	<i>er</i>
3	<i>ir</i>
4	<i>arse</i>
5	<i>erse</i>
6	<i>irse</i>
7	<i>a, an, ó, aron, ará, arán</i> (con proclítico <i>se,</i>

Clase	Sufijos
	verbos <i>-ar</i>)
8	<i>e, en, ió, ieron, erá, erán</i> (con proclítico <i>se</i> , verbos <i>-er</i>)
9	<i>e, en, ió, ieron, -irá, -irán</i> (con proclítico <i>se</i> , verbos <i>-ir</i>)

Para las construcciones adjetivales y adverbiales no se refleja la clase de inclinación (la clase es cero) debido a que son invariables o neutros. Los adjetivos son transformados a la forma estándar (singular masculino); los verbos a la forma infinitiva. Los sustantivos no se reducen a la forma singular, se mantienen ambas formas.

Los rasgos idiomáticos y semánticos se transcriben al archivo secundario en la posición según el esquema de cada archivo.

4.3 Procedimiento para colocaciones modificadoras

Se inicia consultando cada artículo de la base primaria de colocaciones modificadoras. El proceso total se divide en dos pasos.

En el primer paso, se determina toda la información morfológica posible de la palabra clave. Si existe algún marco morfológico se determina su característica morfológica y en caso contrario se determina con base a los sufijos para cada categoría gramatical y clase de inclinación del punto 4.1.

Segundo paso, se obtiene la información morfológica de cada palabra relacionada y se verifica que no haya discrepancia de concordancia entre las palabras relacionadas y la palabra clave. Esto se realiza haciendo uso de las hipótesis para colocaciones modificadoras y se hace la comprobación de la concordancia de género y número.

4.3.1 Hipótesis para colocaciones modificadoras

Sea w la palabra clave y s la palabra relacionada, entonces:

1. Si $P(w) = \text{SUS}$, tenemos que, $\forall i \prod_{i=1}^n P(s_i) = \text{ADJ} \vee \text{CONS_ADJ}$ $n > 0$

2. Si $P(w) = (\text{ADJ} \vee \text{VER} \vee \text{ADV})$, tenemos que, $\forall i \prod_{i=1}^n P(s_i) = \text{ADV} \vee \text{CONS_ADV}$ $n > 0$

Donde CONS_ADJ y CONS_ADV es construcción adjetival y construcción adverbial respectivamente.

4.3.2 Comprobación de la concordancia

Con las hipótesis anteriores comprobamos que las combinaciones concuerden en género y número. Estas comprobaciones se realizan solamente para la primera hipótesis.

Considerando las funciones de género y número (F2 y F3), tenemos que para comprobar la concordancia del género se define:

$$\forall i \prod_{i=1}^n G(s_i) = G(w) \vee G(0) \quad n > 0$$

y para comprobar la concordancia del número se define:

$$\forall i \prod_{i=1}^n N(s_i) = N(w) \vee N(0) \quad n > 0$$

4.4 Procedimiento para colocaciones con valencias verbales

Se inicia consultando cada artículo de la base primaria de colocaciones con valencias verbales. El proceso se divide en dos pasos.

En el primer paso, se determina la información morfológica de la palabra clave con base al punto 4.1.1 y se encuentran los componentes de la palabra relacionada: preposición, tiempo verbal, persona, artículo, posición de la valencia omitida, pronombre *se*, posición de la palabra clave y se obtienen los rasgos morfológicos de la palabra relacionada de acuerdo al punto 4.1.1.

En el segundo paso, se comprueba la concordancia entre el verbo y la persona gramatical.

Para la palabra relacionada, puede contener palabras funcionales y valencia omitida, el verbo debe quedar aislado de las palabras funcionales. Entonces, se procede a buscar la existencia de preposición y de artículo. Se identifica el número de preposición con base al archivo de preposiciones (*Anexo I. Preposiciones*). Si existe el artículo, se identifica si es definido o indefinido. El artículo se ajusta (en género y número) al sustantivo que lo acom-

pañía en el momento que el sistema muestra la combinación. Se procede con la búsqueda de la valencia omitida y la marca de sustitución para la palabra clave.

Al extraer las partes funcionales que enlazan el verbo y el sustantivo, el verbo queda aislado. Una vez aislado el verbo, se determina la existencia del enclítico *se* o como proclítico de oraciones pasivas reflejas. En el primer caso, se elimina el enclítico y se determina la clase de inclinación (véase Tabla 4.3). En el segundo caso, se elimina el pronombre y se determina el tiempo verbal y persona de acuerdo a las desinencias para esta clase (véase el punto 4.2). Una vez identificados los rasgos morfológicos se estandariza el verbo a la forma infinitiva.

Para este archivo de colocaciones se considera la siguiente hipótesis y se hace la comprobación para la persona gramatical.

4.4.1 Hipótesis para colocaciones con valencias verbales

Sea la palabra clave w y la palabra relacionada s .

$$1. \quad P(w) = \text{SUS}, \text{ tenemos que, } \forall i \prod_{i=1}^n P(s_i) = \text{VER} \quad n > 0$$

Toda palabra clave w en este archivo primario es exclusivamente sustantivo y las palabras relacionadas son verbos.

4.4.2 Comprobación de la concordancia

Para los verbos conjugados se realiza la comprobación entre el verbo y el sustantivo. Tomemos en cuenta que por permitir el uso de oraciones pasivas reflejas, estamos obligados a utilizar conjugaciones para la tercera persona del singular y plural. Si el verbo se encuentra conjugado, la comprobación se realiza para el número y la persona.

Se define la siguiente función para la determinación de la persona gramatical del verbo.

Sea **Per** la función que determina la persona gramatical del verbo r y se define como sigue:

$$\mathbf{Per}(r) = v \mid v \in \{1_Per_SING, 2_Per_SING, 3_Per_SING, 1_Per_PLUR, 2_Per_PLUR, 3_Per_PLUR\}$$

Entonces,

$$\mathbf{Per}(s_i) = 3_Per_SING \vee 3_Per_PLUR$$

y para comprobar la concordancia del número se define:

$$N(s_i) = N(w)$$

1_Per_SING, indica primera persona singular, y así con todas las demás personas gramaticales.

4.5 Procedimiento para colocaciones con valencias sustantivales

Se consulta cada artículo de la base primaria de colocaciones con valencias sustantivales. Se determina la información morfológica para la palabra clave con base a los sufijos para cada categoría gramatical (ver el punto 4.1.1). Se procede con la búsqueda de los componentes funcionales de la palabra relacionada.

Para la palabra relacionada, se busca la existencia de la preposición y el artículo. La preposición se empata con el número de preposición correspondiente del archivo de preposiciones (Anexo I). Para realizar la caracterización se considera la siguiente hipótesis.

4.5.1 Hipótesis para colocaciones con valencias sustantivales

Sea la palabra clave w y la palabra relacionada s .

1. Si $P(w) = \text{SUS}$, tenemos que, $\forall i P(s_i) = \text{SUS} \quad n > 0$

Toda palabra clave w en este archivo es exclusivamente sustantivo, así como las palabras relacionadas.

En este archivo no se hace ningún tipo de comprobación de concordancia.

4.6 Procedimiento para hipónimos / hiperónimos

Se consulta cada artículo de la base primaria de hipónimos / hiperónimos. En este archivo sólo existen conceptos. Se determinan los rasgos morfológicos con base a los sufijos para determinar las categorías gramaticales (ver el punto 4.1.1). Esto se realiza para la palabra clase y subclase. Para las etiquetas usadas como clasificaciones no se determina ningún rasgo morfológico.

Para este archivo no se considera ninguna hipótesis ni se comprueba ninguna concordancia.

4.7 Implementación del sistema CrossLexica-Esp

El sistema *CrossLexica-Esp* se desarrolló en C++ Builder 5.0 y el manejador de la base de datos fue Paradox. Se optó por este manejador por ser nativo de C++ Builder y por consiguiente es fácil de implementar. Además, al realizar consultas a la base de datos, se tiene mejor rendimiento que con otros manejadores como Access [35] y utilizar un manejador como SQL Server necesita de mayores requerimientos de hardware [38] que el usuario promedio posee.

Los requerimientos mínimos para la instalación del sistema son:

- Procesador Pentium
- 32 MB de memoria RAM
- Sistema operativo Windows (95/98/2000/NT)
- Espacio disponible en el disco duro 20 MB

4.8 Diagramas de casos de uso

El diagrama de casos de uso del sistema *CrossLexica-Esp* se presenta en la figura 4.1 [37].

La descripción para los casos de usos es la siguiente:

Caso de uso: Formatear colocaciones

Actores: Archivos de colocaciones y archivos formateados de colocaciones.

Propósito: Transformar los archivos de colocaciones a una forma secundaria con rasgos morfológicos y semánticos de cada componente de la colocación.

Resumen: Un artículo es leído del archivo de entrada y se obtienen los rasgos morfológicos de los dos componentes de la colocación. Una vez obtenidos los rasgos, se verifica la concordancia sintáctica y se graba en un archivo secundario.

Flujo de sucesos

1. Inicia con la llegada de un artículo (varias colocaciones con una misma palabra clave).
2. Se determinan los rasgos morfológicos de la palabra clave y de cada palabra relacionada. Para cada colocación se comprueba la concordancia entre los componentes.

- Al terminar de comprobar la concordancia y no encontrar errores, se graba la colocación.

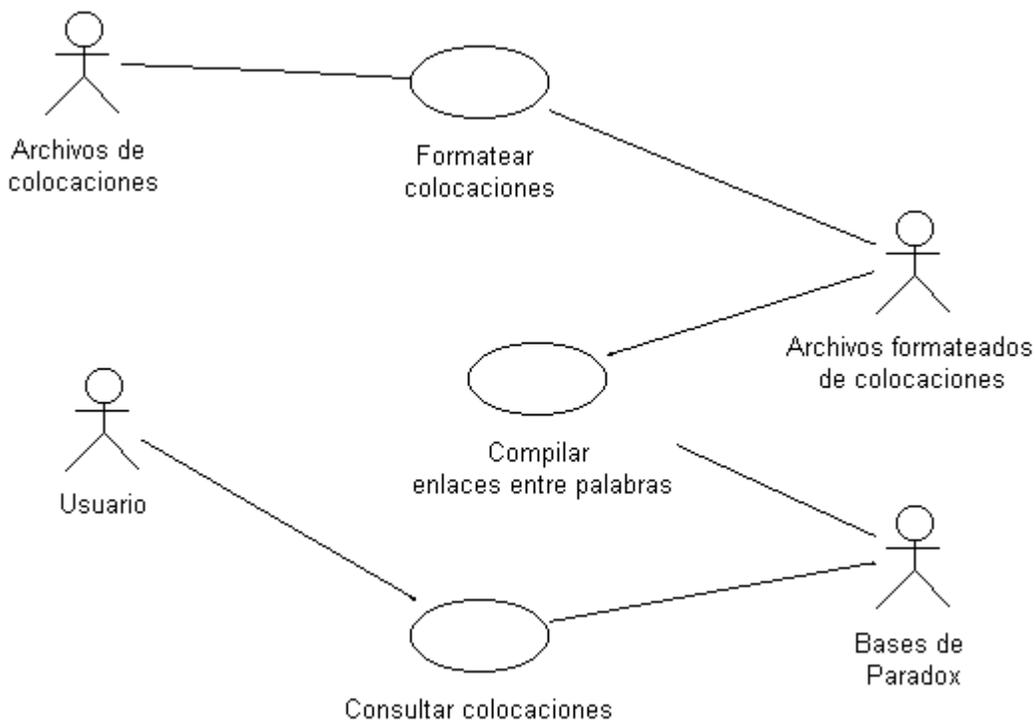


Figura 4.1 Diagrama de casos de uso del sistema CrossLexica-Esp

Caso de uso: Compilar enlaces entre palabras

Actores: Archivos de colocaciones formateados y Bases Paradox.

Propósito: Generar el diccionario general de palabras y la base de datos de enlace sintácticos y semánticos entre palabras.

Resumen: Un artículo es leído del archivo de entrada, se insertan los componentes al diccionario general y a la base de enlaces entre palabras.

Flujo de sucesos

- Inicia con la llegada de un artículo.
- Se separan las colocaciones del artículo y se verifica la existencia de los componentes de la colocación en el diccionario general. Si no existen los componentes se insertan a la base con todas sus características morfológicas.

3. Para todas las colocaciones, se insertan los enlaces entre palabras.

Caso de uso: Consultar colocaciones

Actores: Usuario y Bases de Paradox.

Propósito: Mostrar las colocaciones con los papeles sintácticos y semánticos que desempeñan entre las palabras y generar posibles colocaciones no explícitas en la base de datos.

Resumen: Se introduce una palabra o se selecciona del diccionario general. Con la palabra solicitada, se buscan los enlaces sintácticos y semánticos; posteriormente, se intentan generar nuevas colocaciones con base a la información sintáctica y semántica de las colocaciones previas.

Flujo de sucesos

1. El usuario introduce o selecciona una palabra que pertenece al diccionario general.
2. Se buscan los enlaces sintácticos en la base de datos y se muestran al usuario.
3. Se buscan los enlaces semánticos en la base de datos y se muestran al usuario.
4. Con base a la información anterior, se infieren las nuevas colocaciones. Si se generaron las colocaciones, se muestran al usuario.
5. Termina cuando el usuario cierra la aplicación.

En las figuras 4.2, 4.3 y 4.4 se muestran los diagramas de secuencia para los casos de uso de la figura 4.1.

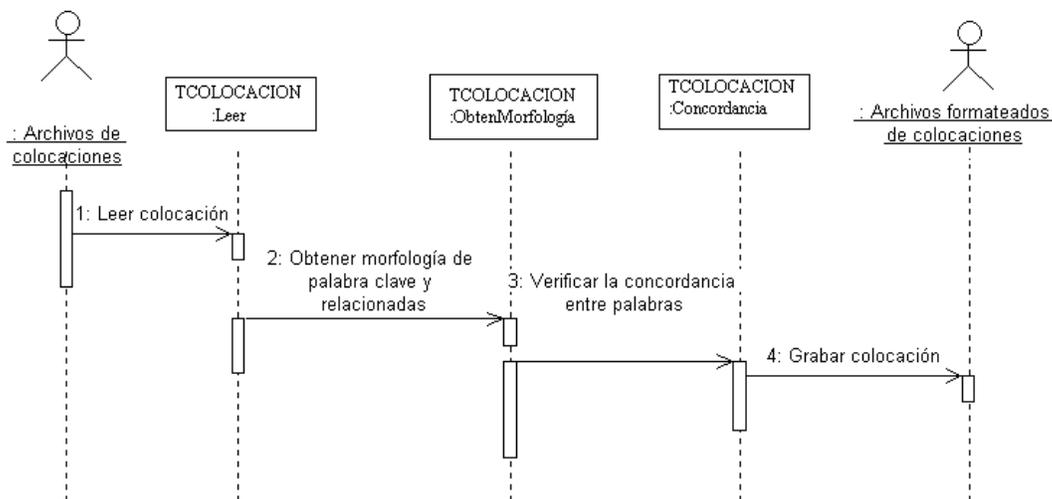


Figura 4.2 Diagrama de secuencia del caso de uso *Formatear colocaciones*

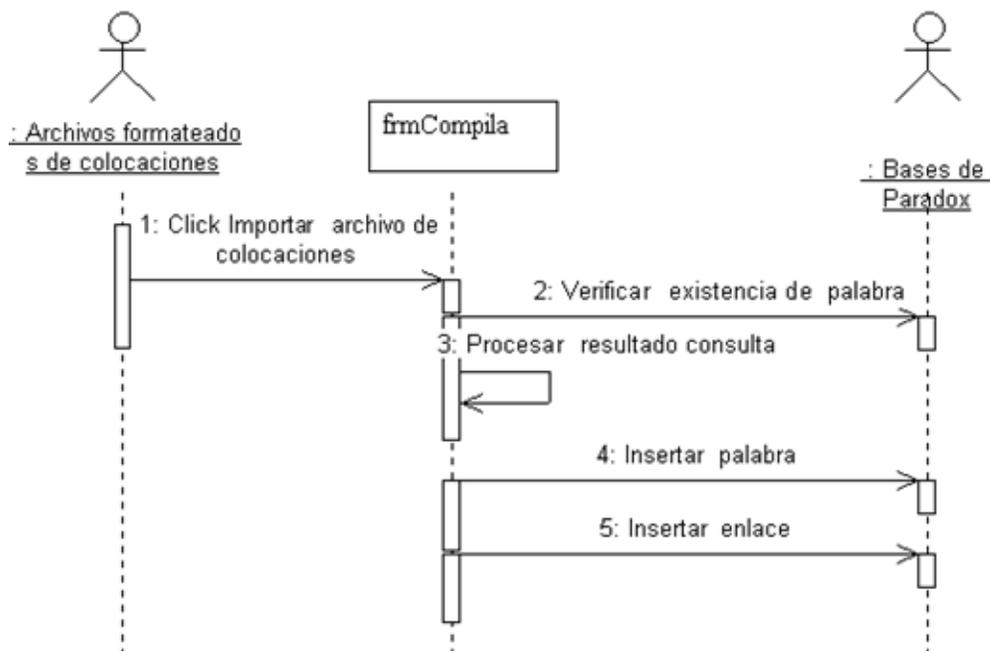


Figura 4.3 Diagrama de secuencia del caso de uso *Compilar enlaces entre palabras*

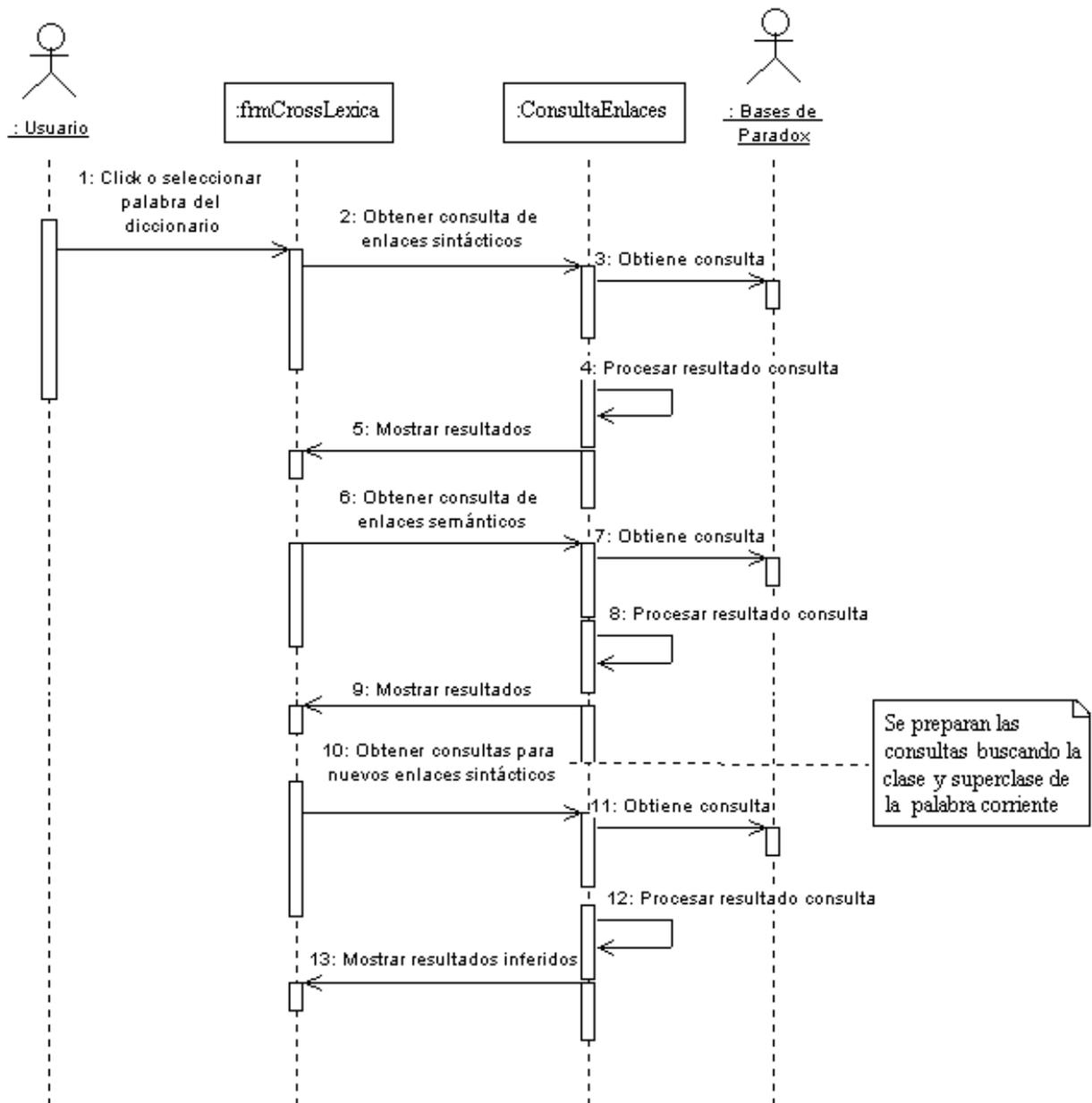


Figura 4.4 Diagrama de secuencia del caso de uso *Consultar colocaciones*

En la figura 4.5 se muestra el diagrama entidad-relación del sistema *CrossLexica-Esp* [36], en esta figura se muestran las relaciones entre las tablas y la cardinalidad entre ellas. Cada tabla mantiene una relación uno–muchos, con respecto a la tabla *Diccionario*. Actualmente, las tablas de colocaciones modificadoras (*ColMod*), sustantivales (*ColSus*), verbales (*ColVerb*) y los hipónimo / hiperónimos entre las palabras (*Conceptos*) tienen aproximadamente 13 mil registros y el diccionario general tiene aproximadamente 5 mil registros.

El volumen de la información es manejable como un estudio que se realizó en el laboratorio de lenguaje natural del CIC [39] y el presentado en [2]. La frecuencia de ocurrencia de las palabras sigue una variante de ley de Zipf. Esta distribución no es sublineal, sino tienen una forma sesgada y se aproxima más al inverso del cuadrado de j , es decir, existe un conjunto pequeño de palabras muy frecuente y muchas que aparecen muy pocas veces, para cualquier idioma [39].

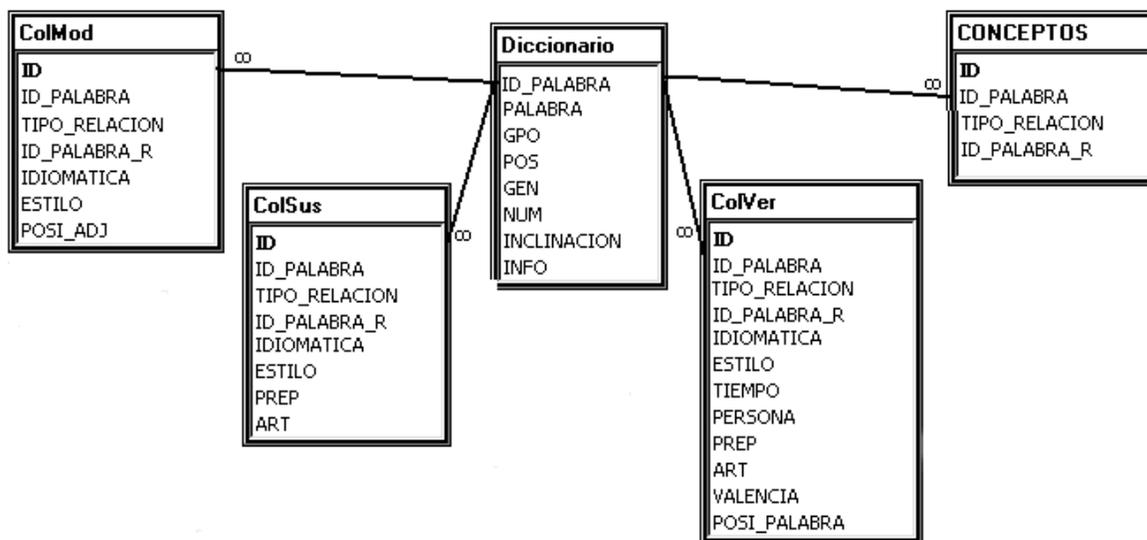


Figura 4.5 Modelo entidad-relación del sistema CrossLexica-Esp

El diccionario de datos de la tabla que contiene las palabras únicas, tabla *Diccionario*, se presenta en la Tabla 4.4.

Los enlaces sintácticos se encuentran en tres tablas. Para los enlaces del tipo modificador, se encuentran codificados en la tabla *ColMod*, el diccionario de datos correspondiente se presenta en la Tabla 4.5. Los enlaces para los complementos sustantivales, se encuentran

codificados en la tabla *ColSus*, el diccionario de datos correspondiente se presenta en la Tabla 4.6. Los enlaces para los complementos verbales, se encuentran codificados en la tabla *ColVerb*, su diccionario de datos es el de la Tabla 4.7.

Los enlaces semánticos de superclase y subclase, se encuentran codificados en la tabla *Conceptos*, el diccionario de datos correspondiente es el de la Tabla 4.8.

Tabla 4.4 Diccionario de datos de la tabla *Diccionario*

	Campo	Tipo de dato	Descripción
1	ID_PALABRA	Numérico	Identificador de la palabra dentro del diccionario. Número consecutivo (llave)
2	PALABRA	Texto(50)	Conjunto de caracteres que definen la palabra.
3	GPO	Numérico	Número consecutivo para identificar las acepciones de la palabra.
4	POS	Numérico	Identificador de parte de la oración.
5	GEN	Numérico	Identificador del género de la palabra.
6	NUM	Numérico	Identificador del número de la palabra.
7	INCLINACION	Numérico	Clase de inclinación de la palabra
8	INFO	Texto(50)	Información semántica de la palabra para ayuda del usuario.

Tabla 4.5 Diccionario de datos de la tabla modificadores: *ColMod*

	Campo	Tipo de dato	Descripción
1	ID	Numérico	Índice de la tabla (llave).
2	ID_PALABRA	Numérico	Identificador de la palabra.
3	TIPO_RELACION	Texto(30)	Papel sintáctico que desempeña la palabra con respecto a su palabra relacionada.
4	ID_PALABRA_R	Numérico	Identificador de la palabra relacionada.
4	IDIOMÁTICA	Texto (1)	Marca para identificar que la combinación de palabras es idiomática.

6	ESTILO	Texto (1)	Marca para identificar que la combinación de palabras se usa de forma especial en los textos.
7	POSI_ADJ	Numérico	Indica si el adjetivo se debe colocar antepuesto a su sustantivo

Tabla 4.6 Diccionario de datos de la tabla de complementos sustantivales: *ColSus*

	Campo	Tipo de dato	Descripción
1	ID	Numérico	Índice de la tabla (llave).
2	ID_PALABRA	Numérico	Identificador de la palabra.
3	TIPO_RELACION	Texto(30)	Papel sintáctico que desempeña la palabra con respecto a su palabra relacionada.
4	ID_PALABRA_R	Numérico	Identificador de la palabra relacionada.
4	IDIOMÁTICA	Texto (1)	Marca para identificar que la combinación de palabras es idiomática.
6	ESTILO	Texto (1)	Marca para identificar que la combinación de palabras se usa de forma especial en los textos.
7	PREP	Numérico	Número de la preposición que conecta a las palabras.
8	ART	Numérico	Número de grupo de artículos (definido o indefinido).

Tabla 4.7 Diccionario de datos de la tabla de complementos verbales: *ColVer*

	Campo	Tipo de dato	Descripción
1	ID	Numérico	Índice de la tabla (llave).
2	ID_PALABRA	Numérico	Identificador de la palabra.
3	TIPO_RELACION	Texto(30)	Papel sintáctico que desempeña la palabra con respecto a la palabra relacionada.

	Campo	Tipo de dato	Descripción
4	ID_PALABRA_R	Numérico	Identificador de la palabra relacionada.
5	IDIOMÁTICA	Texto (1)	Marca para identificar que la combinación de palabras es idiomática.
6	ESTILO	Texto (1)	Marca para identificar que la combinación de palabras se usa de forma especial en los textos.
7	TIEMPO	Numérico	Indica el tiempo de conjugación del verbo.
8	PERSONA	Numérico	Indica la persona de conjugación del verbo.
9	PREP	Numérico	Número de la preposición que conecta a las palabras.
10	ART	Numérico	Número de grupo de artículos (definido o indefinido).
11	VALENCIA	Numérico	Indica la existencia y posición de la posible valencia semántica que se omitió
12	POSI_PALABRA	Numérico	Indica la posición en que se coloca la palabra sustantiva dentro de la colocación

Tabla 4.8 Diccionario de datos de l atabla de clases y subclases: *Conceptos*

	Campo	Tipo de dato	Descripción
1	ID	Numérico	Índice de la tabla (llave).
2	ID_PALABRA	Numérico	Identificador de la palabra.
3	TIPO_RELACION	Texto(30)	Papel sintáctico que desempeña la palabra con respecto a su palabra relacionada.
4	ID_PALABRA_R	Numérico	Identificador de la palabra relacionada.

El funcionamiento del sistema se muestra en el capítulo cinco.

Capítulo 5. Resultados del sistema CrossLexica-Esp

En este capítulo se presentan dos resultados importantes, primero: los resultados obtenidos del formateo de las colocaciones, los que se explican con algunos segmentos de las bases formateadas de colocaciones; segundo: el sistema que demuestra las colocaciones, esto es, la interfaz del sistema *CrossLexica-Esp*, en esta sección se presentan el funcionamiento del sistema.

5.1 Resultados de la caracterización morfológica

Después del proceso de formateo, se obtienen los rasgos morfológicos de las colocaciones, los que se reflejan en las bases secundarias.

En los siguientes puntos se describen los resultados del formateo para cada tipo de colocaciones.

5.1.1 Descripción de los resultados de la base secundaria de colocaciones modificadoras

En el siguiente cuadro se muestra un segmento de la base secundaria de colocaciones modificadoras caracterizada morfológicamente.

+11110	<i>absolutismo</i>
-20030	<i>ilustre</i>
-20010	<i>impuesto</i>
-20050	<i>francés</i>
-20020	<i>provincial</i>
-20020	<i>real</i>
+12120	<i>acción</i>
-20010	<i>armado</i>
-20010	<i>bueno</i>
-20010	<i>combinado</i>
-20000	<i>de gracias</i>
+12120	<i>actividad</i>
-20010	<i>bélico</i>
-20010	<i>bullicioso</i>
-20010	<i>científico</i>
-20010	<i>deportivo</i>
-20010	<i>económico</i>
+11210	<i>actos</i>

RESULTADOS DE LA CARACTERIZACIÓN MORFOLÓGICA

-20010 delictivo
 -20010 público
 -20010 reclamado
 -20010 religioso
 +11110 agente1<CLASS:hombre
 -20010 británico
 -20000 de bolsa
 -20000 de negocios
 -20000 de policía
 +11110 agente2<CLASS:sustancia
 -20010 dañino
 -20000 de limpieza
 -20010 patógeno
 -20010 químico
 -20020 viral
 +20010 agradecido
 -40000 muy
 -40000 poco
 -40000 sumamente
 +11110 caballo
 -20010 fogoso
 -20030 azabache
 +11210 **caballos**
 -20000 **Sde fuerza**
 +12110 cabeza
 -20010 blanco
 -20010 bueno
 -20000& de chorlito
 -20010 descubierto
 +11120 capital1
 -20010 básico
 -20030 circulante
 -20030 constante
 -20010 extranjero
 -20010 financiero
 -20020 fundamental
 -20010 propio
 -20020 social
 -20030 variable
 +12120 capital2
 -20010 mexicano
 +11210 **derecho1<"leyes"**
 -20020 civil
 -20000 de autor
 -20010 humano
 -20020 universal
 +11210 **derecho2<"pagos"**
 -20010 arancelario
 -20000 de aduana
 -20000 de entrada
 +11110 derrame
 -20020 cerebral

CAPÍTULO 5. RESULTADOS DEL SISTEMA *CROSSLEXICA-ESP*

+30010	desarrollar
-40000	armónicamente
-40000	independientemente
+11110	dinero
-20020	constante
-20010	efectivo
-20010	suelto

La primera columna nos indica el papel que desempeña la palabra corriente. Las siguientes cuatro columnas (2-5) muestran la información morfológica de la palabra. La columna 6, sólo para adjetivos, indica si es antepuesto el adjetivo. Las siguientes dos columnas (7 y 8) reflejan la idiomatización y estilo de la combinación de palabras; estas columnas sólo son usadas por las palabras relacionadas.

Por ejemplo, para el siguiente artículo se presenta la siguiente interpretación:

+12120 acción
-20010 armado
-20010 bueno
-20000 de gracias

Para la palabra clave *acción* se tiene su información morfológica la cual es sustantivo (1), femenino (2), singular (1) y la clase de inclinación es de tipo dos (2), esto es, las posibles desinencias son \emptyset /es: *acción* y *acciones*. Véase Tabla 4.1.

Para la palabra relacionada *armado* la interpretación es la siguiente:

POS: Adjetivo (2).

Género: No se refleja (0).

Número: No se refleja (0).

Clase de inclinación: 1.

armado / *armada* / *armados* / *armadas*

Idiomatización: neutra (vacía).

Estilo: neutro (vacío).

Para la palabra relacionada *bueno* la interpretación es la siguiente:

POS: Adjetivo (2).

Género: No se refleja (0).

Número: No se refleja (0).

Clase de inclinación: 1.

bueno / *bueno* / *buenos* / *buenas*

Idiomatización: neutra (vacía).

Estilo: neutro (vacío).

Para la palabra relacionada *de gracias* la interpretación es la siguiente:

POS: Adjetivo (2).

Género: No se refleja (0).

Número: No se refleja (0).

Clase de inclinación: invariable (0).

Idiomatización: neutra (vacía).

Estilo: neutro (vacío).

Obsérvese que los adjetivos están en la forma estándar, y la forma correcta del adjetivo se crea en el momento de generar la combinación. Por ejemplo, para *acción*, al combinarla con el adjetivo *armado*, el adjetivo debe estar en concordancia con el sustantivo que lo acompaña, el cual tiene los rasgos morfológicos: femenino y singular. Por lo tanto *armado* sólo puede adquirir la forma *armada* para construir la combinación correcta, *acción armada*. Las posibles desinencias para las clases de adjetivos se detallan en la Tabla 4.2.

5.1.1.1 Información semántica

La información semántica se mantiene en el archivo secundario para uso futuro. La información semántica se coloca inmediatamente después de la palabra clave, con el marco adecuado (ver el punto 3.3.1.1). Ésta es útil para explicaciones al usuario de la palabra corriente e inferencia de nuevas colocaciones.

Por ejemplo, para la palabra clave *agente* se tienen dos acepciones (*agente*, hombre y *agente*, sustancia), se mantiene la clase semántica a la que pertenece con el marco <CLASS: seguido de la clase. Esta información es de tipo explicativa para el usuario y al mismo tiempo útil para la inferencia de colocaciones, como en *agente1*<CLASS:hombre y *agente2*<CLASS:sustancia.

También se mantiene la información para distinguir las palabras homónimas como el caso anterior. Se mantiene el marco y la explicación de la palabra corriente, <"...". Por ejemplo, para la palabra clave *derecho* tenemos dos acepciones (*derecho* —relativo a la ley— y *derecho* —relativo a los pagos—). Se mantiene la información para explicaciones futuras al usuario, como en *derecho1*<"leyes" y *derecho2*<"pagos".

5.1.1.2 Información estilística

Las características de estilo e idiomatización se mantienen en el archivo secundario como se observa en la combinación *cabeza de chorlito* que nos informa que se usa de manera idiomática (&) y *caballos de fuerza* que es de tipo científica (S). Véase 3.3.1.2 *Parámetros morfológicos*.

La base secundaria de colocaciones modificadoras se encuentra de forma extensa en el *Anexo B.1 Base secundaria de colocaciones modificadoras*.

5.1.2 Descripción de los resultados de la base secundaria de colocaciones con valencias verbales

En el siguiente cuadro se muestra un segmento de la base secundaria de colocaciones con valencias verbales caracterizadas morfológicamente.

+1111	accidente
-3003	00ÿ103prevenir
+1221	aceitunas
-3001&	00>104llegar
+1111	acertijo
-3001	00ÿ203adivinar
+1121	acontecimientos
-3004	00'104adelantar
+1211	adivinanza
-3001	00ÿ203contar
+1211	agua
-3002	00ÿ002beber
-3003	00>125diluir
-3005	00;104meter
-3001	00'104saltar
-3001	00ÿ002tomar
+1221	aguas
-3004	00>003curar
-3001	00ÿ103tomar
+1112	alcohol
-3003	00ÿ032diluir
+1111	alcoholismo
-3003	00ÿ002combatir
+1211	alegría
-3001	00 003brincar
-3001	00 003saltar
-3003	00ÿ002sentir
+1211	almohada
-3001&	00>125consultar
+1111	alumno
-3001	00ÿ203preguntar
+1211	angustia

RESULTADOS DE LA CARACTERIZACIÓN MORFOLÓGICA

-3009	13ÿ02loprimir
+1111	apoyo
-3001	00ÿ002prestar
+1212	apelación
-3002	00ÿ103interponer
+1221	ascuas
-3001&	00;003estar
+1121	cántaros
-3002&	00'003llover
+1112	corazón
-3001	23ÿ001brincar
-3001	23ÿ001saltar
+1121	dados
-3001&	00'104jugar
+1213	luz
-3001&	00'003dar
-3001&	00'104sacar
-3003&	00'103salir
+1111	dinero
-3001	00ÿ103ganar
-3007	23ÿ001disipar
-3007	23ÿ001esfumar
+1111	enigma
-3001	00ÿ203acertar
-3001	00ÿ203dar
+1211	enseñanza
-3001	00ÿ103apoyar
-3001	00'104enfocar
+1111	entusiasmo
-3001	00ÿ103contagiar
+1121	equipos
-3001	00'104integrar
+1122	errores
-3002	00ÿ002cometer
+1211	escalera
-3001	00ÿ103bajar
-3003	00ÿ103subir
+1111	fuego1
-3002	00ÿ103encender
-3002	00ÿ103prender
-3001	00'125echar
+1111	fuego2
-3001	00ÿ103reglar
+1221	fuerzas
-3001	00ÿ002recobrar
+1222	funciones2
-3003	00ÿ002admitir
-3002	00ÿ002ejercer
-3001	00ÿ002realizar
+1221	ganas
-3002&	00ÿ002tener
+1121	gastos

-3002	00>104	correr
+1112		gol
-3001	00ÿ203	marcar
-3002	00ÿ203	meter

La primera columna nos indica el papel que desempeña la palabra corriente. Las siguientes cuatro columnas muestran la información morfológica de la palabra corriente. A partir de la columna seis, las columnas restantes sólo se usan por las palabras relacionadas. Las siguientes dos columnas (6 y 7) reflejan la idiomatización y estilo de la combinación de palabras. Las siguientes dos columnas (8 y 9) indican el tiempo de conjugación del verbo y la persona.

Las siguientes cuatro columnas (10-13) son características para generar correctamente la combinación. Éstas son, el número de la preposición de forma codificada (un byte), número del artículo, posición de la valencia omitida y posición de la palabra clave en la combinación. Posteriormente, se encuentra la palabra corriente seguida de su información semántica.

Por ejemplo, para el artículo de abajo se presenta la siguiente interpretación:

```
+1211      agua
-3002 00ÿ002beber
-3003 00>125diluir
```

Para la palabra clave *agua* se tiene su información morfológica la cual es sustantivo (1), femenino (2), singular (1) y la clase de inclinación es de tipo uno, esto es, las posibles desinencias son \emptyset /s: *agua* y *aguas*.

Para la palabra relacionada *beber* la interpretación es la siguiente.

POS: Verbo (3).

Género: No se refleja.

Número: No se refleja.

Clase de inclinación: 2 (infinitivo terminación *-er*).

Idiomatización: neutra (vacía).

Estilo: neutro (vacío).

Tiempo de conjugación: ninguno (0).

Persona de conjugación: ninguna (0).

Número de preposición: ninguna (de forma codificada, 255).

Artículo: ninguno (0).

Valencia omitida: ninguna (0).

Posición de la palabra clave: 2.

La combinación es *beber agua*.

Para la palabra relacionada *diluir* la interpretación es la siguiente.

POS: Verbo (3).

Género: No se refleja.

Número: No se refleja.

Clase de inclinación: 3 (infinitivo terminación -ir).

Idiomatización: neutra (vacía).

Estilo: neutro (vacío).

Tiempo de conjugación: ninguno (0).

Persona de conjugación: ninguna (0).

Número de preposición: *con* (de forma codificada, 112).

Artículo: artículo definido (1, véase Tabla 3.16 Artículos)

Valencia omitida: 2 (posición dentro de la combinación).

Posición de la palabra clave: 5 (posición dentro de la combinación).

La combinación es *diluir ... con el agua*.

Para el siguiente artículo, las columnas para tiempo y persona (columnas 8 y 9), indican que se debe realizar la conjugación del verbo.

```
+1112      corazón
-3001  23ÿ001brincar
-3001  23ÿ001saltar
```

Para la combinación *corazón* y el verbo *brincar*, con los valores de tiempo y persona (valores: 2 y 3 respectivamente), indica que se debe conjugar el verbo. En este caso, el tiempo es dos (tiempo pasado), y la persona es la tercera (tercera persona del singular), la combinación es *corazón brincó*. De la misma forma se debe conjugar el verbo *saltar* obteniendo la combinación *corazón saltó*.

Para el siguiente artículo, las columnas para tiempo y persona indican que se debe realizar la conjugación del verbo y la clase de inclinación (columna 5) se refiere a un proclítico (valor 7).

```
+1111      dinero
-3001  00ÿ103ganar
-3007  23ÿ001disipar
-3007  23ÿ001esfumar
```

Para la combinación *dinero* con el verbo *disipar* y *esfumarse*, con los valores de tiempo y persona además de la existencia de pronombre *se*, se obtienen las combinaciones *dinero se disipó* y *dinero se esfumó*.

Los rasgos estilísticos e idiomáticos se conservan como en las colocaciones modificadoras.

```
+1213          luz
-3001& 00'003dar
-3001& 00'104sacar
```

Por ejemplo, en las combinaciones *dar luz* y *sacar luz*, se indica que se usa idiomáticamente y las combinaciones son '*dar a luz*' y '*sacar a la luz*'.

La base secundaria de colocaciones con valencias verbales se encuentra de forma extensa en el *Anexo B.2 Base secundaria de colocaciones con valencias verbales*.

5.1.3 Descripción de los resultados de la base secundaria de colocaciones con valencias sustantivales

En el siguiente cuadro se muestra un segmento de la base secundaria de colocaciones con valencias sustantivales caracterizadas morfológicamente.

+1111	adiestramiento
-1211	1fase
+1212	administración1
-1121	1servicios
-1211	1ventaja
-1221	1ventajas
+1112	administrador
-1121	1permisos
+1111	adverbio
-1212	1funcionalidad
+1211	agua
-1112	1nivel
+1111	amo
-1111	1dominio
+1120	análisis2
-1111	1costo
-1111	1pago
-1111	1resultado
+1112	animal
-1211	1naturaleza
-1212	1sensibilidad
-1111	1nombre
-1211	1cola
+1111	año
-1221	1ventas
+1212	aplicación
-1113	1interfaz

RESULTADOS DE LA CARACTERIZACIÓN MORFOLÓGICA

+1111	archivo
-1111	lcontenido
-1211	lcontraseña
-1212	ldescripción
-1212	l extensión
-1111	lformato
-1111	lnombre
-1212	lpetición
-1212	lubicación
+1112	autor
-1111	lnombre
+1212	aviación
-1121	lbombardos
+1112	ayer
-1111	lrecuerdo
+1211	base
-1212	linformación
-1212	lubicación
-1222	lunidades
-1121	ldatos
-1221	lcolumnas
-1211	lconsulta
-1121	lcampos
-1211	lsalida
-1111	lregistro
-1111	lángulo
+1111	enemigo
-1111	ælfuego
+1211	hambre
-1112	ldolor
-1111	El sufrimiento
+1211	naturaleza
-1211	>larmonía
-1111	>lcontacto
-1212	lconservación
-1211	ldefensa
-1121	lelementos
-1111	lenemigo
+1212	oración1<" gramática "
-1212	lconstrucción
-1211	lparte
+1212	oración2<" religión "
-1121	lpasos
+1222	organizaciones
-1211	lnaturaleza
+1111	órgano
-1222	ldimensiones
+1111	país
-1221	láreas
-1122	lciviles
-1111	loeste
-1222	lciudades

-1211	1estructura
-1111	1presidente

La primera columna nos indica el papel que desempeña la palabra corriente. Las siguientes cuatro columnas (2-5) muestran la información morfológica de la palabra corriente. Las siguientes dos columnas (6 y 7) reflejan la idiomatización y estilo de la combinación.

Las siguientes dos columnas (8 y 9) son características para generar correctamente la combinación. Éstas son, el número de la preposición de forma codificada (un byte) y el tipo de artículo.

Por ejemplo, para el siguiente artículo se presenta la siguiente interpretación:

```
+1212   administración1
-1121   1servicios
-1211   1ventaja
-1221   1ventajas
```

Para la palabra clave *administración* se tiene su información morfológica la cual es sustantivo (1), femenino (2), singular (1) y la clase de inclinación es de tipo dos, esto es, las posibles desinencias son \emptyset /es: *administración* \emptyset / *administraciones*.

Para la palabra relacionada *servicios* la interpretación es la siguiente.

POS: Sustantivo (1).
 Género: Masculino (1).
 Número: Plural (2).
 Clase de inclinación: 1 (*servicio* \emptyset / *servicios*).

Idiomatización: neutra (vacía).
 Estilo: neutro (vacío).
 Número de preposición: *de* (de forma codificada, 96).
 Artículo: Artículo definido (1).
 La combinación es ***servicios de la administración***

Para la palabra relacionada *ventaja* la interpretación es la siguiente.

POS: sustantivo (1).
 Género: Femenino (2).
 Número: Singular (1).
 Clase de inclinación: 1 (*ventaja* \emptyset / *ventajas*).

Idiomatización: neutra (vacía).

Estilo: neutro (vacío).

Número de preposición: *de* (de forma codificada, 96).

Artículo: Artículo definido (1).

La combinación es ***ventaja de la administración***.

El artículo se ajusta en número y género de acuerdo al sustantivo que lo acompaña.

Se mantiene la información para las palabras homónimas para ayuda del usuario, como en la palabra clave *oración*, tenemos *oracion1* < "gramática" y *oracion2* < "religión".

La base secundaria de colocaciones con valencias sustantivales se encuentra en el *Anexo B.3. Base secundaria de valencias sustantivales*

5.1.4 Descripción de los resultados de la base secundaria de hipónimos / hiperónimos

En el siguiente cuadro se muestra un segmento de la base secundaria de hipónimos e hiperónimos caracterizados morfológicamente.

+1212acción
-1112crimen
-1111funcionamiento
-1111movimiento
+1121adornos
-1121anillos
-1121aretes
-1122collares
-1121pendientes
+1122animales
-0000animales domésticos
-1121felinos
-1122reptiles
-1122roedores
+0000animales domésticos
-1111cerdo
-1111gato
-1111pájaro
-1111perro
+1121antibióticos
-1211eritromicina
-1211estreptomicina
-1211penicilina
+1111aparato1
-1211aspiradora
-1112iluminador
-1211lavadora
-1112refrigerador
-1112tostador
-1112ventilador

+1221bebidas
-0000bebida caliente
-0000bebida fuerte
-1121refrescos
+0000bebida caliente
-1111cacao
-1111café
-1111chocolate
-1111té
+0000bebida fuerte
-1112licor
-1112ron
-1111vodka
-1112whisky
+1121refrescos
-1211Coca-cola
-1211Fanta
-1211Mirinda
-1211Pepsi Cola

La primera columna indica si es hiperónimo o hipónimo la palabra corriente. Las cuatro columnas siguientes (2-5) definen los rasgos morfológicos de la palabra. Por ejemplo, para el artículo de abajo se tiene la siguiente interpretación:

+1121adornos
 -1121anillos
 -1121aretes
 -1122collares

adornos es sustantivo (1), masculino (1), plural (2), desinencia clase 1 (*adorno*∅ / *adornos*) y es hiperónimo de *anillos*, *aretes* y *collares*.

Se tienen etiquetas de clasificación de hipónimos. Por ejemplo, el hiperónimo (superclase) *animales* contiene los hipónimos (subclases) *animales domésticos*, *felinos*, *reptiles*, *roedores*. Donde se indica que *animales domésticos* es una clasificación (POS es cero). Por lo tanto, esta clasificación debe ser hiperónimo de otros conceptos, como es el caso; los hipónimos para la etiqueta de clasificación ‘*animales domésticos*’ son *cerdo*, *gato*, *pájaro*, *perro*.

No sólo se jerarquiza para las etiquetas de clasificación. También se jerarquiza para conceptos que pueden ser parte de una colocación como es el caso de la clasificación *refrescos* para la que se obtienen sus rasgos morfológicos y tiene los hipónimos *Coca-cola*, *Fanta*, *Mirinda* y *Pepsi Cola*.

La base secundaria de hipónimos e hiperónimos se encuentra de forma extensa en el *Anexo B.4. Base secundaria de hipónimos / hiperónimos*.

5.1.5 Protocolo de errores para la caracterización morfológica

Al caracterizar morfológicamente las colocaciones se tiene un control de los errores. Se cuenta con un protocolo por medio del cual se obtiene un archivo de texto con los errores en el formateo. En este protocolo se explican los errores que se produjeron al formatear las colocaciones. Por ejemplo, los errores producidos en una ejecución del programa fueron los siguientes:

```

Error: COMBINACIÓN SUSTANTIVO - ADJETIVO
Artículo: 3
Palabra clave: accidente
Palabra relacionada: gravemente
POS clave: Sus. POS relacionada: Adv.

Error: CONCORDANCIA EN GÉNERO (Palabra clave y relacionada)
Artículo :39
Palabra clave: alberca
Palabra relacionada: olímpico
Gen clave: Fem. Gen relacionada: Masc.

Error al grabar el artículo: 39
La colocación contiene errores.
LA COLOCACIÓN SE CATALOGÓ:
+1211alberca
-2001 olímpico

Error: CONCORDANCIA EN NÚMERO (Palabra clave y relacionada)
Artículo :40
Palabra clave: aldea
Palabra relacionada: perdidas
Num clave: Sing. Num relacionada: Plur.

Error al grabar el artículo: 40
La colocación contiene errores.
LA COLOCACIÓN SE CATALOGÓ:
+1211aldea
-2001 perdidos

-----
Total de colocaciones leídas:      1259
Total de colocaciones grabadas:    1256
Total de errores: 3
-----

```

El tipo de error se indica. Se pueden generar errores por disconcordancia en género, número o error en la combinación de las palabras esperadas.

Para el error producido se indica el número de artículo en el que ocurrió el error, así como la palabra clave, la palabra relacionada y la categoría que produjo el error. Por ejemplo,

para la palabra *aldea* se muestra el número que es la categoría que tiene discordancia con el número de su palabra relacionada *perdidas*.

Al final, se proporciona un resumen de los errores y de las colocaciones leídas, grabadas y el total de errores producidos.

Si al menos una combinación tiene error en un artículo del diccionario entonces el artículo no se graba en el archivo secundario.

El protocolo de control es la interfaz por medio de la cual sabemos si el formateo de las colocaciones fue exitoso o no. Este control se lleva a cabo para cada una de las conversiones de las bases primarias.

5.2 Sistema CrossLexica-Esp versión beta 1.0

El sistema *CrossLexica-Esp* es la interfaz del usuario en la que se visualizan las colocaciones y los papeles sintácticos y semánticos que tienen cada uno de sus componentes. El usuario puede acceder las combinaciones de palabras por alguno de sus dos componentes.

Las bases secundarias se han compilado y se crearon dos bases de datos: un *diccionario general* y una *base de relaciones entre palabras* (los enlaces entre la palabra clave y sus palabras relacionadas). La base de enlaces contiene todas las relaciones sintácticas y semánticas de cada uno de los componentes de la colocación.

Mostraremos algunos ejemplos con los tipos de relaciones sintácticas y semánticas que se describieron en el capítulo tres.

5.2.1 Funcionamiento del sistema CrossLexica-Esp

El sistema CrossLexica-Esp se muestra en figura 5.1. La petición puede introducirse con desinencias de número o género, el sistema tienen la capacidad de localizar la palabra correcta dentro del diccionario general. Esto sólo se aplica para adjetivos. Para sustantivos ambas formas de palabras (singular y plural) se encuentran en el sistema.

Todas las relaciones encontradas en el sistema se enlistan y se clasifican según el papel sintáctico y semántico.

En las páginas siguientes indicaremos cómo el sistema muestra los tipos de colocaciones que se describieron en esta tesis.

En ventana de la figura 5.1 se puede capturar la palabra que se quiere consultar (1) o seleccionarla de la lista del diccionario general, parte izquierda de la ventana (8). Las combinaciones encontradas para la palabra consultada se muestran en la parte derecha de la ventana (7). Los títulos en negro indican las relaciones sintácticas y semánticas existentes para la palabra consultada (6).

En la parte superior de la ventana se muestra la información morfológica y semántica de la palabra consultada (4 y 5). En este caso, *tono* tiene los rasgos morfológicos: sustantivo, masculino, singular y se refiere a un ambiente familiar o afectivo. Las combinaciones encontradas se presentan en la parte derecha de la ventana (7).

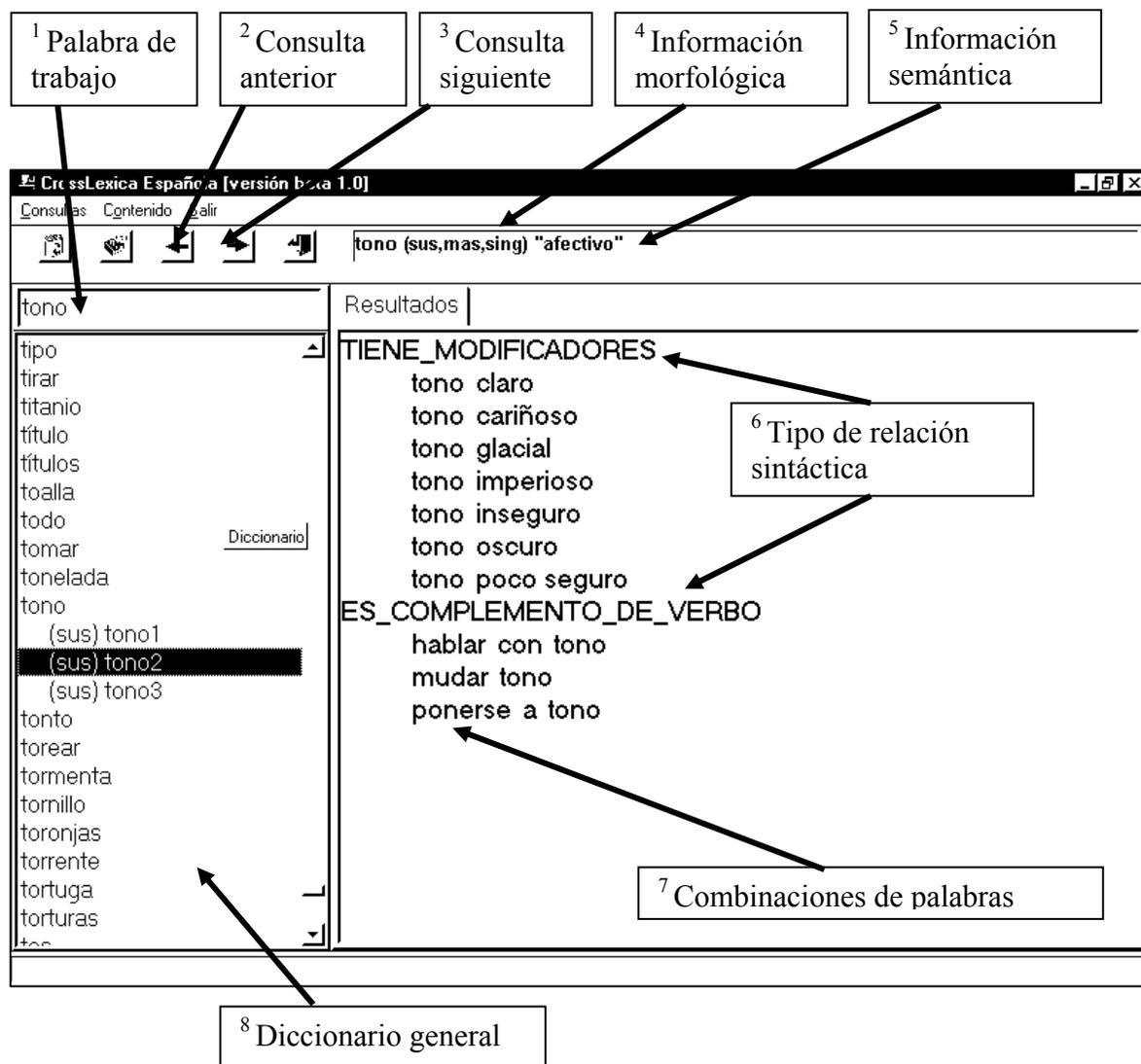


Figura 5.1 Ejemplo de entrada de colocaciones en el sistema CrossLexica-Esp

5.2.2 Ejemplos de colocaciones modificadoras

Como se describió, las colocaciones modificadoras admiten las clases de palabras: sustantivos, adjetivos, verbos y adverbios. En estas colocaciones se muestra cómo son modificadas las palabras o quién las modifica. En los siguientes puntos mostramos algunos ejemplos para estas relaciones sintácticas.

5.2.2.1 Combinaciones sustantivo → adjetivo

Como se describió en el capítulo tres, los modificadores que permiten los sustantivos son únicamente adjetivos o términos multipalabra.

El género y el número de los adjetivos se encuentran en concordancia con el sustantivo que los acompaña y los términos multipalabra son invariables. En la figura 5.2 todos los adjetivos modificadores se encuentran en concordancia con su sustantivo gobernante.

En este caso, se consulta el sustantivo *arma* que tienen los rasgos morfológicos: sustantivo, femenino, singular. Para esta consulta se encuentra la relación sintáctica *tienen modificadores* que son *atómica*, *blanca*, etc.

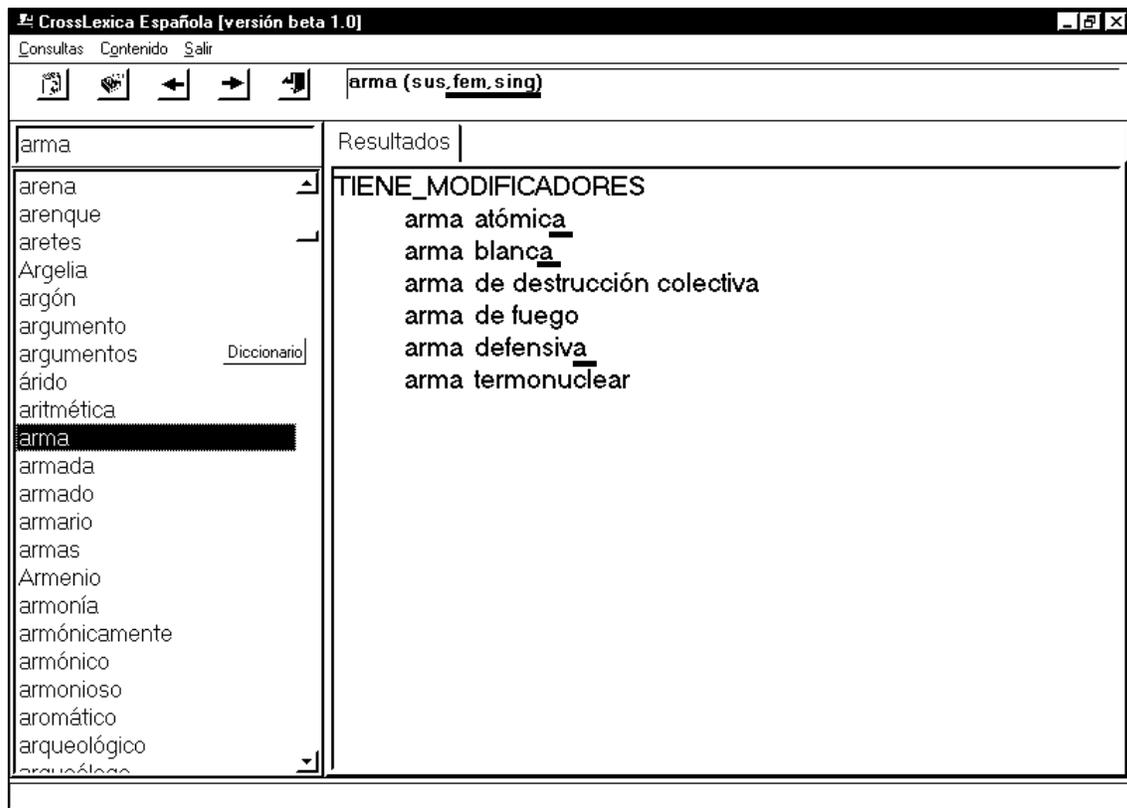


Figura 5.2 Concordancia entre un sustantivo y su adjetivo

De forma inversa, se consulta el adjetivo y se encuentran las palabras que modifica (sustantivos rectores). Los adjetivos se encuentran en concordancia en género y número para cada sustantivo que lo rige. En la figura 5.3 se consulta el adjetivo *blanco* y se encuentran todos los sustantivos para los cuales es modificador (*arma*, *armas* y *verso*).

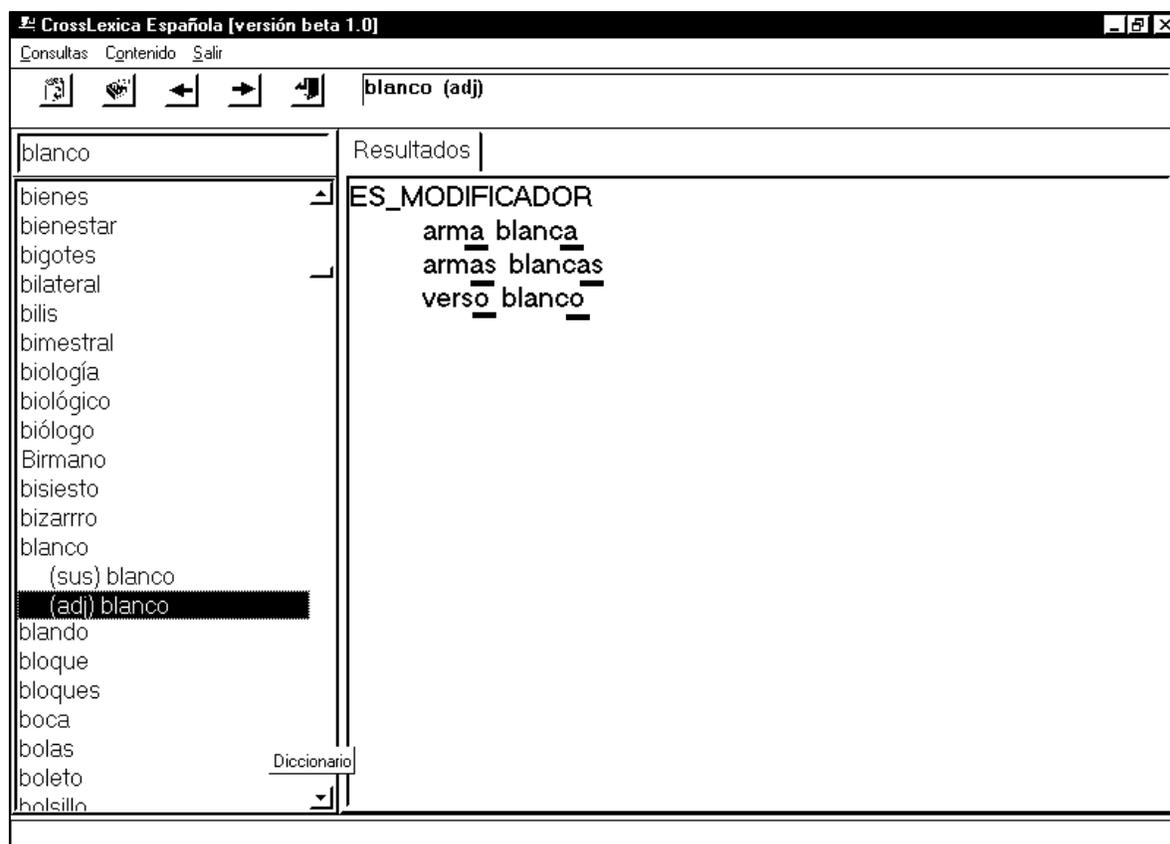


Figura 5.3 Concordancia entre un adjetivo y varios sustantivos

5.2.2.2 Combinaciones adjetivo → adverbio

Se consulta el adjetivo y se encuentran los adverbios o términos multipalabra modificadores para este adjetivo. Por ejemplo, en la figura 5.4 se muestran las palabras modificadoras para el adjetivo *astuto* que son: *muy* y *como un diablo*.

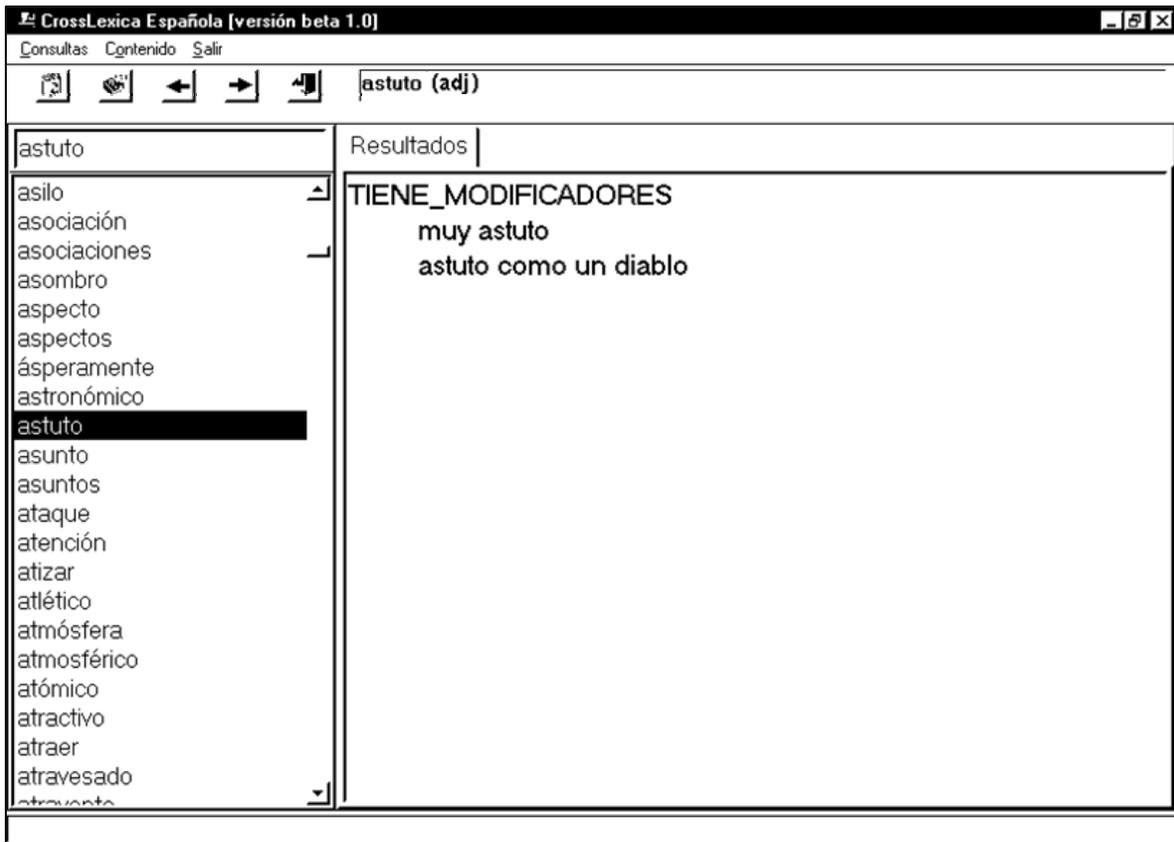


Figura 5.4 Combinación de un adjetivo y sus modificadores adverbiales

Para el otro modo de uso, se consulta el adverbio *bien*, figura 5.5, y se localizan todas las combinaciones para las cuales el adverbio es modificador. Aquí se muestra más de una relación sintáctica para la palabra consultada.

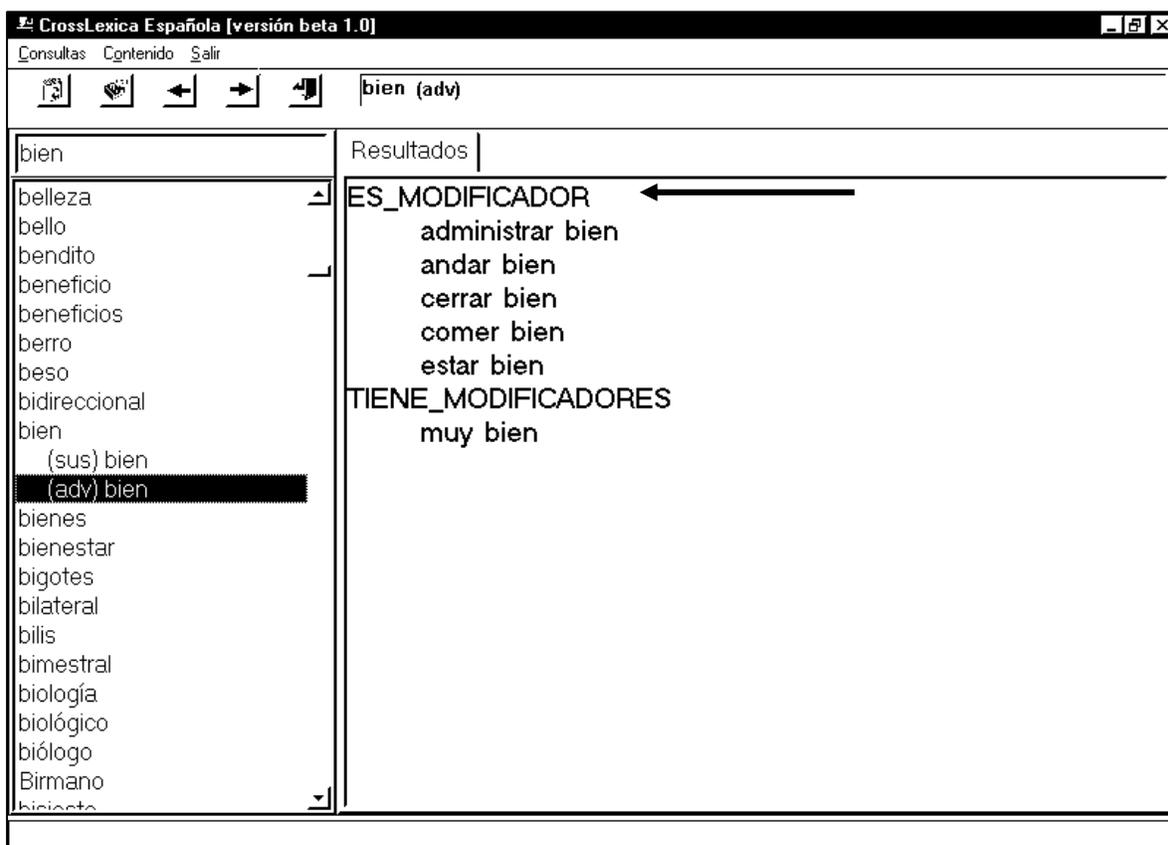


Figura 5.5 Combinación de un adverbio y los verbos que modifica

5.2.2.3 Combinaciones verbo → adverbio

Para estas combinaciones, se consulta el verbo y se encuentran las palabras que lo modifican (sólo adverbios o términos multipalabra). Por ejemplo, en la figura 5.6 se muestran las palabras relacionadas con el verbo *comer*, que son sus modificadores: *bien* y *mucho*. De forma inversa, se pueden consultar los adverbios y encontrar las palabras que modifica, como en las combinaciones adjetivo → adverbio (figura 5.5).

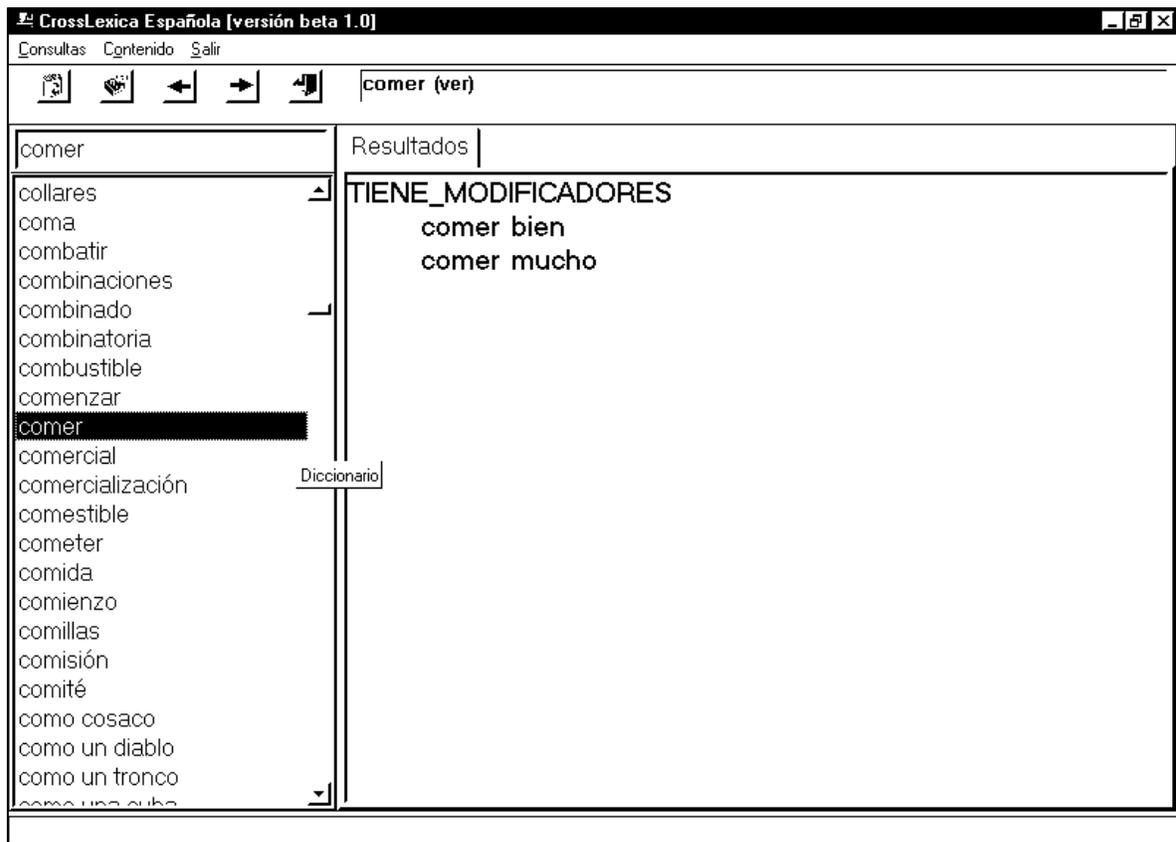


Figura 5.6 Combinación de un verbo y sus adverbios modificadores

5.2.3 Ejemplos de colocaciones con valencias verbales

Para las colocaciones con valencias verbales se muestran los complementos del verbo o sus sujetos, así como las relaciones recíprocas. Mostramos algunos ejemplos para estas relaciones sintácticas.

5.2.3.1 Combinaciones verbo → sustantivo

Para estas combinaciones se muestran dos tipos de relación sintáctica, los complementos que tiene el verbo y los sujetos que tiene el verbo. En la figura 5.7, el verbo *abrir* se encuentra conjugado; esta construcción es un tipo de oración de pasiva refleja. Véase el punto 3.2.2.

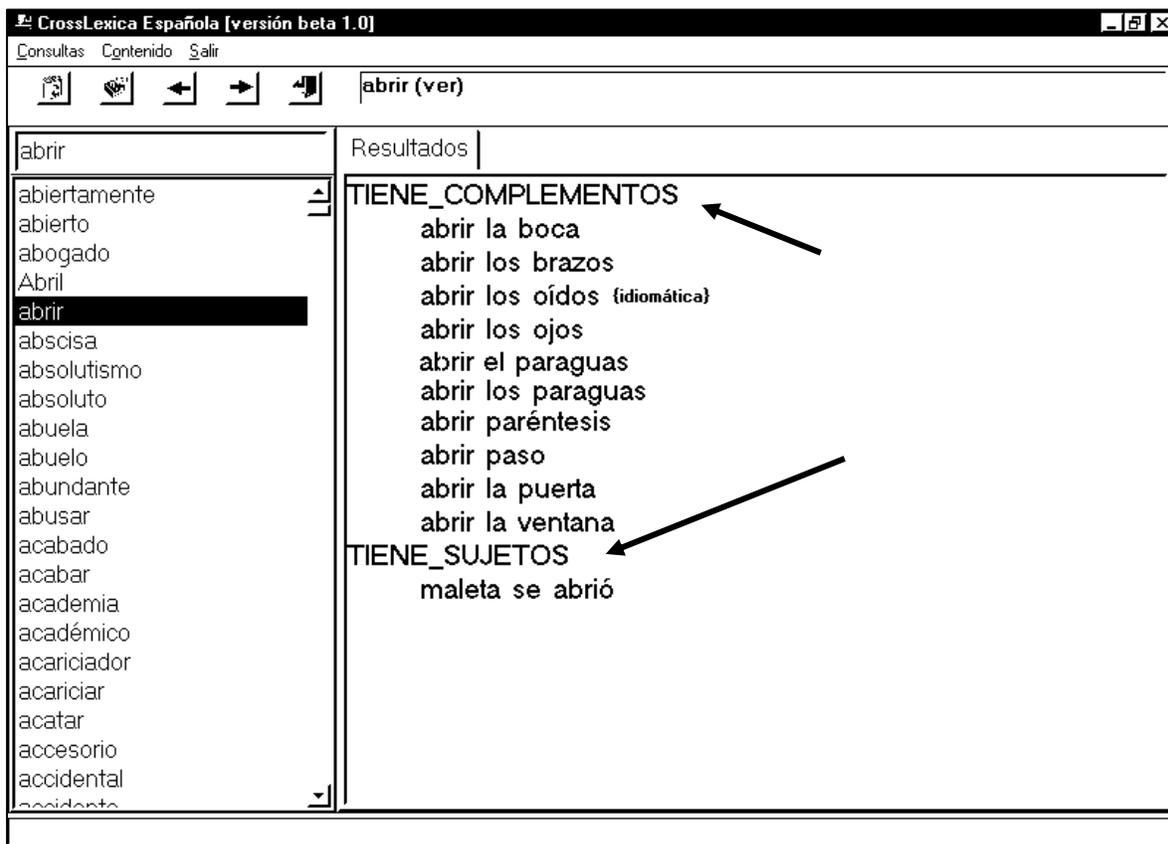


Figura 5.7 Combinación de un verbo con sus complementos y sujetos

De forma inversa, en la figura 5.8, la palabra consultada *visita* que tiene los rasgos morfológicos: sustantivo, femenino, singular. Ésta es un sustantivo complementario para el verbo *hacer*. Además, se muestran los modificadores que tienen la palabra consultada.

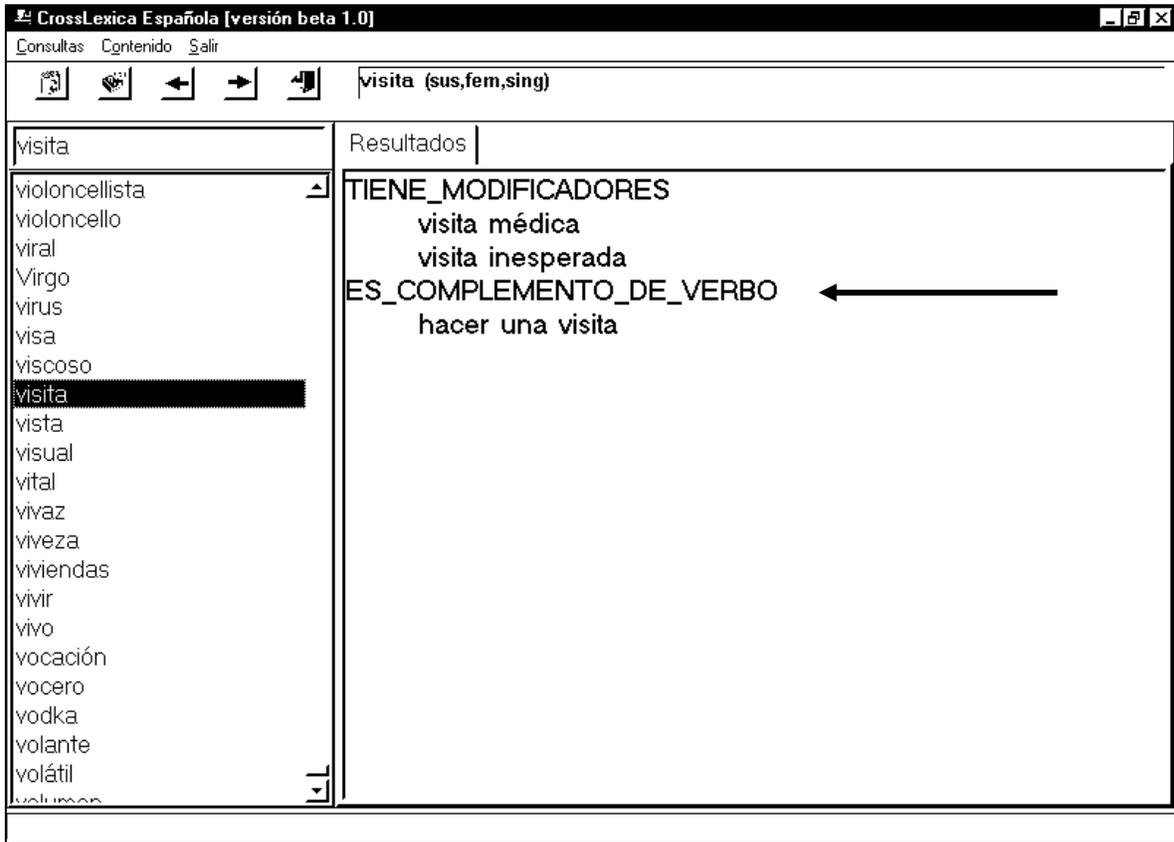


Figura 5.8 Sustantivo como complemento verbal

5.2.4 Ejemplos de colocaciones con valencias sustantivales

Para las colocaciones con valencias sustantivales se muestran los complementos del sustantivo y su recíproco. Véase el punto 3.2.4.

5.2.4.1 Combinaciones sustantivo → sustantivo

En la figura 5.9 se muestran las relaciones sintácticas que se mencionaron anteriormente y la nueva relación *es complemento de sustantivo*. En la ventana siguiente, se muestran los complementos sustantivales para *naturaleza* con rasgos morfológicos: sustantivo, femenino, singular.

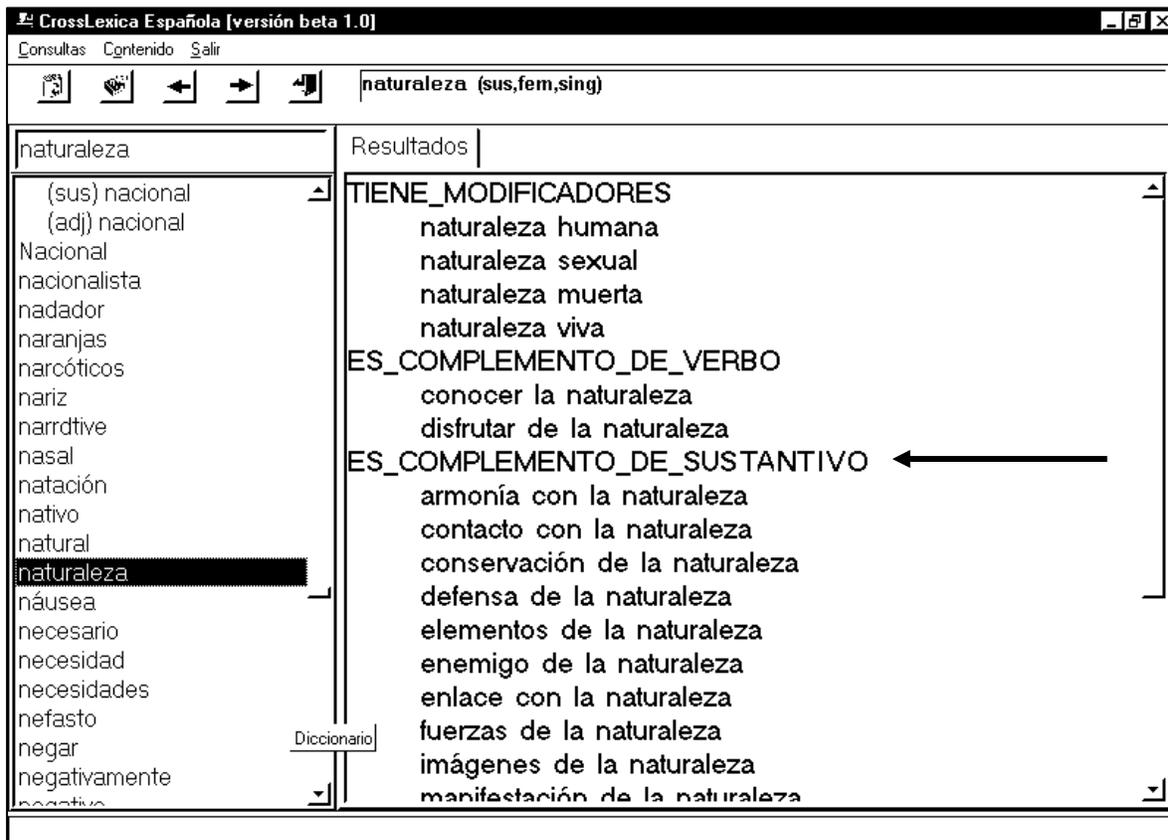


Figura 5.9 Sustantivo como complemento sustantival

De forma inversa se puede consultar el sustantivo de la izquierda, de *naturaleza*, obteniéndose la relación recíproca *tiene complementos*.

5.2.5 Ejemplos de combinaciones de palabras con información semántica

La información para las diferentes acepciones, proporciona la noción en qué situaciones podemos usar la combinación de palabras. En el primer caso, figura 5.10, las combinaciones hacen referencia al ambiente musical. La palabra *tono* tiene los rasgos morfológicos: sustantivo, masculino, singular.

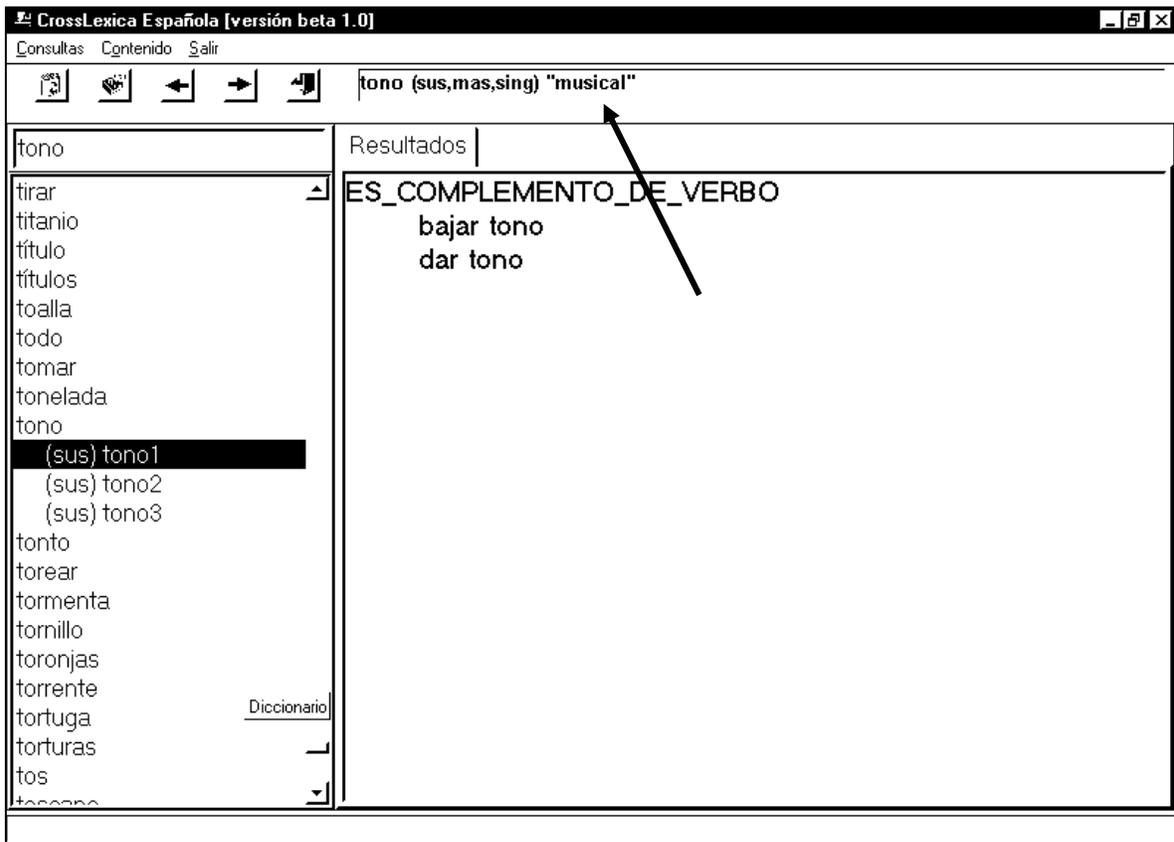


Figura 5.10 Información adicional para palabras homónimas (primer caso)

En el segundo caso, figura 5.11, las combinaciones hacen referencia a un ambiente amistoso o familiar.

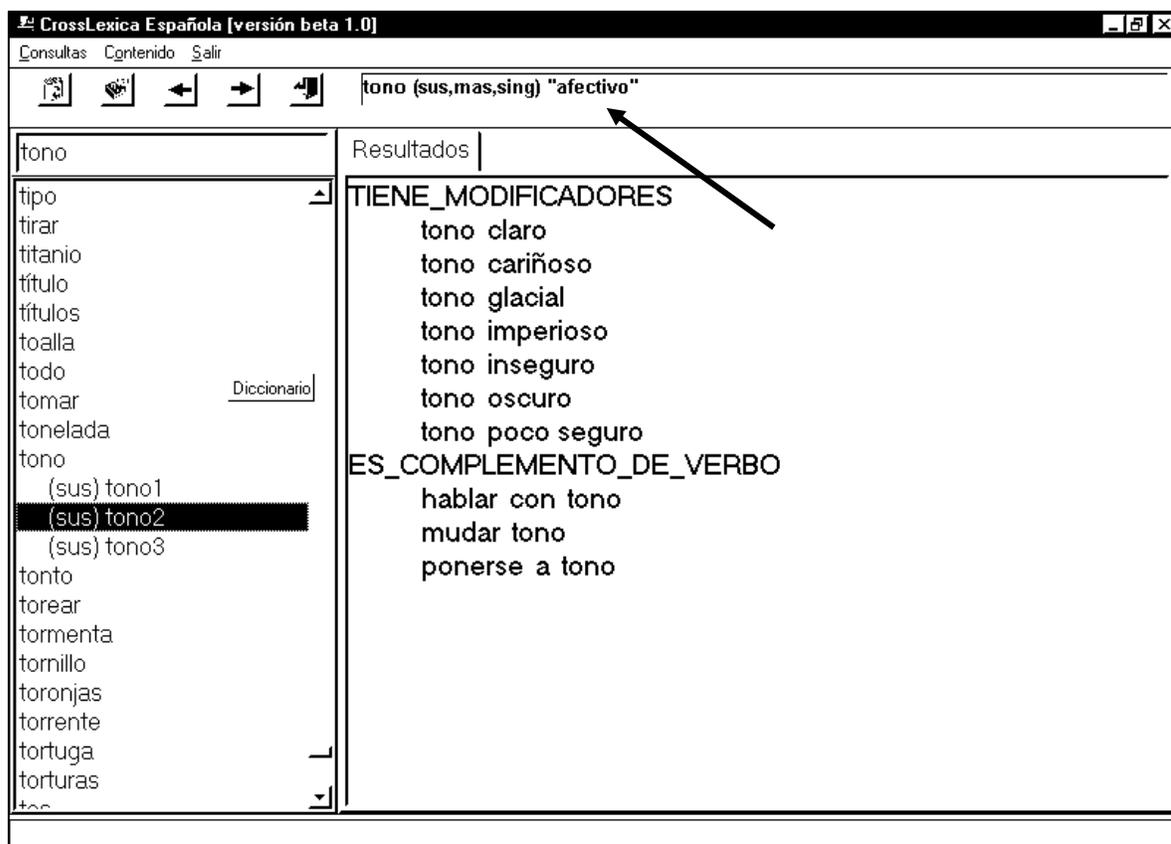


Figura 5.11 Información adicional para palabras homónimas (segundo caso)

En el tercer caso, figura 5.12, las combinaciones hacen referencia al cuerpo.

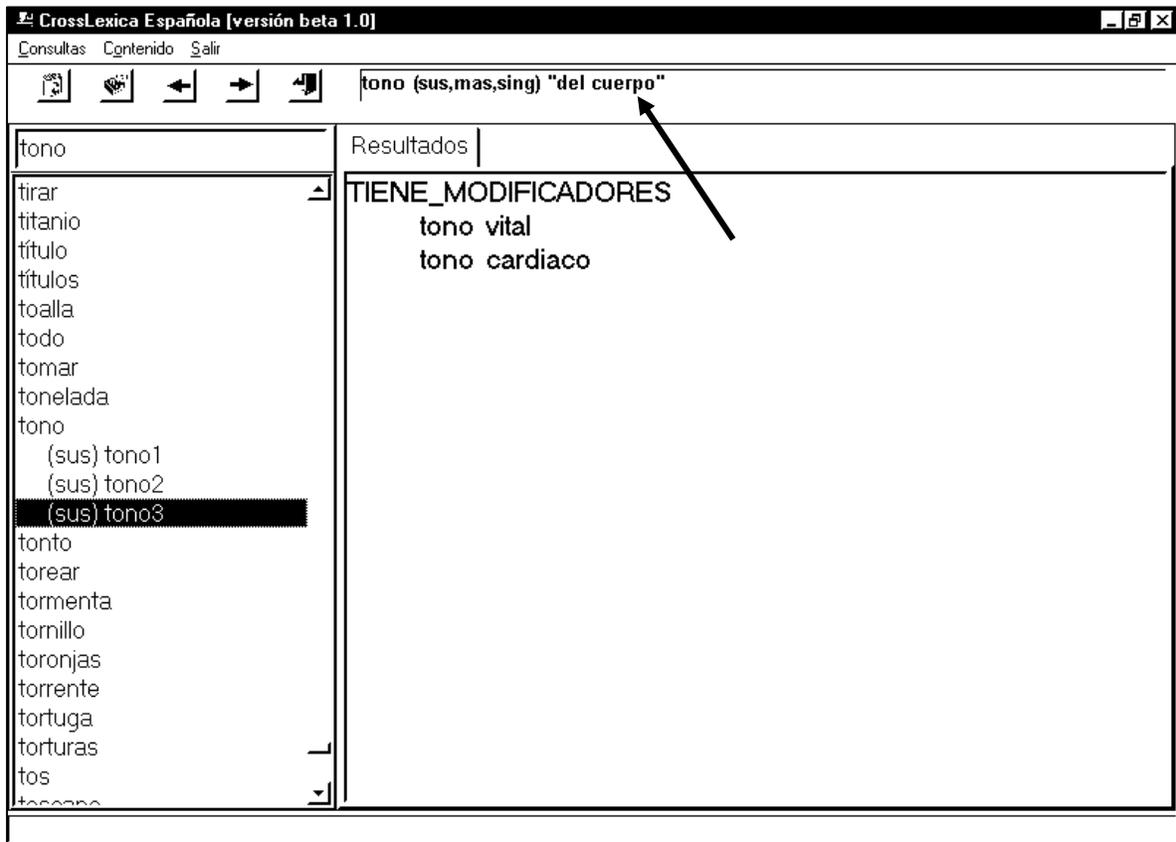
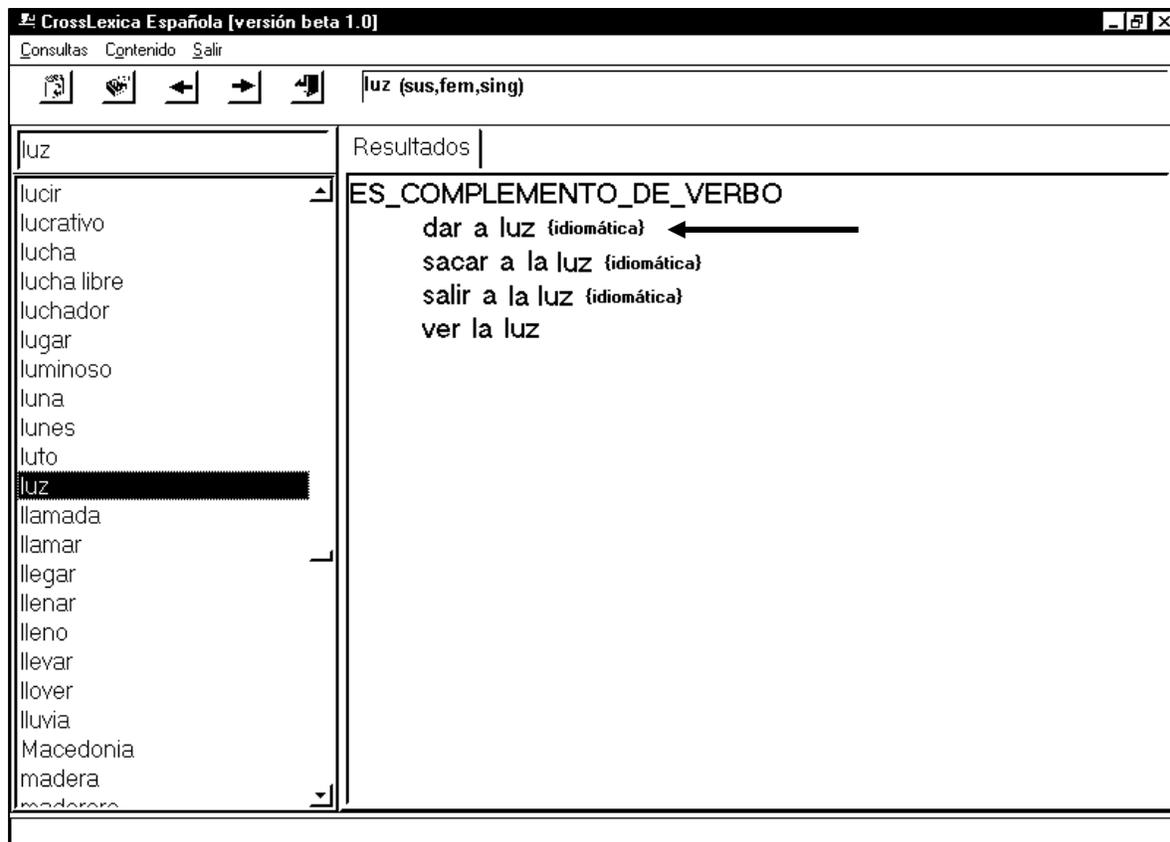


Figura 5.12 Información adicional para palabras homónimas (tercer caso)

5.2.6 Ejemplos de combinación de palabras con idiomatización y estilo

El uso de las combinaciones de palabras de forma idiomática se muestra con una etiqueta al final de la combinación, figura 5.13. Ésta nos indica si la combinación es idiomática o se-



mi-idiomática. La ausencia de etiqueta indica que es una combinación neutral.

Figura 5.13 Combinaciones con información idiomática

El estilo de uso de la combinación se muestra al final de la combinación, figura 5.14. En este caso, *caballos de fuerza* se usa generalmente en textos científicos o técnicos. La ausencia de etiqueta indica que es una combinación neutral.

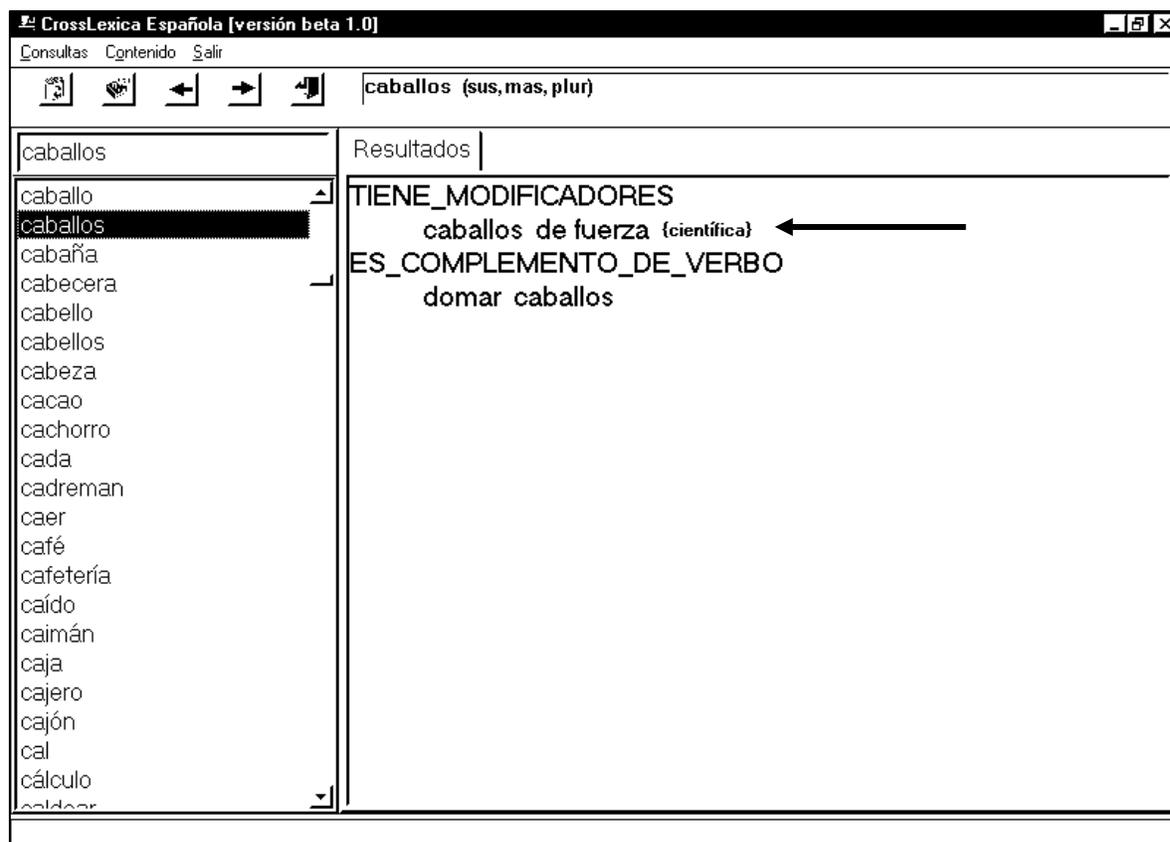


Figura 5.14 Combinaciones con información estilística

5.2.7 Ejemplos de relaciones semánticas

Las relaciones semánticas que presenta el sistema es una jerarquía de conceptos como se ve en la figura 5.15. *Flores* es hiperónimo para *azucenas*, *callas*, *dalias*, etc. y el concepto *flora* es hiperónimo de *flores*. Los conceptos se muestran jerarquizados a partir del concepto consultado y se divide en dos partes: los hipónimos escalonados hacia abajo y los hiperónimos escalonados hacia arriba.

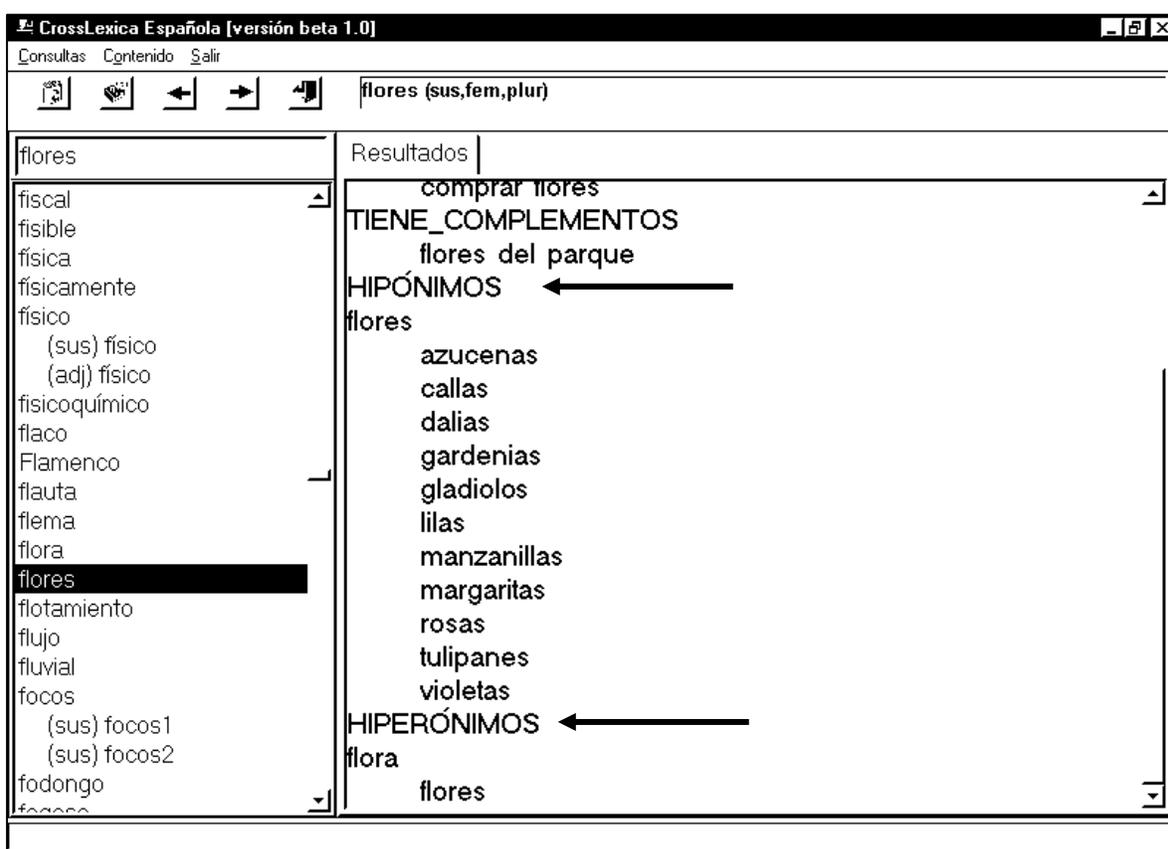


Figura 5.15 Combinaciones con relaciones semánticas

5.2.8 Inferencia de colocaciones

El sistema es capaz de inferir colocaciones que no están explícitamente en la base de colocaciones. Esto se logra a través de la información semántica de hipónimos / hiperónimos, la clase a la cual pertenece una palabra. Por ejemplo, para la palabra *flores*, se expresan las siguientes combinaciones *flores aromáticas*, *flores hermosas*, *flores exóticas*. Sabiendo que *azucenas* ES_UN tipo de *flor*, podemos inferir tales colocaciones para *azucenas*: *azucenas aromáticas*, *azucenas hermosas*, *azucenas exóticas*, y para todas aquellas palabras que pertenezcan a la clase.

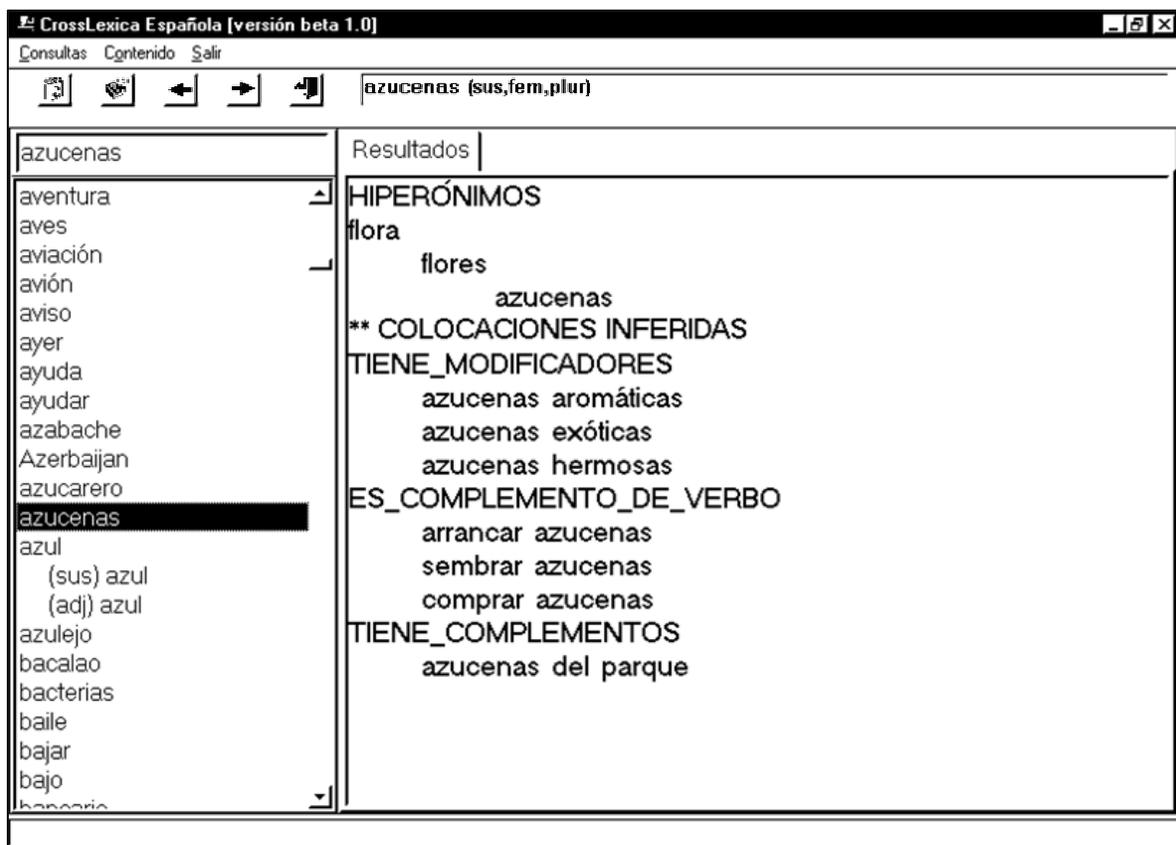


Figura 5.16 Inferencia de colocaciones para *azucenas*

Para *tulipanes*, otra subclase de *flores*, tenemos las combinaciones generadas de la figura 5.17.

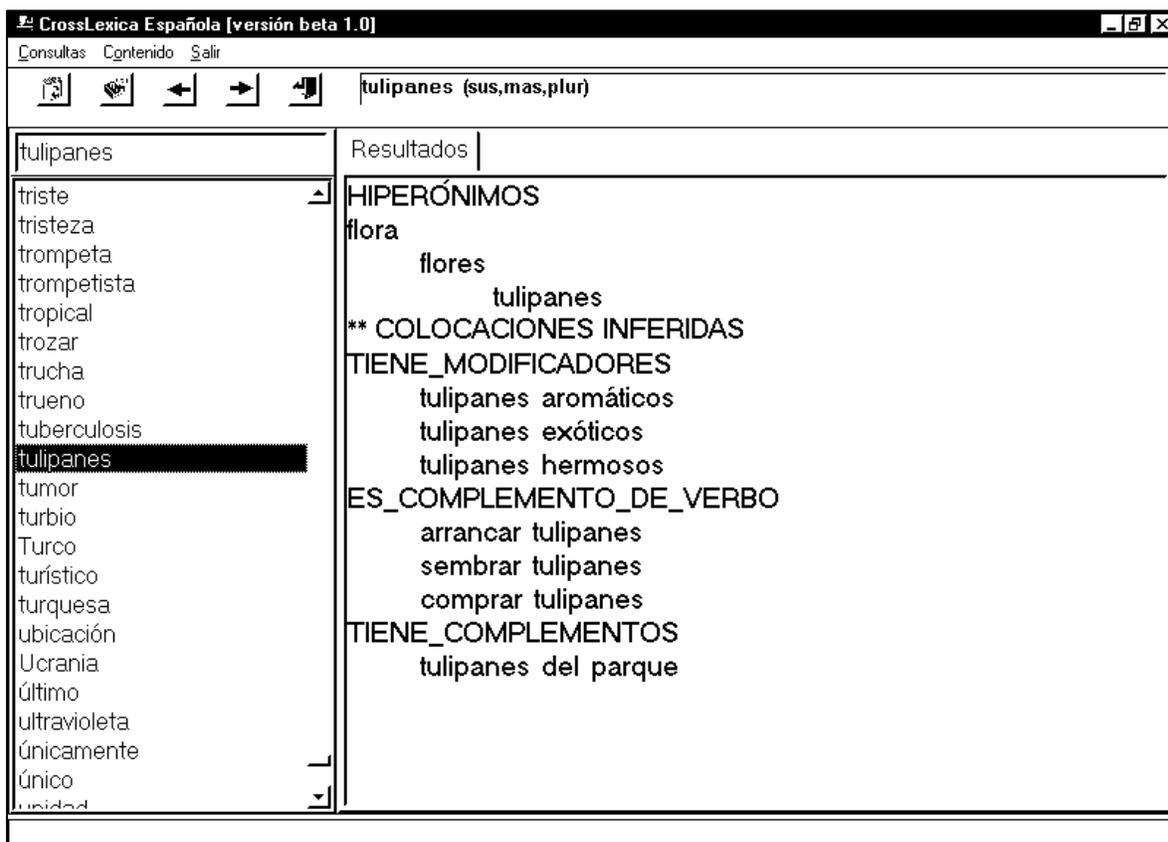


Figura 5.17 Inferencia de colocaciones para *tulipanes*

Las combinaciones de palabras que son completamente fijas, es decir, las combinaciones idiomáticas, no se puede realizar la inferencia con ellas. Las combinaciones idiomáticas únicamente pueden tener esos componentes que lo constituyen y no otros, es decir, no se pueden sustituir sus elementos.

El proceso de inferencia no garantiza que las combinaciones generadas sean usadas muy frecuentemente por las personas. Esto es solamente una forma de enriquecer la base de colocaciones.

En la figura 5.18 se muestran las relaciones sintácticas y semánticas para *bebidas*. Se muestran todos los hipónimos y los niveles jerárquicos que representan entre ellos. Para la palabra *Coca-cola* ES_UN tipo de *refresco* y a su vez *refresco* ES_UN tipo de *bebida*. Las combinaciones encontradas en el proceso de inferencia son sustituidas por su correspondiente hipónimo o hiperónimo para formar las nuevas combinaciones como se muestra en la figura 5.19.

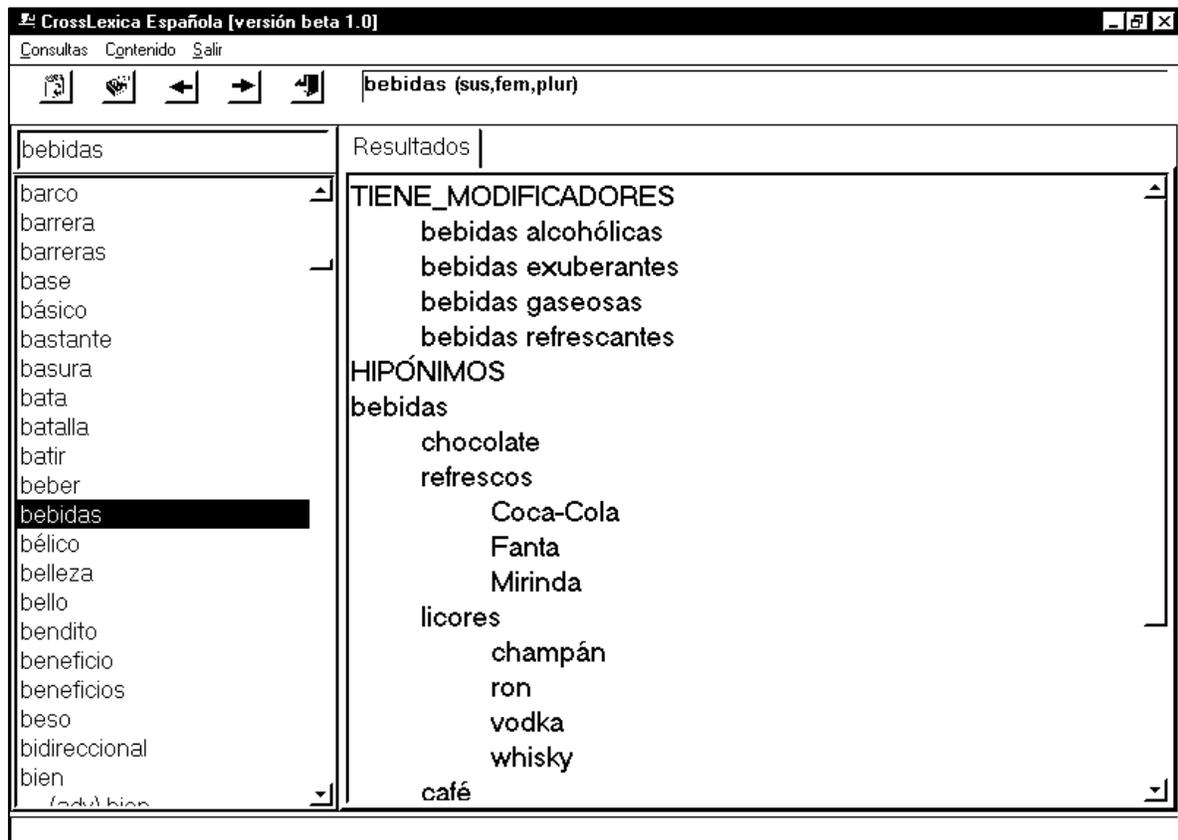


Figura 5.18 Relaciones sintácticas y semánticas para *bebidas*

En la figura 5.19 se muestra la inferencia de nuevas colocaciones para *Coca-Cola*. Se ve que las colocaciones fueron generadas a través de sus hiperónimos *bebidas* y *refrescos*.

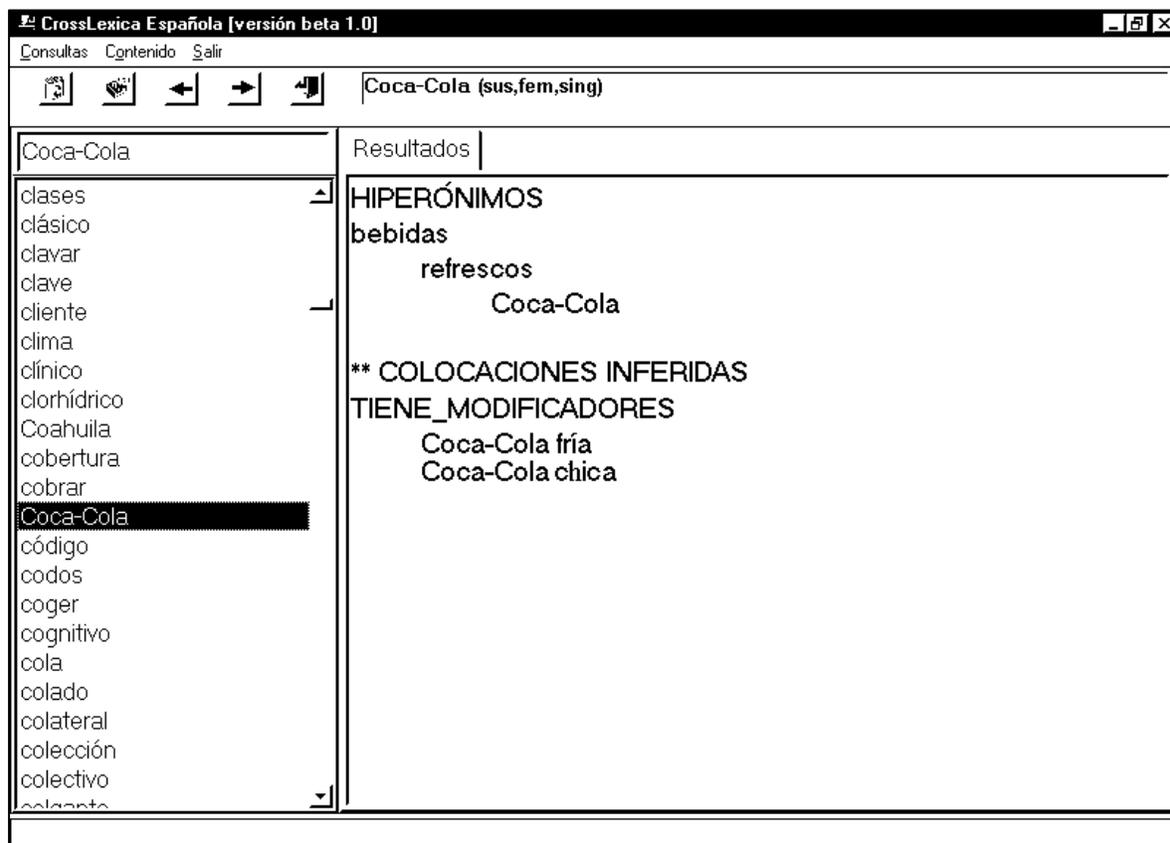


Figura 5.19 Inferencia de colocaciones para *Coca-Cola*

Capítulo 6. Conclusiones

En este capítulo se presentan las conclusiones de la tesis, se enumeran las contribuciones y se proporcionan algunas sugerencias para trabajo futuro.

En esta tesis se demostró que para el aprendizaje informatizado de un idioma extranjero, así como para resolver algunas tareas de lingüística computacional: desambiguación de sentidos de palabras, traducción automática, analizadores sintácticos, etc., es necesario el conocimiento de las colocaciones.

Para español no existía ninguna base de colocaciones. Con el presente trabajo, ya se tiene una base de más de 10 millares de colocaciones que fueron recopiladas de forma manual. Se tiene un promedio de 4.5 palabras relacionadas por palabra clave.

Se desarrollaron las herramientas necesarias para el formateo de los tres tipos de colocaciones con relaciones sintácticas (modificadores, valencias verbales y valencias sustantivales) y para la relación semántica de hipónimos / hiperónimos.

Las herramientas de formateo son esenciales para la obtención de la información morfológica para cada componente de la colocación. Se demostró que la morfología de las palabras es necesaria para la correcta presentación de las combinaciones, esto se llevó acabo haciendo uso de la concordancia sintáctica.

Se proporcionaron las tablas de desinencias y archivos auxiliares (preposiciones, adverbios, excepciones a las reglas generales propuestas, etc.) necesarios para la determinación morfológica de las colocaciones.

Se compilaron las colocaciones formateadas y se generaron dos bases de datos. Una base de datos que contiene los componentes de las colocaciones (el diccionario general) y otra base que contiene los enlaces sintácticos y semánticos entre los componentes de la colocación (base de enlaces entre palabras).

Se desarrolló la interfaz del sistema *CrossLexica-Esp* para la demostración de las colocaciones y los papeles sintácticos y semánticos que desempeñan las palabras en las combinaciones.

Se demostró que la inferencia es una característica más del sistema, por medio de la cual se generan colocaciones que no han sido explícitamente capturadas. Sin embargo, estas colo-

caciones generadas no son absolutamente confiables, es decir, que se usen con alta frecuencia por una persona nativa del lenguaje; en general, las colocaciones generadas son buenas. Las colocaciones generadas mediante la inferencia sólo están disponibles en tiempo de ejecución del sistema.

6.1 Contribuciones

Con base a los resultados obtenidos, podemos considerar que las principales contribuciones de la presente tesis son las siguientes:

1. Las herramientas que transforman las colocaciones a una forma secundaria con los rasgos morfológicos y semánticos de cada uno de sus componentes.
2. El sistema que maneja las colocaciones y la morfología para generar las combinaciones de palabras en concordancia correcta. Además de la inferencia de nuevas colocaciones con base a las clases y subclases de las palabras.
3. Un esquema para español que representa los enlaces sintácticos y semánticos entre las combinaciones de palabras.
4. Un diccionario pequeño de colocaciones para la demostración del funcionamiento de la interfaz y de las herramientas de formateo.

6.2 Publicaciones

Miranda-Jiménez S., Bolshakov I. *Base de datos de un diccionario combinatorio para Español: Utilitaria para modificadores*. Memorias del Congreso Internacional de Computación CIC'2002, Vol. 2, México, pp 267-278.

6.3 Trabajo futuro

Dentro de las tareas futuras consideramos que es importante agregar otras relaciones semánticas como son los sinónimos y antónimos. Por medio de los cuales la propiedad de inferencia enriquecería aún más nuestra base de colocaciones. Por ejemplo, haciendo uso de los sinónimos más cercanos para realizar la inferencia. El verbo *mover* dentro de sus sinónimos se encuentra *agitar*. En combinaciones como *mover la mano*, *mover la cabeza* y *mover la bandera* se inferirían nuevas combinaciones sustituyendo el verbo y obteniéndose combinaciones como: *agitar la mano*, *agitar la cabeza* y *agitar la bandera*.

Agregar otras relaciones sintácticas también importantes para español tales como *verbo* → *verbo: comenzar* (→ *a*) → *trabajar*, *terminar* (→ *de*) → *trabajar*, *saber* → *ver*; pares coordinados de palabras como: (*a*) *diestra* y *siniestra*, (*a*) *pan* y *agua*, *ahora* y *siempre*; y algunas otras.

También se puede vincular a un diccionario bilingüe (por ejemplo, inglés-español), para traducir las colocaciones a sus equivalentes ingleses, al tomar en consideración las bases de colocaciones para inglés.

Se puede vincular el diccionario de colocaciones con EuroWordNet español, que se basa en su antecesor WordNet [15], para obtener las relaciones de sinonimia, antonimia, y *parte_de* (las partes que tienen un concepto como *árbol* que tiene *raíz*, *hojas*, *ramas*, etc.) y dar posiblemente explicaciones de la palabra consultada con base a los componentes semánticos en que se estructura WordNet.

Anexos

Anexo A. Bases primarias de colocaciones

Anexo A.1 Base primaria de colocaciones modificadoras

<i>absolutismo</i>	<i>combinada</i>
<i>ilustre</i>	<i>de gracias</i>
<i>impuesto</i>	<i>democrática</i>
<i>francés</i>	<i>deshonesta</i>
<i>provincial</i>	<i>destructora</i>
<i>real</i>	<i>dramática</i>
<i>acceso</i>	<i>directa</i>
<i>aleatorio</i>	<i>educativa</i>
<i>directo</i>	<i>ejecutada</i>
<i>exclusivo</i>	<i>física</i>
<i>eficaz</i>	<i>gubernamental</i>
<i>fácil</i>	<i>humana</i>
<i>gratuito</i>	<i>inmediata</i>
<i>libre</i>	<i>judicial</i>
<i>permitido</i>	<i>jurídica</i>
<i>público</i>	<i>libre</i>
<i>principal</i>	<i>militar</i>
<i>privado</i>	<i>nacional</i>
<i>prohibido</i>	<i>nacionalista</i>
<i>restringido</i>	<i>natural</i>
<i>rápido</i>	<i>penal</i>
<i>accidente</i>	<i>popular</i>
<i>aéreo</i>	<i>perfecta</i>
<i>casual</i>	<i>pública</i>
<i>cruel</i>	<i>prometida</i>
<i>estúpido</i>	<i>social</i>
<i>fatal</i>	<i>verbal</i>
<i>geográfico</i>	<i>acciones</i>
<i>general</i>	<i>arbitrarias</i>
<i>grave</i>	<i>correctivas</i>
<i>humano</i>	<i>ordinarias</i>
<i>industrial</i>	<i>concretas</i>
<i>involuntario</i>	<i>bélicas</i>
<i>mortal</i>	<i>políticas</i>
<i>natural</i>	<i>violentas</i>
<i>peligroso</i>	<i>militares</i>
<i>penoso</i>	<i>específicas</i>
<i>terrible</i>	<i>individuales</i>
<i>serio</i>	<i>penales</i>
<i>acción</i>	<i>puntuales</i>
<i>armada</i>	<i>minerías</i>
<i>buena</i>	<i>heroicas</i>

ANEXO A.1 BASE PRIMARIA DE COLOCACIONES MODIFICADORAS

<i>encaminadas</i>	<i>cianhídrico</i>
<i>necesarias</i>	<i>cítrico</i>
<i>previstas</i>	<i>clorhídrico</i>
<i>indignas</i>	<i>fénico</i>
<i>simples</i>	<i>láctico</i>
<i>inútiles</i>	<i>nítrico</i>
<i>positivas</i>	<i>sulfúrico</i>
<i>unilaterales</i>	<i>acoger</i>
<i>especiales</i>	<i>con brazos abiertos</i>
<i>implícitas</i>	<i>favorablemente</i>
<i>urbanas</i>	<i>acostarse</i>
<i>delictivas</i>	<i>terriblemente</i>
<i>irregulares</i>	<i>actitud</i>
<i>constructivas</i>	<i>crítica</i>
<i>aceite</i>	<i>agresiva</i>
<i>de anís</i>	<i>benévola</i>
<i>de bacalao</i>	<i>expectante</i>
<i>de cañadón</i>	<i>expectativa</i>
<i>de girasol</i>	<i>negligente</i>
<i>de hígado</i>	<i>negativa</i>
<i>de linaza</i>	<i>pacífica</i>
<i>de maíz</i>	<i>política</i>
<i>de máquinas</i>	<i>positiva</i>
<i>de oliva</i>	<i>hostil</i>
<i>de ricino</i>	<i>heroica</i>
<i>quemado</i>	<i>actividad</i>
<i>esencial</i>	<i>bélica</i>
<i>comestible</i>	<i>bulliciosa</i>
<i>producido</i>	<i>científica</i>
<i>barato</i>	<i>deportiva</i>
<i>vegetal</i>	<i>económica</i>
<i>puro</i>	<i>física</i>
<i>mineral</i>	<i>mecánica</i>
<i>vegetal</i>	<i>tecnológica</i>
<i>aceites</i>	<i>vital</i>
<i>pesados</i>	<i>ganadera</i>
<i>volátiles</i>	<i>política</i>
<i>utilizados</i>	<i>humana</i>
<i>industriales</i>	<i>comercial</i>
<i>acero</i>	<i>industrial</i>
<i>colado</i>	<i>cultural</i>
<i>cromado</i>	<i>artística</i>
<i>inoxidable</i>	<i>profesional</i>
<i>martín</i>	<i>laboral</i>
<i>templado</i>	<i>agrícola</i>
<i>pulido</i>	<i>musical</i>
<i>caliente</i>	<i>productiva</i>
<i>forjado</i>	<i>militar</i>
<i>recubierto</i>	<i>sexual</i>
<i>ácido</i>	<i>cerebral</i>
<i>arsénico</i>	<i>biológica</i>
<i>bórico</i>	<i>petrolera</i>
<i>carbónico</i>	<i>mental</i>

ANEXO A.1 BASE PRIMARIA DE COLOCACIONES MODIFICADORAS

actividades
científicas
deportivas
económicas
manufactureras
tecnológicas
culturales
comerciales
políticas
agrícolas
humanas
sociales
productivas
educativas
musicales
religiosas
ilícitas
paralelas
terroristas
importantes
laborales
financieras
recreativas
públicas
caritativas
concretas
programadas
principales
empresariales
gubernamentales
conjuntas
bancarias
estudiantiles
inmediatas

acto
a la bandera
brutal
cívico
de violencia
delictivo
grandote
proselitista
público
reclamado
religioso
sexual
seguido
único
segundo
primero
solemne
social
político

oficial
central
realizado
terrorista
ilícito
importante
relevante
precioso
ilegal
violento

actos
delictivos
públicos
reclamados
religiosos
conjuntos
humanos
culturales
terroristas
oficiales
judiciales
centrales
violentos
musicales
contrarios
formales
criminales
aislados
heroicos
incalificables
inconscientes
indebidos
perversos
prohibidos

acuerdo
bilateral
común
condicional
eventual
final
parlamentario
político
zonal
general
comercial
nacional
firmado
tácito
logrado
internacional
mutuo
definitivo
global

ANEXO A.1 BASE PRIMARIA DE COLOCACIONES MODIFICADORAS

<i>preliminar</i>	<i>labrados</i>
<i>completo</i>	<i>ridículos</i>
<i>secreto</i>	<i>superfluos</i>
<i>formal</i>	<i>importados</i>
<i>verbal</i>	<i>grabados</i>
<i>matrimonial</i>	<i>pequeños</i>
<i>legal</i>	<i>inútiles</i>
<i>parcial</i>	<i>preciosos</i>
<i>honorable</i>	<i>bellos</i>
<i>adelantos</i>	<i>florales</i>
<i>científicos</i>	<i>orientales</i>
<i>importantes</i>	<i>advertencia</i>
<i>serios</i>	<i>oportuna</i>
<i>tecnológicos</i>	<i>primera</i>
<i>técnicos</i>	<i>última</i>
<i>logrados</i>	<i>seria</i>
<i>médicos</i>	<i>previa</i>
<i>adiestramiento</i>	<i>afectar</i>
<i>básico</i>	<i>sustancialmente</i>
<i>canino</i>	<i>directamente</i>
<i>militar</i>	<i>negativamente</i>
<i>profesional</i>	<i>indirectamente</i>
<i>administración1</i>	<i>físicamente</i>
<i>central</i>	<i>gravemente</i>
<i>de empresas</i>	<i>agente1<hombre</i>
<i>doméstica</i>	<i>británico</i>
<i>estatal</i>	<i>de bolsa</i>
<i>federal</i>	<i>de negocios</i>
<i>local</i>	<i>de policía</i>
<i>municipal</i>	<i>del servicio de contraespionaje</i>
<i>pública</i>	<i>invasor</i>
<i>territorial</i>	<i>secreto</i>
<i>centralizada</i>	<i>encubierto</i>
<i>general</i>	<i>fiscal</i>
<i>administración2</i>	<i>oficial</i>
<i>de justicia</i>	<i>infiltrado</i>
<i>administrar</i>	<i>judicial</i>
<i>bien</i>	<i>político</i>
<i>adecuadamente</i>	<i>agente2<sustancia</i>
<i>libremente</i>	<i>dañino</i>
<i>independientemente</i>	<i>de limpieza</i>
<i>admiración</i>	<i>patógeno</i>
<i>gran</i>	<i>químico</i>
<i>profunda</i>	<i>viral</i>
<i>pública</i>	<i>infeccioso</i>
<i>ciega</i>	<i>erosivo</i>
<i>absoluta</i>	<i>extraño</i>
<i>amorosa</i>	<i>tóxico</i>
<i>extraña</i>	<i>venenoso</i>
<i>adornos</i>	<i>agotamiento</i>
<i>personales</i>	<i>físico</i>
<i>corporales</i>	<i>mental</i>
<i>navideños</i>	<i>agradecido</i>

ANEXO A.1 BASE PRIMARIA DE COLOCACIONES MODIFICADORAS

muy	pluviales
poco	residuales
sumamente	subterráneas
eternamente	termales
agua	territoriales
carbónica	profundas
clara	superficiales
corriente	interiores
cristalina	corrientes
& de cepas	estancadas
& de colonia	salobres
de frutas	salvajes
de mar	internacionales
de mesa	saladas
de purga	sucias
de régimen	jurisdiccionales
de roca	tropicales
dulce	quietas
estancada	aire
fresca	condicionado
fría	cálido
& fuerte	cargado
mineral	comprimido
muerta	contaminado
& nieve	diáfano
& oxigenada	fresco
pesada	húmedo
poco profunda	libre
potable	líquido
purgante	caliente
viva	puro
caliente	seco
salada	cargado
bendita	pesado
sucia	limpio
pura	helado
limpia	aires
negra	& de gran señor
contaminada	condicionados
abundante	marinos
espumosa	alambre
aguardiente	de púas
destilado	espinoso
fuerte	galvanizado
aguas	alarma
de creciente	falsa
de menguante	contra incendio
de uso	social
frescas	acústica
interiores	general
jurisdiccionales	alarmas
marinas	sensibles
negras	contra incendio

ANEXO A.1 BASE PRIMARIA DE COLOCACIONES MODIFICADORAS

protectoras	máxima
electrónicas	inaccesible
alberca	determinada
olímpica	media
aldea	superior
perdida	total
remota	reglamentaria
alergia	alusión
de contacto	indirecta
respiratoria	transparente
retardada	ama
alfabetización	de casa
completa	de llaves
alfabeto	amar
cirílico	apasionadamente
fonético	ambiente
latino	cargado
algodón	húmedo
en rama	informal
hidrófilo	ruidoso
&S pólvora	sideral
algoritmos	ámbito
S genéticos	cultural
paralelos	económico
alimentación	educativo
adecuada	hispanico
balanceada	laboral
balanceada	nacional
de grasa	político
de motores	religioso
grasa	riguroso
rigurosa	social
vegetariana	ambulancia
alimento	de asistencia pública
vegetal	de correos
alma	sanitaria
atravesada	amenaza
& de caballo	de guerra
& de cántaro	grande
& de Dios	grave
en pena	amigo
almacenes	personal
grandes	de la infancia
almanaque	mejor
de pared	viejo
exfoliados	íntimo
alteraciones	fiel
cerebrales	inseparable
alternativa	leal
mejor	sincero
posible	cercano
verdadera	especial
altura	amistad

ANEXO A.1 BASE PRIMARIA DE COLOCACIONES MODIFICADORAS

<i>de los pueblos</i>	<i>platónico</i>
<i>indisoluble</i>	<i>propio</i>
<i>sincera</i>	<i>recíproco</i>
<i>tierna</i>	<i>sexual</i>
<i>profunda</i>	<i>amplio</i>
<i>pura</i>	<i>más</i>
<i>amor</i>	<i>muy</i>
<i>carnal</i>	<i>análisis1<SING</i>
<i>eterno</i>	<i>clínico</i>
<i>falso</i>	<i>cualitativo</i>
<i>filial</i>	<i>cuantitativo</i>
<i>íntimo</i>	<i>espectral</i>
<i>loco</i>	<i>estadístico</i>
<i>maternal</i>	<i>gramatical</i>
<i>mutuo</i>	<i>literario</i>
<i>pasajero</i>	<i>numérico</i>
<i>pasional</i>	<i>profundo</i>

Anexo A.2 Base primaria de colocaciones con valencias verbales

accidente	asilo
prevenir el ~	buscar ~
aceitunas	conceder ~
& llegar con las ~	aspectos
acertijo	cubrir los ~
adivinar un ~	asuntos
acontecimientos	inmiscuirse en los ~
adelantarse a los ~	ataque
adivinanza	lanzarse al ~
contar una ~	atención
admiración	atraer ~
extasiarse de ~	dedicar ~
adquisición	distracer ~
facilitar la ~	prestar ~
agua	atmósfera
diluir ... con el ~	caldear la ~
meterse en el ~	automóvil
saltar al ~	caer bajo un ~
aguas	autoridad
curarse con ~	aplastar ... con la ~
tomar las ~	auxilio
alcohol	prestar ~
diluir ~ ...	aviso
alcoholismo	dar el ~
combatir el ~	escribir el ~
alegría	ayuda
brincar de ~	prestar ~
saltar de ~	bandera
sentir ~	arriar la ~
alternativas	enarbolar la ~
ofrecer ~	izar la ~
almohada	baño
& consultar ... con la ~	meterse en el ~
alumno	barba
preguntar un ~	cortarse la ~
angustia	dejarse la ~
~ ... se oprime	& hacer la ~
apoyo	barrera
prestar ~	saltar la ~
apelación	barreras
interponer la ~	poner ~
árbol	basura
subirse a un ~	tirar ... a la ~
armario	batalla
rebuscar en el ~	ganar la ~
asco	bienes
dar ~	embargar los ~
ascuas	bigotes
& estar en ~	retorcerse los ~

ANEXO A.2 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS VERBALES

tener los ~	hacer la ~
teñir los ~	camino
boca	ir por el ~
abrir la ~	cántaros
tapar la ~ ...	& llover a ~
blanco	cal
dar en el ~	revocar ... con ~
errar el ~	calentura
fallar el ~	tener ~
bolas	campo
& meter ~	vivir en el ~
bolsillo	canall
rebuscar en el ~	compartir un ~
bomba	candidato
dar la ~	seleccionar al ~
bombilla	capa
~ luce	cubrir ... con una ~
borrachera	cara
andar de ~	cambiar de ~
bosque	dar la ~
ir por el ~	mentir a la ~
bostezo	mudar de ~
lanzar un ~	reírse en la ~
botas	romper la ~
quitarse las ~	cárcel
brasero	extenuar ~
atizar el ~	meter ... en el ~
brazos	cargo
abrir los ~	conquistar el ~
echarse ... en los ~	tener al ~
retorcer los ~	cargos
broma	desempeñar los ~
tomar a ~	carrera
caballos	hacer ~
domar ~	casas
cabellos	instalar en las ~
trancar los ~	instalar por las ~
cabeza	categoría
acariciar la ~	bajar de ~
agachar la ~	descender de ~
golpearse la ~	celos
inclinarse la ~	& tener ~
mover la ~	ceño
& perder la ~	fruncir ~
rascarse la ~	centros
volver la ~	~ se establecen
calor	establecer ~
& tener ~	certificado
calle	obtener el ~
ir por la ~	mostrar el ~
cajón	charla
rebuscar en el ~	conceder una ~
cama	conferir una ~

ANEXO A.2 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS VERBALES

dar una ~	expresar en el ~
chispas	expresarse en el ~
& echar ~	controversias
ciclo	dirimir las ~
cerrar el ~	convenios
clases1	celebrar ~
dar ~	convicción
impartir las ~	actuar por ~
& matar ~	convocatoria
codos	expedir ~
& hablar por los ~	cooperación
colores	prestar la ~
~ brotaron ...	corazón
sacar los ~ ...	~ brincó
comida	~ saltó
abalanzarse a la ~	cortina
comisión	bajar la ~
desempeñar la ~	cosas
comodidades	colocar las ~
ofrecer las ~	costillas
computadora	& medir las ~
usar una ~	crédito
conciencia	conceder un ~
remuerde la ~	cuidado
concierto	& tener ~
dar un ~	ciudad
conclusión	se edificó la ~
dar ~	ciudadanía
condecoraciones	~ se pierde
aceptar ~	perder la ~
usar ~	cuchara
conexión	& meter la ~
tener ~	cuello
confianza	bajar el ~
tener ~	cuenta
conflicto	poner a ~
poner en ~	meter a ~
conflictos	saldar una ~
generar ~	& tener en ~
confusión	cueros
reprimir la ~	ponerse en ~
congreso	culpa
convocar un ~	& tener la ~
conocimiento	cultura
perder el ~	difundir la ~
consecuencia	cumpleaños
sacar una ~	celebrar el ~
consecuencias	festejar el ~
sacar ~	cumplido
tener ~	decir un ~
contactos	hacer un ~
mantener ~	cursos
contrato	coordinar ~

ANEXO A.2 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS VERBALES

programar ~	enseñar los ~
datos	dificultades
& jugar a los ~	cansar ~
daño	dimisión
hacer ~	presentar la ~
daños	dinero
indemnizar los ~	ganar el ~
reparar los ~	~ se disipó
datos	~ se esfumó
mostrar ~	discurso
deber	concluir el ~
reconocer el ~	pronunciar un ~
deberes	disparo
dar ~	~ se oyó
dedo	disputa
engarabitar un ~	meterse a la ~
&C chuparse el ~	desuso
dedos	caer en ~
mover los ~	deuda
deducción	liquidar una ~
hacer ~	pagar una ~
defensa	dinero
fortalecer la ~	malgastar el ~
delito	disponibilidad
conjuguar el ~	tener ~
democracia	dolor
lograr ~	aguantar ~
derecho	causar ~
tener el ~	gritar de ~
ganar el ~	& tener ~
ganarse el ~	tranquilizar ~
derechos	efecto
tener los ~	surtir ~
ganar los ~	ejemplo
pisotear los ~	seguir el ~ ...
derrota	tomar ... en ~
causar una ~	elecciones
infligir una ~	votar en las ~
descanso	embargo
conceder el ~	levantar el ~
privar el ~	enemigo
deseos	aniquilar al ~
arder en ~	lanzarse contra ~
desorden	pasarse al ~
estar en ~	empleados
destreza	nombrar los ~
sobresalir ... por ~	empleo
superar ... por ~	desempeñar el ~
deudas	empresa
contraer ~	cargar la ~
días	enfermedad
& correr los ~	evitar la ~
dientes	preservar contra la ~

ANEXO A.2 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS VERBALES

preservar de ~
enigma
acertar un ~
dar un ~
enseñanza
apoyar la ~
enfocar a la ~
entusiasmo
contagiar el ~
equipos
integrar a los ~
errores
cometer ~

escalera
bajar la ~
bajar por la ~
subir la ~
subir por la ~
escándalo
provocar un ~
espaldas
caer de ~
espanto
causar ~
estancias
organizar ~

Anexo A.3 Base primaria de colocaciones con valencias sustantivales

adiestramiento	consulta de la
fase del	campos de la
administración1	salida de la
servicios de la	registro de la
ventaja de la	ángulo de la
ventajas de la	cachorro
administrador	carácter del
permisos del	campamento
adverbio	metros del
funcionalidad del	campo
agua	sintaxis del
nivel del	parte del
amo	valor del
dominio del	nombre del
análisis2<PLUR	capital1
costo de los	nombre de la
pago de los	casa
resultado de los	persona de la
ángel	dueño de la
& cabello de	cerebro
animal	funcionamiento del
naturaleza del	función del
sensibilidad del	ciudadano
nombre del	deberes del
cola del	garantías del
año	obligación del
ventas del	obligaciones del
aplicación	ciudadanos
interfaz de la	deberes de los
archivo	garantías de los
contenido del	obligaciones de los
contraseña del	cliente
descripción del	obligaciones del
extensión del	derechos del
formato del	petición del
nombre del	computadora
petición del	configuración de la
ubicación del	parte de la
autor	reparación de la
nombre del	conector
aviación	funcionamiento del
bombardeos de la	archivo del
ayer	características del
recuerdo del	etiquetas del
base	consulta
información de la	resultados de la
ubicación de la	controlador
datos de la	especificación del
columnas de la	correa

ANEXO A.3 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

extremo de la
 límite de la
 corazón
 sonidos del
 cuello
 piel del
 cuerpo
 material del
 diente
 extremo del
 extracción del
 superficie del
 dirección
 elementos de la
 disciplina
 resultados de la
 disco
 canciones del
 lado del
 distancia
 mitad de la
 documento
 construcción del
 empleado
 manual del
 encabezado
 parte del
 enemigo
 fuego contra el
 entrenamiento
 modalidad del
 equipo
 nombre del
 especificación
 revisión de la
 escándalo
 chispa del
 burlas por el
 éxito
 semillas del
 camino del
 fin
 principio del
 fuego
 & prueba de
 formato
 parte del
 formulario
 controles del
 datos del
 fracción
 párrafo de la
 gobierno

credibilidad del
 efectividad del
 política del
 vocero del
 grupo
 control del
 recursos del
 guerra
 comienzo de la
 fin de la
 hambre
 dolor del
 sufrimiento debido al
 hipótesis<SING
 demostración de la
 planteamiento de la
 hocico
 presa del
 infantería
 ataque de la
 retirada de la
 información
 contenido de la
 recuperación de la
 instalaciones
 recorrido por las
 instrucción
 formato de la
 ejecución de la
 utilización de la
 instrumental
 esterilidad del
 inteligencia
 estatus de la
 intervención
 día de la
 juego
 traducción del
 lengua
 lesiones de la
 líneal
 parámetro de la
 lista
 elemento de la
 libertad
 amante de la
 clima de
 lucha por la
 mano
 dorso de la
 borde de la
 manos
 & juego de

ANEXO A.3 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

<i>masa</i>	<i>organización</i>
<i>valores de la</i>	<i>informativos de la</i>
<i>método</i>	<i>documentos de la</i>
<i>detalle del</i>	<i>organizaciones</i>
<i>estructura del</i>	<i>naturaleza de las</i>
<i>explicación del</i>	<i>órgano</i>
<i>pasos del</i>	<i>dimensiones del</i>
<i>miembro</i>	<i>origen</i>
<i>nombre del</i>	<i>nombre del</i>
<i>perfil del</i>	<i>definición del</i>
<i>minerales</i>	<i>página</i>
<i>características de los</i>	<i>propiedades de la</i>
<i>detección de los</i>	<i>dirección de la</i>
<i>descripción de los</i>	<i>país</i>
<i>distribución de</i>	<i>áreas del</i>
<i>propiedades de los</i>	<i>civiles del</i>
<i>procesamiento de</i>	<i>oeste del</i>
<i>mundo</i>	<i>ciudades del</i>
<i>fin del</i>	<i>estructura del</i>
<i>partes del</i>	<i>presidente del</i>
<i>pueblos del</i>	<i>paquete</i>
<i>resto del</i>	<i>tamaño del</i>
<i>rincones del</i>	<i>peso del</i>
<i>naturaleza</i>	<i>parámetro</i>
<i>armonía con la</i>	<i>nombre del</i>
<i>contacto con la</i>	<i>valor del</i>
<i>conservación de la</i>	<i>parque</i>
<i>defensa de la</i>	<i>árboles del</i>
<i>elementos de la</i>	<i>flores del</i>
<i>enemigo de la</i>	<i>palomas del</i>
<i>enlace con la</i>	<i>pasado</i>
<i>fuerzas de la</i>	<i>tema del</i>
<i>imágenes de la</i>	<i>hechos del</i>
<i>manifestación de la</i>	<i>paz</i>
<i>maravillas de la</i>	<i>amante de la</i>
<i>postales de la</i>	<i>pensamiento</i>
<i>protección de la</i>	<i>funcionamiento del</i>
<i>reservas de la</i>	<i>perro</i>
<i>sonidos de la</i>	<i>ánimo del</i>
<i>odio</i>	<i>transgresiones del</i>
<i>amor al</i>	<i>atención del</i>
<i>batalla contra el</i>	<i>trasero del</i>
<i>causas del</i>	<i>confianza del</i>
<i>presa del</i>	<i>movimiento del</i>
<i>opción</i>	<i>reacción del</i>
<i>valores de la</i>	<i>pista</i>
<i>oración1<"lingüística"</i>	<i>inicio de la</i>
<i>construcción de la</i>	<i>poblado</i>
<i>parte de la</i>	<i>habitantes del</i>
<i>oración2<"religión"</i>	<i>programa</i>
<i>pasos de la</i>	<i>ejemplares del</i>
<i>orden</i>	<i>nombre del</i>
<i>significado de la</i>	<i>protocolo</i>

ANEXO A.3 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

monitorización del	inicio de la
versión del	final de la
prueba	mapa de la
descripción de la	punto de la
ejemplos de la	recorrido de la
explicación de la	resto de la
importancia de la	sección
proceso de la	coordinadora de la
resultados de la	índice de la
publicación	información de la
avance de la	inicio de la
características de la	ejemplo de la
contenido de la	jefe de la
desarrollo de la	objetivos de la
ejemplares de la	organigrama de la
fracaso de la	personal de la
historia de la	serie
índice de la	número de la
motivo de la	productos de la
retraso de la	transcurso de la
responsables de la	secuencia de la
ventajas de la	servidor
pueblo	nombre del
canto del	variables del
correo del	sistema
defensor del	archivos del
familias del	datos del
habitantes del	estado del
héroes del	nombre del
teatro del	presentación del
partido del	sociedad
voz del	aspectos de la
puerta	base de la
diseño de la	cambio de la
manija de la	obligaciones de la
marco de la	sonido
red	importancia del
exploración de la	información del
equipo de la	origen del
información de la	tipo de
usuarios de la	sonidos
paquetes de la	arte de los
registro	articulación de los
copia del	amplificación de los
creación del	característica de los
ruta	creador de los
búsqueda de la	descripción de los
continuación de la	determinación de los
detalles de la	estudio de los
determinación de la	grabación de los
etapas de la	mezcla de los
evaluación de la	mundos de los
extremo de la	origen de los

ANEXO A.3 BASE PRIMARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

suscripción

*ampliación de la
cancelación de la
características de la
condiciones de la
confirmación de la
datos de la
detalles de la
duración de la
fecha de la
importe de la
pago de la
período de la*

*precio de la
renovación de la
vencimiento de la
vigencia de la
sustancias
comercialización de
estudio de las
manejo de las
nomenclatura de las
paso de las
sustantivo
modificación del
teorema*

Anexo A.4 Base primaria de hipónimos / hiperónimos

<i>adornos</i>	<i>bebida caliente</i>
<i> anillos</i>	<i>bebida fuerte</i>
<i> aretes</i>	<i> refrescos</i>
<i> collares</i>	<i>bebida caliente</i>
<i> pendientes</i>	<i> cacao</i>
<i>animales</i>	<i> café</i>
<i> animales domésticos</i>	<i> chocolate</i>
<i> felinos</i>	<i> té</i>
<i> reptiles</i>	<i>bebida fuerte</i>
<i> roedores</i>	<i> champán</i>
<i>animales domésticos</i>	<i> licor</i>
<i> cerdo</i>	<i> ron</i>
<i> ganado</i>	<i> vodka</i>
<i> gato</i>	<i> whisky</i>
<i> pájaro</i>	<i>refrescos</i>
<i> perro</i>	<i> Coca-cola</i>
<i>antibióticos</i>	<i> Fanta</i>
<i> eritromicina</i>	<i> Mirinda</i>
<i> estreptomycin</i>	<i> Pepsi Cola</i>
<i> penicilina</i>	<i>carne</i>
<i>aparatos</i>	<i> bisté</i>
<i> aspiradora</i>	<i> chuleta</i>
<i> iluminador</i>	<i> liebre</i>
<i> lavadora</i>	<i> pollo</i>
<i> refrigerador</i>	<i> salchichas</i>
<i> tostador</i>	<i>casa habitacional</i>
<i> ventilador</i>	<i> apartamento</i>
<i>autor</i>	<i> cabaña</i>
<i> crítico</i>	<i> casa</i>
<i> escritor</i>	<i> choza</i>
<i> periodista</i>	<i>categoría gramatical</i>
<i> poeta</i>	<i> caso</i>
<i>baile</i>	<i> género</i>
<i> charlestón</i>	<i> modo</i>
<i> cumbia</i>	<i> número</i>
<i> lambada</i>	<i>cereales</i>
<i> polca</i>	<i> arroz</i>
<i> samba</i>	<i> avena</i>
<i> tango</i>	<i> cebada</i>
<i>bayas</i>	<i> trigo</i>
<i> arándano</i>	<i>ciencia</i>
<i> cereza</i>	<i> ciencia de la humanidad</i>
<i> frambuesa</i>	<i> ciencia exacta</i>
<i> fresa</i>	<i> ciencia médica</i>
<i> grosella</i>	<i> ciencia natural</i>
<i> grosella</i>	<i> ciencia técnica</i>
<i> morera</i>	<i>ciencia de la humanidad</i>
<i> zarza</i>	<i> antropología</i>
<i>bebidas</i>	<i> arqueología</i>

ANEXO A.4 BASE PRIMARIA DE HIPÓNIMOS / HIPERÓNIMOS

economía	combustible
filología	aceite
filosofía	alcohol
historia	carbón
jurisprudencia	gas
periodismo	gasolina
ciencia exacta	leña
cibernética	petróleo
física	querosén
informática	constelación
matemática	Cangrejo
ciencia natural	Dragón
biología	Galaxia
botánica	Gemelos
genética	Pegaso
geología	Virgo
geografía	construcciones
química	casa habitacional
ciencia técnica	castillo
electrónica	hotel
informática	kiosco
metalurgia	pabellón
minería	residencia
cítricos	teatro
limones	continente
mandarinas	África
naranjas	América
toronjas	Asia
color	Australia
amarillento	Europa
amarillo	coordenada
ámbar	abscisa
anaranjado	elevación
azul	rango
blanco	cuadro
castaño	grabando
crema	paisaje
durazno	retrato
esmeralda	cuerpo gobernante
índigo azul	administración
lavanda	directorio
limón	gobierno
mostaza	deporte
mulato	acrobacia
negro	atletismo
paja	gimnasia
púrpura	lucha libre
rojo	natación
rosa	deportista
turquesa	corredor
verde	disparador
verdoso	futbolista
violeta	gimnasta

ANEXO A.4 BASE PRIMARIA DE HIPÓNIMOS / HIPERÓNIMOS

luchador	disentería
nadador	gripe
saltador	hipertensión
tenista	hipotensión
día de semana	infarto
domingo	SIDA
jueves	tuberculosis
lunes	venérea
martes	escuela
miércoles	academia
sábado	instituto
viernes	universidad
diplomático	especialista
cónsul	antropólogo
delegado	arqueólogo
embajador	artista
dispositivo óptico	banquero
microscopio	biólogo
prismático	científico
telescopio	economista
documento	físico
acuerdo	geógrafo
aplicación	geólogo
certificado	historiador
contrato	lingüista
declaración	maestro
declaración	matemático
decreto	médico
demanda	profesor
diploma	programador
disertación	químico
informe	psicólogo
inventario	estación
licencia	invierno
protocolo	otoño
registro	primavera
tratado	verano
edad	estrato
adolescencia	campesinos
infancia	clase media
juventud	empleados
niñez	gerentes
empresa de comida	intelectuales
bar	obreros
cafetería	extremos
pizzería	máximo
restaurante	mínimo
enfermedad	óptimo
apoplejía	felinos
asma	gato
cáncer	león
cólera	leopardo

ANEXO A.4 BASE PRIMARIA DE HIPÓNIMOS / HIPERÓNIMOS

lince	crédito
panteras	forma instructiva
puma	conferencia
tigre	exposición
fenómeno	libro
calma	seminario
eclipse	foro
sequedad	conferencia
tempestad	congreso
terremoto	convención
tiempo	reunión
tormenta	simposio
viento	taller
fenómeno social	frutas
desarmamento	bayas
desempleo	ciruelas
estancamiento	cítricos
evolución	duraznos
movilización	manzanas
privatización	peras
prosperidad	piña
reacción	plátanos
represiones	uvas
retroceso	fuerzas armadas
revolución	acorazadas
figura de tres dimensión	armada
cilindro	fuerza aérea
cono	infantería
elipsoide	función geométrica
poliedro	cosecante
filología	coseno
lingüística	cotangente
flora	seno
árbol	tangente
arbusto	gas
cereal	aire
césped	amoníaco
flores	argón
flores	hidrógeno
azucenas	metano
callas	neón
dalias	nitrógeno
gardenias	oxígeno
gladiolos	gastronomía
lilas	huevos
manzanillas	queso
margaritas	salchicha
rosas	género1
tulipanes	femenino
violetas	masculino
fondo	neutro
certificado	género2
cheque	boceto

ANEXO A.4 BASE PRIMARIA DE HIPÓNIMOS / HIPERÓNIMOS

ensayo	Esloveno
folletín	Sueco
historia	Turco
memorias	Vasco
narrativo	iglesias
novela	católica
poema	griega
poesía	ortodoxa
guarnición	romana
arroz	inferencia lógica
pasta	deducción
patatas fritas	inducción
puro	instrumento de mesa
idioma	azucarero
Afgano	cuchara
Albanés	cuchillo
Alemán	salero
Árabe	servilletero
Armenio	tenedor
Azerbaiján	intervalo de tiempo
Bielorruso	año
Birmano	década
Búlgaro	día
Catalán	época
Checo	hora
Chino	mes
Coreano	milenio
Eslovaco	minuto
Español	segundo
Estonio	semana
Finlandés	siglo
Flamenco	instrumento musical
Francés	acordeón
Griego	clarinete
Hebreo	contrabajo
Holandés	flauta
Húngaro	guitarra
idioma artificial	piano
idioma extranjero	saxofón
Inglés	tambor
Islandés	trompeta
Italiano	violín
Latino	violonchelo
lenguaje de programación	jefe
Lituano	decano
Mongol	diputado
Noruego	director
Polaco	gerente
Portugués	ministro
Rumano	presidente
Ruso	rector
Serbio	joyería

ANEXO A.4 BASE PRIMARIA DE HIPÓNIMOS / HIPERÓNIMOS

<i>brillantes</i>	<i>ladrillos</i>
<i>collar</i>	<i>linóleo</i>
<i>jurisprudencia</i>	<i>mármol</i>
<i>abogado</i>	<i>vidrio</i>
<i>juez</i>	<i>materiales radiactivos</i>
<i>lenguaje de programación</i>	<i>plutonio</i>
<i>Ada</i>	<i>uranio</i>
<i>Algol</i>	<i>médico</i>
<i>C</i>	<i>audiólogo</i>
<i>C++</i>	<i>cirujano</i>
<i>Cobol</i>	<i>dentista</i>
<i>Fortran</i>	<i>dermatólogo</i>
<i>Java</i>	<i>neurólogo</i>
<i>Pascal</i>	<i>oculista</i>
<i>Prolog</i>	<i>oncólogo</i>
<i>limpiador</i>	<i>pediatra</i>
<i>champú</i>	<i>proctólogo</i>
<i>detergente</i>	<i>reumatólogo</i>
<i>jabón</i>	<i>terapeuta</i>
<i>línea</i>	<i>medida de longitud</i>
<i>curva</i>	<i>centímetro</i>
<i>espiral</i>	<i>decímetro</i>
<i>hipérbola</i>	<i>kilómetro</i>
<i>ondulada</i>	<i>micra</i>
<i>parábola</i>	<i>milímetro</i>
<i>recta</i>	<i>milla</i>
<i>líquido</i>	<i>pie</i>
<i>ácido</i>	<i>pulgada</i>
<i>agua</i>	<i>medida de peso</i>
<i>alcalino</i>	<i>gramo</i>
<i>lugar</i>	<i>kilogramo</i>
<i>ciudad</i>	<i>libra</i>
<i>pueblo</i>	<i>onza</i>
<i>suburbio</i>	<i>tonelada</i>
<i>mar</i>	<i>mensaje</i>
<i>Mar Adriático</i>	<i>carta</i>
<i>Mar Báltico</i>	<i>telegrama</i>
<i>Mar Blanco</i>	<i>mes</i>
<i>Mar Caribeño</i>	<i>mes de invierno</i>
<i>Mar Mediterráneo</i>	<i>mes de otoño</i>
<i>Mar Negro</i>	<i>mes de verano</i>
<i>Mar Rojo</i>	<i>mes primaveral</i>
<i>marinero</i>	<i>mes de invierno</i>
<i>almirante</i>	<i>Diciembre</i>
<i>capitán</i>	<i>Enero</i>
<i>compañero</i>	<i>Febrero</i>
<i>contraaestre</i>	<i>mes de otoño</i>
<i>material para construcción</i>	<i>Noviembre</i>
<i>alfombra</i>	<i>Octubre</i>
<i>arena</i>	<i>Septiembre</i>
<i>azulejo</i>	<i>mes de verano</i>
<i>granito</i>	<i>Agosto</i>
<i>hormigón</i>	<i>Julio</i>

ANEXO A.4 BASE PRIMARIA DE HIPÓNIMOS / HIPERÓNIMOS

<i>Junio</i>	<i>metal precioso</i>
<i>mes primaveral</i>	<i>oro</i>
<i>Abril</i>	<i>plata</i>
<i>Marzo</i>	<i>platino</i>
<i>Mayo</i>	<i>material crudo</i>
<i>metal</i>	<i>algodón</i>
<i>metal básico</i>	<i>forraje</i>
<i>metal férreo</i>	<i>lana</i>
<i>metal precioso</i>	<i>madera</i>
<i>metal básico</i>	<i>maíz</i>
<i>aluminio</i>	<i>microorganismos</i>
<i>bronce</i>	<i>bacterias</i>
<i>cobre</i>	<i>microbios</i>
<i>estaño</i>	<i>virus</i>
<i>latón</i>	<i>miembro de frase</i>
<i>mercurio</i>	<i>asunto</i>
<i>níquel</i>	<i>circunstancia</i>
<i>titanio</i>	<i>complemento</i>
<i>metal férreo</i>	<i>objeto</i>
<i>acero</i>	<i>predicado</i>
<i>hierro</i>	<i>sujeto</i>

Anexo B. Bases secundarias de colocaciones

Anexo B.1 Base secundaria de colocaciones modificadoras

+11110	<i>absolutismo</i>	-20010	<i>específico</i>
-20030	<i>ilustre</i>	-20020	<i>individual</i>
-20010	<i>impuesto</i>	-20020	<i>penal</i>
-20050	<i>francés</i>	-20020	<i>puntual</i>
-20020	<i>provincial</i>	-20010	<i>minero</i>
-20020	<i>real</i>	-20010	<i>heroico</i>
+12120	<i>acción</i>	-20010	<i>encaminado</i>
-20010	<i>armado</i>	-20010	<i>necesario</i>
-20010	<i>bueno</i>	-20030	<i>prevista</i>
-20010	<i>combinado</i>	-20010	<i>indigno</i>
-20000	<i>de gracias</i>	-20030	<i>simple</i>
-20010	<i>democrático</i>	-20020	<i>inútil</i>
-20010	<i>deshonesto</i>	-20010	<i>positivo</i>
-20050	<i>destructor</i>	-20020	<i>unilateral</i>
-20010	<i>dramático</i>	-20020	<i>especial</i>
-20010	<i>directo</i>	-20010	<i>implícito</i>
-20010	<i>educativo</i>	-20010	<i>urbano</i>
-20010	<i>ejecutado</i>	-20010	<i>delictivo</i>
-20010	<i>físico</i>	-20020	<i>irregular</i>
-20020	<i>gubernamental</i>	-20010	<i>constructivo</i>
-20010	<i>humano</i>	+11110	<i>aceite</i>
-20010	<i>inmediato</i>	-20000	<i>de anís</i>
-20020	<i>judicial</i>	-20000	<i>de bacalao</i>
-20010	<i>jurídico</i>	-20000	<i>de cañadón</i>
-20030	<i>libre</i>	-20000	<i>de girasol</i>
-20020	<i>militar</i>	-20000	<i>de hígado</i>
-20020	<i>nacional</i>	-20000	<i>de linaza</i>
-20030	<i>nacionalista</i>	-20000	<i>de maíz</i>
-20020	<i>natural</i>	-20000	<i>de máquinas</i>
-20020	<i>penal</i>	-20000	<i>de oliva</i>
-20020	<i>popular</i>	-20000	<i>de ricino</i>
-20010	<i>perfecto</i>	-20010	<i>quemado</i>
-20010	<i>público</i>	-20020	<i>esencial</i>
-20010	<i>prometido</i>	-20030	<i>comestible</i>
-20020	<i>social</i>	-20010	<i>producido</i>
-20020	<i>verbal</i>	-20010	<i>barato</i>
+12220	<i>acciones</i>	-20020	<i>vegetal</i>
-20010	<i>arbitrario</i>	-20010	<i>puro</i>
-20010	<i>correctivo</i>	-20020	<i>mineral</i>
-20010	<i>ordinario</i>	-20020	<i>vegetal</i>
-20040	<i>concrete</i>	+11210	<i>aceites</i>
-20010	<i>bélico</i>	-20010	<i>pesado</i>
-20010	<i>político</i>	-20020	<i>volátil</i>
-20010	<i>violento</i>	-20010	<i>utilizado</i>
-20020	<i>militar</i>	-20020	<i>industrial</i>

ANEXO B.1 BASE SECUNDARIA DE COLOCACIONES MODIFICADORAS

+11110	acero	-20020	mental
-20010	colado	+12220	actividades
-20010	cromado	-20010	científico
-20030	inoxidable	-20010	deportivo
-20020	martín	-20010	económico
-20010	templado	-20010	manufacturero
-20010	pulido	-20010	tecnológico
-20030	caliente	-20020	cultural
-20010	forjado	-20020	comercial
-20010	recubierto	-20010	político
+11110	ácido	-20030	agrícola
-20010	arsénico	-20010	humano
-20010	bórico	-20020	social
-20010	carbónico	-20010	productivo
-20010	cianhídrico	-20010	educativo
-20010	cítrico	-20020	musical
-20010	clorhídrico	-20010	religioso
-20010	fénico	-20010	ilícito
-20010	láctico	-20010	paralelo
-20010	nítrico	-20030	terrorista
-20010	sulfúrico	-20030	importante
+30020	acoger	-20020	laboral
-40000	con brazos abiertos	-20010	financiero
-40000	favorablemente	-20010	recreativo
+30040	acostar	-20010	público
-40000	terriblemente	-20010	caritativo
+12120	actividad	-20040	concrete
-20010	bélico	-20010	programado
-20010	bullicioso	-20020	principal
-20010	científico	-20020	empresarial
-20010	deportivo	-20020	gubernamental
-20010	económico	-20010	conjunto
-20010	físico	-20010	bancario
-20010	mecánico	-20020	estudiantil
-20010	tecnológico	-20010	inmediato
-20020	vital	+11110	acto
-20010	ganadero	-20000	a la bandera
-20010	político	-20020	brutal
-20010	humano	-20010	cívico
-20020	comercial	-20000	de violencia
-20020	industrial	-20010	delictivo
-20020	cultural	-20040	grandote
-20010	artístico	-20030	proselitista
-20020	profesional	-20010	público
-20020	laboral	-20010	reclamado
-20030	agrícola	-20010	religioso
-20020	musical	-20020	sexual
-20010	productivo	-20010	seguido
-20020	militar	-20010	único
-20020	sexual	-20010	segundo
-20020	cerebral	-20010	primero
-20010	biológico	-20030	solemne
-20010	petrolero	-20020	social

ANEXO B.1 BASE SECUNDARIA DE COLOCACIONES MODIFICADORAS

-20010	político	-20010	definitivo
-20020	oficial	-20020	global
-20020	central	-20020	preliminar
-20010	realizado	-20010	completo
-20030	terrorista	-20010	secreto
-20010	ilícito	-20020	formal
-20030	importante	-20020	verbal
-20030	relevante	-20020	matrimonial
-20010	precioso	-20020	legal
-20020	ilegal	-20020	parcial
-20010	violento	-20030	honorable
+11210	actos	+11210	adelantos
-20010	delictivo	-20010	científico
-20010	público	-20030	importante
-20010	reclamado	-20010	serio
-20010	religioso	-20010	tecnológico
-20010	conjunto	-20010	técnico
-20010	humano	-20010	logrado
-20020	cultural	-20010	médico
-20030	terrorista	+11110	adiestramiento
-20020	oficial	-20010	básico
-20020	judicial	-20010	canino
-20020	central	-20020	militar
-20010	violento	-20020	profesional
-20020	musical	+12120	administración1
-20010	contrario	-20020	central
-20020	formal	-20000	de empresas
-20020	criminal	-20010	doméstico
-20010	aislado	-20020	estatal
-20010	heroico	-20020	federal
-20030	incalificable	-20020	local
-20030	inconsciente	-20020	municipal
-20010	indebido	-20010	público
-20010	perverso	-20020	territorial
-20010	prohibido	-20010	centralizado
+11110	acuerdo	-20020	general
-20020	bilateral	+12120	administración2
-20020	común	-20000	de justicia
-20020	condicional	+30010	administrar
-20020	eventual	-40000	bien
-20020	final	-40000	adecuadamente
-20010	parlamentario	-40000	libremente
-20010	político	-40000	independientemente
-20020	zonal	+12110	advertencia
-20020	general	-20010	oportuno
-20020	comercial	-20010	primero
-20020	nacional	-20010	último
-20010	firmado	-20010	serio
-20010	tácito	-20010	previo
-20010	logrado	+30010	afectar
-20020	internacional	-40000	sustancialmente
-20010	mutuo	-40000	directamente

ANEXO B.1 BASE SECUNDARIA DE COLOCACIONES MODIFICADORAS

-40000	negativamente	-20030	dulce
-40000	indirectamente	-20010	estancado
-40000	físicamente	-20010	fresco
-40000	gravemente	-20010	frío
+30010	agarrar	-20030&	fuerte
-40000	al toro por los cuernos	-20020	mineral
+11110	agente1<CLASS:hombre	-20010	muerto
-20010	británico	-20030&	nieve
-20000	de bolsa	-20010&	oxigenado
-20000	de negocios	-20010	pesado
-20000	de policía	-20000	poco profunda
-20000	del servicio de contra-espionaje	-20030	potable
-20050	invasor	-20030	purgante
-20010	secreto	-20010	vivo
-20010	encubierto	-20030	caliente
-20020	fiscal	-20010	salado
-20020	oficial	-20010	bendito
-20010	infiltrado	-20010	sucio
-20020	judicial	-20010	puro
-20010	político	-20010	limpio
+11110	agente2<CLASS:sustancia	-20010	negro
-20010	dañino	-20010	contaminado
-20000	de limpieza	-20030	abundante
-20010	patógeno	-20010	espumoso
-20010	químico	+12210	aguas
-20020	viral	-20000	de creciente
-20010	infeccioso	-20000	de menguante
-20010	erosivo	-20000	de uso
-20010	extraño	-20010	fresco
-20010	tóxico	-20020	interior
-20010	venenoso	-20020	jurisdiccional
+11110	agotamiento	-20010	marino
-20010	físico	-20010	negro
-20020	mental	-20020	pluvial
+20010	agradecido	-20020	residual
-40000	muy	-20010	subterráneo
-40000	poco	-20020	termal
-40000	sumamente	-20020	territorial
-40000	eternamente	-20010	profundo
+12110	agua	-20020	superficial
-20010	carbónico	-20020	interior
-20010	claro	-20030	corriente
-20030	corriente	-20010	estancado
-20010	cristalino	-20030	salobre
-20000&	de cepas	-20030	salvaje
-20000&	de colonia	-20020	internacional
-20000	de frutas	-20010	salado
-20000	de mar	-20010	sucio
-20000	de mesa	-20020	jurisdiccional
-20000	de purga	-20020	tropical
-20000	de régimen	-20040	quiete
-20000	de roca	+11110	aire
		-20010	acondicionado

ANEXO B.1 BASE SECUNDARIA DE COLOCACIONES MODIFICADORAS

-20010	cálido	+11110	alimento
-20010	cargado	-20020	vegetal
-20010	comprimido	+12110	alma
-20010	contaminado	-20010	atravesado
-20010	diáfano	-20000&	de caballo
-20010	fresco	-20000&	de cántaro
-20010	húmedo	-20000&	de Dios
-20030	libre	-20000	en pena
-20010	líquido	+12220	alteraciones
-20030	caliente	-20020	cerebral
-20010	puro	+12110	altura
-20010	seco	-20010	máximo
-20010	cargado	-20030	inaccesible
-20010	pesado	-20010	determinado
-20010	limpio	-20010	medio
-20010	helado	-20020	superior
+11220	aires	-20020	total
-20000&	de gran señor	-20010	reglamentario
-20010	acondicionado	+12120	alusión
-20010	marino	-20010	indirecto
+11110	alambre	-20030	transparente
-20000	de púas	+12110	ama
-20010	espinoso	-20000	de casa
-20010	galvanizado	-20000	de llaves
+12110	alarma	+30010	amar
-20010	falso	-40000	apasionadamente
-20000	contra incendio	+11110	ambiente
-20020	social	-20010	cargado
-20010	acústico	-20010	húmedo
-20020	general	-20020	informal
+12210	alarmas	-20010	ruidoso
-20030	sensible	-20020	sideral
-20000	contra incendio	+11110	ámbito
-20050	protector	-20020	cultural
-20010	electrónico	-20010	económico
+12110	alergia	-20010	educativo
-20000	de contacto	-20010	hispanico
-20010	respiratorio	-20020	laboral
-20010	retardado	-20020	nacional
+12120	alfabetización	-20010	político
-20040	complete	-20010	religioso
+11110	alfabeto	-20010	riguroso
-20010	cirílico	-20020	social
-20010	fonético	+12110	ambulancia
-20010	latino	-20000	de asistencia pública
+11120	algodón	-20000	de correos
-20000	en rama	-20010	sanitario
-20010	hidrófilo	+11110	amigo
-20010&	Spólvora	-20020	personal
+11210	algoritmos	-20000	de la infancia
-20010	Sgenético	-20020	mejor
-20010	paralelo	-20010	viejo

ANEXO B.1 BASE SECUNDARIA DE COLOCACIONES MODIFICADORAS

-20010	<i>íntimo</i>	-20010	<i>falso</i>
-20020	<i>fiel</i>	-20020	<i>filial</i>
-20030	<i>inseparable</i>	-20010	<i>íntimo</i>
-20020	<i>leal</i>	-20010	<i>loco</i>
-20010	<i>sincero</i>	-20020	<i>maternal</i>
-20010	<i>cercano</i>	-20010	<i>mutuo</i>
-20020	<i>especial</i>	-20010	<i>pasajero</i>
+12120	<i>amistad</i>	-20020	<i>pasional</i>
-20000	<i>de los pueblos</i>	-20010	<i>platónico</i>
-20030	<i>indisoluble</i>	-20010	<i>propio</i>
-20010	<i>sincero</i>	-20010	<i>recíproco</i>
-20010	<i>tierno</i>	-20020	<i>sexual</i>
-20010	<i>profundo</i>	+20010	<i>amplio</i>
-20010	<i>puro</i>	-40000	<i>más</i>
+11120	<i>amor</i>	-40000	<i>muy</i>
-20020	<i>carnal</i>		
-20010	<i>eterno</i>		

Anexo B.2 Base secundaria de colocaciones con valencias verbales

+1111	accidente	+1111	asilo
-3003	00ÿ103prevenir	-3001	00ÿ002buscar
+1221	aceitunas	-3002	00ÿ002conceder
-3001&	00>104llegar	+1121	aspectos
+1111	acertijo	-3003	00ÿ103cubrir
-3001	00ÿ203adivinar	+1121	asuntos
+1121	acontecimientos	-3006	00;104inmiscuir
-3004	00'104adelantar	+1111	ataque
+1211	adivinanza	-3004	00'104lanzar
-3001	00ÿ203contar	+1212	atención
+1212	admiración	-3002	00ÿ002atraer
-3004	00 003extasiar	-3001	00ÿ002dedicar
+1212	adquisición	-3002	00ÿ002distraer
-3001	00ÿ103facilitar	-3001	00ÿ002prestar
+1211	agua	+1211	atmósfera
-3003	00>125diluir	-3001	00ÿ103caldear
-3005	00;104meter	+1112	automóvil
-3001	00'104saltar	-3002	00™204caer
+1221	aguas	+1212	autoridad
-3004	00>003curar	-3001	00>125aplastar
-3001	00ÿ103tomar	+1111	auxilio
+1112	alcohol	-3001	00ÿ002prestar
-3003	00ÿ032diluir	+1111	aviso
+1111	alcoholismo	-3001	00ÿ103dar
-3003	00ÿ002combatir	-3003	00ÿ103escribir
+1211	alegría	+1211	ayuda
-3001	00 003brincar	-3001	00ÿ002prestar
-3001	00 003saltar	+1211	bandera
-3003	00ÿ002sentir	-3001	00ÿ103arriar
+1221	alternativas	-3001	00ÿ103enarbolar
-3002	00ÿ002ofrecer	-3001	00ÿ103izar
+1211	almohada	+1111	baño
-3001&	00>125consultar	-3005	00;104meter
+1111	alumno	+1211	barba
-3001	00ÿ203preguntar	-3004	00ÿ103cortar
+1211	angustia	-3004	00ÿ103dejar
-3009	13ÿ021oprimir	-3002&	00ÿ103hacer
+1111	apoyo	+1211	barrera
-3001	00ÿ002prestar	-3001	00ÿ103saltar
+1212	apelación	+1221	barreras
-3002	00ÿ103interponer	-3002	00ÿ002poner
+1112	árbol	+1211	basura
-3006	00'204subir	-3001	00'125tirar
+1111	armario	+1211	batalla
-3001	00;104rebuscar	-3001	00ÿ103ganar
+1111	asco	+1122	bienes
-3001	00ÿ002dar	-3001	00ÿ103embargar
+1221	ascuas	+1121	bigotes
-3001&	00;003estar	-3005	00ÿ103retorcer

ANEXO B.2 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS VERBALES

-3002	00ÿ103	tener	-3002	00ÿ103	hacer
-3003	00ÿ103	teñir	+1111		camino
+1211		boca	-3003	00«104	ir
-3003	00ÿ103	abrir	+1121		cántaros
-3001	00ÿ143	tapar	-3002&	00'003	llover
+1111		blanco	+1212		cal
-3001	00ÿ103	dar	-3001	00>024	revocar
-3001	00ÿ103	errar	+1211		calentura
-3001	00ÿ103	efallar	-3002	00ÿ002	tener
+1221		bolas	+1111		campo
-3002&	00ÿ002	meter	-3003	00;104	vivir
+1111		bolsillo	+1112		canall
-3001	00;104	rebuscar	-3003	00ÿ203	compartir
+1211		bomba	+1111		candidato
-3001	00ÿ103	dar	-3001	00'104	seleccionar
+1211		bombilla	+1211		capa
-3003	13ÿ001	lucir	-3003	00>225	cubrir
+1211		borrachera	+1211		cara
-3001	00 003	andar	-3001	00 003	cambiar
+1111		bosque	-3001	00ÿ103	dar
-3003	00«104	ir	-3003	00'104	mentir
+1111		bostezo	-3001	00 003	mudar
-3001	00ÿ203	lanzar	-3006	00;104	reir
+1221		botas	-3002	00ÿ103	romper
-3004	00ÿ103	quitar	+1212		cárcel
+1111		brasero	-3001	00ÿ002	extenuar
-3001	00ÿ103	atizar	-3002	00;125	meter
+1121		brazos	+1111		cargo
-3003	00ÿ103	abrir	-3001	00ÿ103	conquistar
-3004	00;125	echar	-3002	00'104	tener
-3002	00ÿ103	retorcer	+1121		cargos
+1211		broma	-3001	00ÿ103	desempeñar
-3001	00'003	tomar	+1211		carrera
+1121		caballos	-3002	00ÿ002	hacer
-3001	00ÿ002	domar	+1221		casas
+1121		cabellos	-3001	00;104	instalar
-3001	00ÿ103	trancar	-3001	00«104	instalar
+1211		cabeza	+1211		categoría
-3001	00ÿ103	acariciar	-3001	00 003	bajar
-3001	00ÿ103	agachar	-3002	00 003	descender
-3004	00ÿ103	golpear	+1121		celos
-3001	00ÿ103	inclinarse	-3002&	00ÿ002	tener
-3002	00ÿ103	mover	+1111		ceño
-3002&	00ÿ103	perder	-3003	00ÿ002	fruncir
-3004	00ÿ103	rascar	+1121		centros
-3002	00ÿ103	volverse	-3008	16ÿ001	establecer
+1112		calor	-3002	00ÿ002	establecer
-3002&	00ÿ002	tener	+1111		certificado
+1211		calle	-3002	00ÿ103	obtener
-3003	00«104	ir	-3001	00ÿ103	mostrar
+1112		cajón	+1211		charla
-3001	00;104	rebuscar	-3002	00ÿ203	conceder
+1211		cama	-3003	00ÿ203	conferir

ANEXO B.2 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS VERBALES

-3001	00ÿ203	dar	-3001	00;104	expresar
+1221		chispas	-3004	00;104	expresar
-3001&	00ÿ002	echar	+1221		controversias
+1111		ciclo	-3003	00ÿ103	dirimir
-3001	00ÿ103	cerrar	+1121		convenios
+1221		clases1	-3001	00ÿ002	celebrar
-3001	00ÿ002	dar	+1212		convicción
-3003	00ÿ103	impartir	-3001	00«003	actuar
-3001&	00ÿ002	matar	+1211		convocatoria
+1121		codos	-3003	00ÿ002	expedir
-3001&	00«104	hablar	+1212		cooperación
+1122		colores	-3001	00ÿ103	prestar
-3001	26ÿ031	brotar	+1112		corazón
-3001	00ÿ143	sacar	-3001	23ÿ001	brincar
+1211		comida	-3001	23ÿ001	saltar
-3004	00'104	abalanzar	+1211		cortina
+1212		comisión	-3001	00ÿ103	bajar
-3001	00ÿ103	desempeñar	+1221		cosas
+1222		comodidades	-3001	00ÿ103	colocar
-3002	00ÿ103	ofrecer	+1221		costillas
+1211		computadora	-3003&	00ÿ103	medir
-3001	00ÿ203	usar	+1111		crédito
+1211		conciencia	-3002	00ÿ203	conceder
-3002	13ÿ103	remorder	+1111		cuidado
+1111		concierto	-3002&	00ÿ002	tener
-3001	00ÿ203	dar	+1212		ciudad
+1212		conclusión	-3007	23ÿ104	edificar
-3001	00ÿ002	dar	+1211		ciudadanía
+1222		condecoraciones	-3008	13ÿ001	perder
-3001	00ÿ002	aceptar	-3002	00ÿ103	perder
-3001	00ÿ002	usar	+1211		cuchara
+1212		conexión	-3002&	00ÿ103	meter
-3002	00ÿ002	tener	+1111		cuello
+1211		confianza	-3001	00ÿ103	bajar
-3002	00ÿ002	tener	+1211		cuenta
+1111		conflicto	-3002	00'003	poner
-3002	00;003	poner	-3002	00'003	meter
+1121		conflictos	-3001	00ÿ203	saldar
-3001	00ÿ002	generar	-3002&	00;003	tener
+1212		confusión	+1121		cueros
-3003	00ÿ103	reprimir	-3005	00;003	poner
+1111		congreso	+1211		culpa
-3001	00ÿ203	convocar	-3002&	00ÿ103	tener
+1111		conocimiento	+1211		cultura
-3002	00ÿ103	perder	-3003	00ÿ103	difundir
+1211		consecuencia	+1121		cumpleaños
-3001	00ÿ203	sacar	-3001	00ÿ103	celebrar
+1221		consecuencias	-3001	00ÿ103	festejar
-3001	00ÿ002	sacar	+1111		cumplido
-3002	00ÿ002	tener	-3003	00ÿ203	decir
+1121		contactos	-3002	00ÿ203	hacer
-3002	00ÿ002	mantener	+1121		cursos
+1111		contrato	-3001	00ÿ002	coordinar

ANEXO B.2 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS VERBALES

-3001	00ÿ002	programar	+1222	dificultades		
+1121		datos	-3001	00ÿ002	cansar	
-3001&	00'104	jugar	+1212		dimisión	
+1111		daño	-3001	00ÿ103	presentar	
-3002	00ÿ002	hacer	+1111		dinero	
+1121		daños	-3001	00ÿ103	ganar	
-3001	00ÿ103	indemnizar	-3007	23ÿ001	disipar	
-3001	00ÿ103	reparar	-3007	23ÿ001	esfumar	
+1121		datos	+1111		discurso	
-3001	00ÿ002	mostrar	-3003	00ÿ103	concluir	
+1112		deber	-3001	00ÿ203	pronunciar	
-3002	00ÿ103	reconocer	+1211		disputa	
+1122		deberes	-3005	00'104	meter	
-3001	00ÿ002	dar	+1111		desuso	
+1111		dedo	-3002	00;003	caer	
-3001	00ÿ203	engarabitar	+1211		deuda	
+1121		dedos	-3001	00ÿ203	liquidar	
-3002	00ÿ103	mover	-3001	00ÿ203	pagar	
+1212		deducción	+1111		dinero	
-3002	00ÿ002	hacer	-3001	00ÿ103	malgastar	
+1211		defensa	+1212		disponibilidad	
-3002	00ÿ103	fortalecer	-3002	00ÿ002	tener	
+1111		delito	+1112		dolor	
-3001	00ÿ103	conjugar	-3001	00ÿ002	aguantar	
+1211		democracia	-3001	00ÿ002	causar	
-3001	00ÿ002	lograr	-3001	00	003	gritar
+1111		derecho	-3002&	00ÿ002	tener	
-3002	00ÿ103	tener	-3001	00ÿ002	tranquilizar	
-3001	00ÿ103	ganar	+1111		efecto	
-3004	00ÿ103	ganar	-3003	00ÿ002	surtir	
+1121		derechos	+1111		ejemplo	
-3002	00ÿ103	tener	-3003	00ÿ143	seguir	
-3001	00ÿ103	ganar	-3001	00;024	tomar	
-3001	00ÿ103	pisotear	+1222		elecciones	
+1211		derrota	-3001	00;104	votar	
-3001	00ÿ203	causar	+1111		embargo	
-3003	00ÿ203	infligir	-3001	00ÿ103	levantar	
+1111		descanso	+1111		enemigo	
-3002	00ÿ103	conceder	-3001	00'104	aniquilar	
-3001	00ÿ103	privar	-3004	00æ003	lanzar	
+1121		deseos	-3004	00'104	pasar	
-3002	00;003	arder	+1121		empleados	
+1112		desorden	-3001	00ÿ103	nombrar	
-3001	00;003	estar	+1111		empleo	
+1211		destreza	-3001	00ÿ103	desempeñar	
-3003	00×024	sobresalir	+1211		empresa	
-3001	00×024	superar	-3001	00ÿ103	cargar	
+1221		deudas	+1212		enfermedad	
-3002	00ÿ002	contraer	-3001	00ÿ103	evitar	
+1121		días	-3001	00æ104	preservar	
-3002&	00ÿ103	correr	-3001	00	003	preservar
+1121		dientes	+1111		enigma	
-3001	00ÿ103	enseñar	-3001	00ÿ203	asertar	

ANEXO B.2 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS VERBALES

-3001	00ÿ203	dar	+1122	errores
+1211		enseñanza	-3002	00ÿ002
-3001	00ÿ103	apoyar	+1211	escalera
-3001	00'104	enfocar	-3001	00ÿ103
+1111		entusiasmo	-3001	00«104
-3001	00ÿ103	contagiar	-3003	00ÿ103
+1121		equipos	-3003	00«104
-3001	00'104	integrar		

Anexo B.3 Base secundaria de colocaciones con valencias sustantivales

+1222	actividades	-1212	linformación
-1111	0cambio	-1212	lubicación
+1212	actitud	-1222	lunidades
-1111	0cambio	-1121	ldatos
+1111	adiestramiento	-1221	lcolumnas
-1211	lfase	-1211	lconsulta
+1212	administración1	-1121	lcampos
-1121	lservicios	-1211	lsalida
-1211	lventaja	-1111	lregistro
-1221	lventajas	-1111	lángulo
+1112	administrador	+1111	cachorro
-1121	lpermisos	-1112	lcarácter
+1111	adverbio	+1111	campamento
-1212	lfuncionalidad	-1121	lmetros
+1211	agua	+1111	campo
-1112	lnivel	-1200	lsintaxis
+1111	amo	-1211	lparte
-1111	ldominio	-1112	lvalor
+1120	análisis2	-1111	lnombre
-1111	lcosto	+1212	capital1
-1111	lpago	-1111	lnombre
-1111	lresultado	+1211	casa
+1112	ángel	-1211	lpersona
-1111&	0cabello	-1111	ldueño
+1112	animal	+1111	cerebro
-1211	lnaturaleza	-1111	lfuncionamiento
-1212	lsensibilidad	-1212	lfunción
-1111	lnombre	+1111	ciudadano
-1211	lcola	-1122	ldeberes
+1111	año	-1221	lgarantías
-1221	lventas	-1212	lobligación
+1212	aplicación	-1222	lobligaciones
-1113	linterfaz	+1121	ciudadanos
+1111	archivo	-1122	ldeberes
-1111	lcontenido	-1221	lgarantías
-1211	lcontraseña	-1222	lobligaciones
-1212	ldescripción	+1111	cliente
-1212	lextensión	-1222	lobligaciones
-1111	lformato	-1121	lderechos
-1111	lnombre	-1212	lpetición
-1212	lpetición	+1211	computadora
-1212	lubicación	-1212	lconfiguración
+1112	autor	-1211	lparte
-1111	lnombre	-1212	lreparación
+1212	aviación	+1112	conector
-1121	lbombardeos	-1111	lfuncionamiento
+1112	ayer	-1111	larchivo
-1111	lrecuerdo	-1221	lcaracterísticas
+1211	base	-1221	letiquetas

ANEXO B.3 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

+1211	consulta	+1111	formato
-1121	lresultados	-1211	lparte
+1112	controlador	+1111	formulario
-1212	lespecificación	-1122	lcontroles
+1211	correa	-1121	ldatos
-1111	lxtremo	+1212	fracción
-1111	límite	-1111	lpárrafo
+1112	corazón	+1111	gobierno
-1121	lsonidos	-1212	lcredibilidad
+1111	cuello	-1212	lefectividad
-1212	lpiel	-1211	lpolítica
+1111	cuerpo	-1111	lvocero
-1112	lmaterial	+1111	grupo
+1111	diente	-1112	lcontrol
-1111	lxtremo	-1121	lrecursos
-1212	lextracción	+1211	guerra
-1211	lsuperficie	-1111	lcomienzo
-1111	lxtremo	-1112	lfin
-1211	lperdida	+1211	hambre
+1212	dirección	-1112	ldolor
-1121	lelementos	-1111	El sufrimiento
+1211	disciplina	+1210	hipótesis
-1121	lresultados	-1212	ldemostración
+1111	disco	-1111	lplanteamiento
-1222	lcanciones	+1111	hocico
-1111	llado	-1211	lpresa
+1211	distancia	+1211	infantería
-1212	lmitad	-1111	lataque
+1111	documento	-1211	lretirada
-1212	lconstrucción	+1212	información
+1111	empleado	-1111	lcontenido
-1112	lmanual	-1212	lrecuperación
+1111	encabezado	+1222	instalaciones
-1211	lparte	-1111	«lrecorrido
+1111	enemigo	+1212	instrucción
-1111	ælfuego	-1111	lformato
+1111	entrenamiento	-1212	lejecución
-1221	lalturas	-1212	lutilización
-1212	lmodalidad	+1112	instrumental
+1111	equipo	-1212	lesterilidad
-1111	lnombre	+1211	inteligencia
+1212	especificación	-1100	lstatus
-1212	lrevisión	+1212	intervención
+1111	escándalo	-1111	ldía
-1211	lchispa	+1111	juego
-1221	«lburlas	-1212	ltraducción
+1111	éxito	+1211	lengua
-1221	lsemillas	-1222	llesiones
-1111	lcamino	+1211	líneal
+1112	fin	-1111	lparámetro
-1111	lprincipio	+1111	lista
+1111	fuego1	-1111	lelemento
-1211&	0prueba	+1212	libertad

ANEXO B.3 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

-1111	lamante	-1111	0sentimiento
-1111	0clima	+1212	opción
-1211	<1lucha	-1122	lvalores
+1211	mano	+1212	oración1<"lingüística"
-1111	ldorso	-1212	lconstrucción
-1111	lborde	-1211	lparte
+1221	manos	+1212	oración2<"religión"
-1111&	0juego	-1121	lpasos
+1211	masa	+1112	orden
-1122	lvalores	-1111	lsignificado
+1111	método	+1212	organización
-1111	ldetalle	-1121	linformativos
-1211	lestructura	-1121	ldocumentos
-1212	lexplicación	+1222	organizaciones
-1121	lpasos	-1211	lnaturaleza
+1111	miembro	+1111	órgano
-1111	lnombre	-1222	ldimensiones
-1112	lperfil	+1112	origen
+1122	minerales	-1111	lnombre
-1221	lcaracterísticas	-1212	ldefinición
-1212	ldetección	+1211	página
-1212	ldescripción	-1222	lpropiedades
-1212	0distribución	-1212	ldirección
-1222	lpropiedades	+1111	país
-1111	0procesamiento	-1221	láreas
+1111	mundo	-1122	lciviles
-1112	lfin	-1111	loeste
-1221	lpartes	-1222	lciudades
-1121	lpueblos	-1211	lestructura
-1111	lresto	-1111	lpresidente
-1122	lrincones	+1111	paquete
+1211	naturaleza	-1111	ltamaño
-1211	>larmonía	-1111	lpeso
-1111	>lcontacto	+1111	parámetro
-1212	lconservación	-1111	lnombre
-1211	ldefensa	-1112	lvalor
-1121	lelementos	+1111	parque
-1111	lenemigo	-1122	lárboles
-1111	>lenlace	-1222	lflores
-1221	lfuerzas	-1221	lpalomas
-1122	limágenes	+1111	pasado
-1212	lmanifestación	-1111	ltema
-1221	lmaravillas	-1121	lhechos
-1122	lpostales	+1213	paz
-1212	lprotección	-1111	lamante
-1221	lreservas	+1111	pensamiento
-1121	lsonidos	-1111	lfuncionamiento
+1111	odio	+1111	perro
-1112	'lamor	-1111	lánimo
-1211	elbatalla	-1222	ltransgresiones
-1221	lcausas	-1212	latención
-1212	0expresión	-1111	ltraseo
-1211	lpresa	-1211	lconfianza

ANEXO B.3 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

-1111	lmovimiento	-1211	lcopia
-1212	lreacción	-1212	lcreación
+1111	pista	+1211	ruta
-1111	linicio	-1211	lbúsqueda
+1111	poblado	-1212	lcontinuación
-1121	lhabitantes	-1122	ldetalles
+1111	programa	-1212	ldeterminación
-1122	lejemplares	-1221	letapas
-1111	lnombre	-1212	levaluación
+1111	protocolo	-1111	lextremo
-1212	lmonitorización	-1111	linicio
-1212	lversión	-1112	lfinal
+1211	prueba	-1111	lmapa
-1212	ldescripción	-1111	lpunto
-1121	lejemplos	-1111	lrecorrido
-1212	lexplicación	-1111	lresto
-1211	limportancia	+1212	sección
-1111	lproceso	-1211	lcoordinadora
-1121	lresultados	-1111	líndice
+1212	publicación	-1212	linformación
-1111	lavance	-1111	linicio
-1221	lcaracterísticas	-1111	lejemplo
-1111	lcontenido	-1111	ljefe
-1111	ldesarrollo	-1121	lobjetivos
-1122	lejemplares	-1111	lorganigrama
-1111	lfracaso	-1112	lpersonal
-1211	lhistoria	+1211	serie
-1111	líndice	-1111	lnúmero
-1111	lmotivo	-1121	lproductos
-1111	lretraso	-1111	ltranscurso
-1122	lresponsables	-1211	lsecuencia
-1221	lventajas	+1112	servidor
+1111	pueblo	-1111	lnombre
-1111	lcanto	-1122	lvariables
-1111	lcorreo	+1111	sistema
-1112	ldefensor	-1121	larchivos
-1221	lfamilias	-1121	ldatos
-1121	lhabitantes	-1111	lestado
-1121	lhéroes	-1111	lnombre
-1111	lteatro	-1212	lpresentación
-1111	lpartido	+1212	sociedad
-1213	lvoz	-1121	laspectos
+1211	puerta	-1211	lbase
-1111	ldiseño	-1111	lcambio
-1211	lmanija	-1222	lobligaciones
-1111	lmarco	+1111	sonido
+1212	red	-1211	limportancia
-1212	lexploración	-1212	linformación
-1111	lequipo	-1112	lorigen
-1212	linformación	-1111	lorigen
-1121	lusuarios	+1121	sonidos
-1121	lpaquetes	-1111	larte
+1111	registro	-1212	larticulación

ANEXO B.3 BASE SECUNDARIA DE COLOCACIONES CON VALENCIAS SUSTANTIVALES

-1212	<i>lamplificación</i>	-1221	<i>lcaracterísticas</i>
-1211	<i>lcaracterística</i>	-1222	<i>lcondiciones</i>
-1112	<i>lcreador</i>	-1212	<i>lconfirmación</i>
-1212	<i>ldescripción</i>	-1121	<i>ldatos</i>
-1212	<i>ldeterminación</i>	-1122	<i>ldetalles</i>
-1111	<i>lestudio</i>	-1212	<i>lduración</i>
-1212	<i>lgrabación</i>	-1211	<i>lfecha</i>
-1211	<i>lmezcla</i>	-1111	<i>limporte</i>
-1121	<i>lmundos</i>	-1111	<i>lpago</i>
-1112	<i>lorigen</i>	-1111	<i>lperíodo</i>
+1212	<i>suscripción</i>	-1111	<i>lprecio</i>
-1212	<i>lampliación</i>	-1212	<i>lrenovación</i>
-1212	<i>lcancelación</i>	-1111	<i>lvencimiento</i>

Anexo B.4 Base secundaria de hipónimos / hiperónimos

+1121adornos	-0000bebida caliente
-1121anillos	-0000bebida fuerte
-1121aretes	-1121refrescos
-1122collares	+0000bebida caliente
-1121pendientes	-1111cacao
+1122animales	-1111café
-0000animales domésticos	-1111chocolate
-1121felinos	-1111té
-1122reptiles	+0000bebida fuerte
-1122roedores	-1112champán
+0000animales domésticos	-1112licor
-1111cerdo	-1112ron
-1111ganado	-1111vodka
-1111gato	-1112whisky
-1111pájaro	+1121refrescos
-1111perro	-1211Coca-cola
+1121antibióticos	-1211Fanta
-1211eritromicina	-1211Mirinda
-1211estreptomicina	-1200Pepsi Cola
-1211penicilina	+0000casa habitacional
+1111aparato1	-1111apartamento
-1211aspiradora	-1211cabaña
-1112iluminador	-1211casa
-1211lavadora	-1211choza
-1112refrigerador	+0000categoría gramatical
-1112tostador	-1111caso
-1112ventilador	-1111género
+1112autor	-1111modo
-1111crítico	-1111número
-1112escritor	+1122cereales
-1111periodista	-1113arroz
-1111poeta	-1211avena
+1111baile	-1211cebada
-1112charleston	-1111trigo
-1211cumbia	+1211ciencia
-1211lambada	-0000ciencia de la humanidad
-1211polca	-0000ciencia exacta
-1211samba	-0000ciencia médica
-1111tango	-0000ciencia natural
+1221bayas	-0000ciencia técnica
-1111arándano	+0000ciencia de la humanidad
-1211cereza	-1211antropología
-1211frambuesa	-1211aequeología
-1211fresa	-1211economía
-1211grosella	-1211filología
-1211grosella	-1211filosofía
-1211morera	-1211historia
-1211zarza	-1211jurisprudencia
+1221bebidas	-1111periodismo

ANEXO B.4 BASE SECUNDARIA DE HIPÓNIMOS / HIPERÓNIMOS

+0000ciencia exacta	-1211leña
-1211cibernética	-1111petroleo
-1211física	-1112querosén
-1211informática	+1212constelación
-1211matemática	-1111Cangrejo
+0000ciencia natural	-1112Dragón
-1211biología	-1211Galaxia
-1211botánica	-1121Gemelos
-1211genética	-1111Pegaso
-1211geología	-1111Virgo
-1211geografía	+1222construcciones
-1211química	-0000casa habitacional
+0000ciencia técnica	-1111castillo
-1211electrónica	-1112hotel
-1211informática	-1111kiosco
-1211metalurgia	-1112pabellón
-1211minería	-1211residencia
+1121cítricos	-1111teatro
-1122limones	+1111continente
-1221mandarinas	-1211África
-1221naranjas	-1211América
-1221toronjas	-1211Asia
+1112color	-1211Australia
-1111amarillento	-1211Europa
-1111amarillo	+1211coordenada
-1112ambar	-1211abscisa
-1111anaranjado	-1212elevación
-1112azul	-1111rango
-1111blanco	+1111cuadro
-1111castaño	-1111grabando
-1211crema	-1111paisaje
-1111durazno	-1111retrato
-1211esmeralda	+0000cuerpo gobernante
-1000índigo azul	-1212administración
-1211lavanda	-1111directorío
-1112limón	-1111gobierno
-1211mostaza	+1111deporte
-1111mulato	-1211acrobatía
-1111negro	-1111atletismo
-1211paja	-1211gimnasia
-1211purpura	-1000lucha libre
-1111rojo	-1212natación
-1211rosa	+1111deportista
-1211turquesa	-1112corredor
-1111verde	-1112disparador
-1111verdoso	-1111futbolista
-1211violeta	-1111gimnasta
+1111combustible	-1112luchador
-1111aceite	-1112nadador
-1112alcohol	-1112saltador
-1111carbono	-1111tenista
-1121gas	+0000día de semana
-1211gasolina	-1111domingo

ANEXO B.4 BASE SECUNDARIA DE HIPÓNIMOS / HIPERÓNIMOS

-1100jueves	-1211venérea
-1100lunes	+1211escuela
-1100martes	-1211academia
-1100miércoles	-1111instituto
-1111sábado	-1212universidad
-1100viernes	+1111especialista
+1111diplomático	-1111antropólogo
-1112cónsul	-1111arqueólogo
-1111delegado	-1111artista
-1112embajador	-1111banquero
+0000dispositivo óptico	-1111biólogo
-1111microscopio	-1111científico
-1111prismático	-1111economista
-1111telescopio	-1111físico
+1111documento	-1111geógrafo
-1111acuerdo	-1111geólogo
-1212aplicación	-1112historiador
-1111certificado	-1111lingüista
-1111contrato	-1111maestro
-1212declaración	-1111matemático
-1212declaración	-1111médico
-1111decreto	-1112profesor
-1211demanda	-1112programador
-1111diploma	-1111químico
-1212disertación	-1111sicólogo
-1111informe	+1212estación
-1111inventario	-1111invierno
-1211licencia	-1111otoño
-1111protocolo	-1211primavera
-1111registro	-1111verano
-1111tratado	+1111estrato
+1212edad	-1121campesinos
-1211adolescencia	-1000clase media
-1211infancia	-1121empleados
-1212juventud	-1121gerentes
-1213niñez	-1122intelectuales
+0000empresa de comida	-1121obreros
-1112bar	+1121estremos
-1211cafetería	-1111máximo
-1211pizzeria	-1111mínimo
-1111restaurante	-1111óptimo
+1212enfermedad	+1121felinos
-1211apoplejía	-1111gato
-1211asma	-1112león
-1112cáncer	-1111leopardo
-1211cólera	-1111lince
-1211disentería	-1221panteras
-1211gripe	-1111puma
-1212hipertensión	-1111tigre
-1112hipotension	+1111fenómeno
-1111infarto	-1211calma
-1111sida	-1111eclipse
-1200tuberculosis	-1212sequedad

ANEXO B.4 BASE SECUNDARIA DE HIPÓNIMOS / HIPERÓNIMOS

-1212tempestad	-1111congreso
-1111terremoto	-1212convención
-1111tiempo	-1212reunión
-1211tormenta	-1111simposio
-1111viento	-1112taller
+0000fenómeno social	+1221frutas
-1111desarmamento	-1221bayas
-1111desempleo	-1221ciruelas
-1111estancamiento	-1121cítricos
-1212evolución	-1121duraznos
-1212movilización	-1221manzanas
-1212privatización	-1221peras
-1212prosperidad	-1211piña
-1212reacción	-1121plátanos
-1222represiones	-1221uvas
-1111retroceso	+0000fuerzas armadas
-1212revolución	-1221acorazadas
+0000figura de tres dimensión	-1211armada
-1111cilindro	-1000fuerza aérea
-1111cono	-1211infantería
-1111elipsoide	+0000función geométrica
-1111poliedro	-1211cosecante
+1211filología	-1111coseno
-1211lingüística	-1211cotangente
+1211flora	-1111seno
-1112árbol	-1111tangente
-1111arbusto	+1121gas
-1112cereal	-1111aire
-1212césped	-1111amoníaco
-1222flores	-1112argón
+1222flores	-1111hidrógeno
-1221azucenas	-1111metano
-1221callas	-1112neón
-1221dalias	-1111nitrógeno
-1221gardenias	-1111oxígeno
-1221gladiolas	+1211gastronomía
-1221lilas	-1121huevos
-1221manzanillas	-1111queso
-1221margaritas	-1211salchicha
-1221rosas	+1111género1
-1122tulipanes	-1111femenino
-1221violetas	-1111masculino
+1111fondo	-1111neutro
-1111certificado	+1111género2
-1111cheque	-1111boceto
-1111crédito	-1111ensayo
+0000forma instructiva	-1112folletín
-1211conferencia	-1211historia
-1212exposición	-1221memorias
-1111libro	-1111narrdtive
-1111seminario	-1211novela
+1111foro	-1211novella
-1211conferencia	-1111poema

ANEXO B.4 BASE SECUNDARIA DE HIPÓNIMOS / HIPERÓNIMOS

-1211poesía	-1211romana
+1212guarnición	+1000inferencia lógica
-1113arroz	-1212deducción
-1211pasta	-1212inducción
-1000patatas fritas	+0000instrumento de mesa
-1111puro	-1111azucarero
+1111idioma	-1211cuchara
-1111Afgano	-1111cuchillo
-1111Albanés	-1111salero
-1112Alemán	-1111servilletero
-1111Árabe	-1112tenedor
-1111Armenio	+0000intervalo de tiempo
-1112Azerbaijan	-1111año
-1111Beloruso	-1211década
-1111Birmano	-1111día
-1111Búlgaro	-1211época
-1112Catalán	-1211hora
-1111Checo	-1121mes
-1111Chino	-1111milenio
-1111Coreano	-1111minuto
-1111Eslovaco	-1111segundo
-1112Español	-1211semana
-1111Estonio	-1111siglo
-1111Finlandés	+0000intrumento musical
-1111Flamenco	-1112acordeón
-1111Francés	-1111clarinete
-1111Griego	-1111contrabajo
-1111Hebreo	-1211flauta
-1111Holandés	-1211guitarra
-1111Húngaro	-1111piano
-0000idioma artificial	-1112saxofón
-0000idioma extranjero	-1112tambor
-1111Inglés	-1211trompeta
-1111Islandés	-1112violín
-1111Italiano	-1111violoncello
-1111Latino	+1111jefe
-0000lenguaje de programación	-1111decano
-1111Lituano	-1111diputado
-1112Mongol	-1112director
-1111Noruego	-1111gerente
-1111Polaco	-1111ministro
-1111Portugués	-1111presidente
-1111Rumano	-1112rector
-1111Ruso	+1211joyería
-1111Serbio	-1121brillantes
-1111Slovenio	-1112collar
-1111Sueco	+1211jurisprudencia
-1111Turco	-1111abogado
-1111Vasco	-1113juez
+1221iglesias	+1112limpiador
-1211católica	-1111champoo
-1211griega	-1111detergente
-1211ortodoxa	-1112jabón

ANEXO B.4 BASE SECUNDARIA DE HIPÓNIMOS / HIPERÓNIMOS

+1211línea
-1211curva
-1212espiral
-1211hipérbola
-1211ondulada
-1211parábola
-1211recta
+1111líquido
-1111ácido
-1211agua
-1111alcalino
+1112lugar
-1212ciudad
-1111pueblo
-1111suburbio
+1112mar
-1100Mar Adriático
-1100Mar Báltico
-1100Mar Blanco
-1100Mar Caribeño
-1100Mar Mediterráneo
-1100Mar Negro
-1100Mar Rojo
+1111marinero
-1111almirante
-1112capitán
-1111compañero
-1111contraestre

Anexo I. Preposiciones

```
#####  
#  
## Centro de Investigación en Computación - IPN  
## Autor : Sabino Miranda  
## Archivo: HUPREPO.txt  
## Creado : 10-SEP-2002  
## Descripción:  
## Este archivo contiene las preposiciones extendidas para español  
## el orden es descendente debido a que algunas preposiciones se veri-  
ficarían  
## antes que otras con características similares  
## (Obs)  
## Ejemplo:  
##     de  
##     de cara a  
##     de parte de  
## Correcto:  
##     de parte de  
##     de cara a  
##     de  
##  
#####  
###  
1. tras de  
2. tras  
3. tocante a  
4. sobre  
5. so pretexto de  
6. so pena de  
7. so capa de  
8. so  
9. sin  
10. según  
11. salvo  
12. pro  
13. por medio de  
14. por lo que respecta a  
15. por intermedio de  
16. por entre  
17. por encima de  
18. por debajo de  
19. por cuenta de  
20. por conducto de  
21. por arriba de  
22. por amor de  
23. por amor a  
24. por  
25. pese a que  
26. para con  
27. para  
28. menos  
29. mediante  
30. luego de  
31. lejos de  
32. junto con  
33. junto a  
34. incluso  
35. hasta  
36. hacia  
37. gracias a  
38. fuera de  
39. frente a  
40. extra  
41. excepto  
42. entre  
43. enfrente de  
44. encima de  
45. en virtud de  
46. en vez de  
47. en torno de  
48. en torno a  
49. en relación con  
50. en razón de  
51. en razón a  
52. en pro de
```

ANEXO I. PREPOSICIONES

53. en pos de
54. en orden a
55. en obsequio de
56. en obsequio a
57. en nombre de
58. en mitad de
59. en memoria de
60. en medio de
61. en materia de
62. en lugar de
63. en los límites de
64. en igual de
65. en gracia a
66. en función de
67. en fuerza de
68. en favor de
69. en dirección a
70. en descargo de
71. en derredor de
72. en defecto de
73. en cuanto a
74. en consideración a
75. en concordancia con
76. en cambio de
77. en calidad de
78. en atención a
79. en aras de
80. en
81. durante
82. donde
83. detrás de
84. después de
85. desde
86. dentro de
87. delante de
88. debido a
89. debajo de
90. de parte de
91. de frente a
92. de entreguerras
93. de entre
94. de cara a
95. de acuerdo con
96. de
97. contra
98. conforme a
99. con vista a
100. con respecto a
101. con referencia a
102. con motivo de
103. con lo dispuesto en
104. con idea de
105. con honores de
106. con fundamento en
107. con el fin de que
108. con el fin de
109. con cargo a
110. con carácter de
111. con arreglo a
112. con
113. comparativamente a
114. cerca de
115. cara a
116. cabe
117. bajo
118. apud
119. aparte de
120. antes de
121. ante
122. amén de
123. allende
124. alrededor de
125. al nivel de
126. al modo de
127. al lado de
128. al frente de
129. al filo de
130. al decir de
131. al calor de
132. al cabo de
133. al borde de
134. adonde
135. además de
136. ad
137. acerca de
138. abajo de
139. ab
140. a través de
141. a título de
142. a ruego de
143. a retaguardia de
144. a reserva de
145. a razón de
146. a raíz de
147. a propósito de
148. a pesar de
149. a partir de
150. a modo de
151. a merced de
152. a manera de
153. a lo largo de
154. a la vuelta de
155. a la vista de
156. a la manera de
157. a la altura de
158. a instancias de

ANEXO I. PREPOSICIONES

159.a instancia
160.a fuerza de
161.a flor de
162.a fin de que
163.a fin de
164.a favor de
165.a expensas de
166.a eso de
167.a diferencia de

168.a despecho de
169.a cuenta de
170.a costa de
171.a consecuencia de
172.a causa de
173.a cargo de
174.a cambio de
175.a base de
176.a

255 *No existe preposición

Anexo II. Adverbios

```
#####  
## Centro de Investigación en Computación - IPN  
## Autor : Sabino Miranda  
## Archivo: HUAdv.txt  
## Creado : 10-Dic-2001  
## Descripción:  
## Este archivo contiene todos los adverbios:  
## lugar, tiempo, cantidad, modo, afirmativos, negativos, duda y cantidad.  
## Las locuciones adverbiales se determinan en el programa son las construcciones adver-  
biales.  
#####  
### Adverbios de lugar  
#####  
#  
abajo                aparte                doquier  
acá                  aquí                 doquiera  
acullá               arriba               encima  
adelante             atrás                enfrente  
adentro              cerca                extramuros  
adónde              debajo              fuera  
afuera               delante              intramuros  
ahí                  dentro               junto  
alrededor            detrás               lejos  
allá                 dónde  
allí                 dondequiera  
#####  
## Adverbios de tiempo  
#####  
ahora                cuando quiera       mientras  
ahorita              después              nunca  
anoche               entonces             pasado mañana  
antaño               entremedias         pronto  
anteayer             entremedio           siempre  
antes                hogaño              tarde  
aún                  hoy                  temprano  
ayer                 luego                todavía  
cuando               mañana               ya  
#####  
## Adverbios de cantidad
```

ANEXO II. ADVERBIOS

#####

#

algo	excepto	nada
bastante	harto	poco
casi	más	salvo
cuan	medio	sobremanera
cuán	menos	tan
cuánto	mero	tanto
cuasi	mucho	
demasiado	muy	

#####

Adverbios de modo

#####

adrede	claro	incluso
anteanoche	cual	mal
apenas	deprisa	máxime
aposta	despacio	presto
aprisa	enhoramala	recién
así	enseguida	sólo
bajo	exclusive	tal
bien	expreso	verbigracia
como	gratis	
como quiera	inclusive	

#####

Adverbios de afirmación

#####

#

asimismo	sí	también
cierto	siquiera	viceversa

#####

Adverbios de negación

#####

#

jamás	ni siquiera	tampoco
nanay	no	
ni	nunca jamás	

#####

Adverbios de duda

#####

acaso
quizá
quizás

ANEXO II. ADVERBIOS

```
#####  
## Adverbios diversos  
#####  
mediante
```

Anexo III. Archivos Auxiliares

Los archivos auxiliares están compilados en este anexo. Cada archivo inicia con su respectivo encabezado seguido de su contenido.

```
#####
## Centro de Investigación en Computación - IPN
## Autor : Sabino Miranda
## Archivo: HUAPOCO.txt
## Creado : 6-May-2002
## Descripción:
## Este archivo contiene las palabras apócopas
##
#####
algún                cualquier                tercer
buen                 gran                    san
mal                  ningún
cien                 primer
#####
## Centro de Investigación en Computación - IPN
## Autor : Sabino Miranda
## Archivo: HUEADJ.txt
## Creado : 10-Dic-2001
## Descripción:
## Este archivo contiene la lista de palabras que son excepciones a la regla general
## para identificar adjetivos tales como: clemente, vehemente, etc.
##
#####
# SUFIJO -ado
#####
agrado                condado                entarimado
alumnado              consulado              entorchado
apostolado           contrapicado          entramado
arado                 costado                envarado
archiducado          dado                   episcopado
arcontado            damasquinado          estado
arzobispado          dechado               estrado
bocado               desagrado              estudiantado
brazado              desaguisado           externado
brocado              desenfado              fado
calafateado          ducado                fotograbado
campesinado          electorado            grado
carbonado            emparrado             granado
catecumenado         emperchado            hado
cayado               empiñonado            hígado
collado              enfado                hipermercado
```

ANEXO III. ARCHIVOS AUXILIARES

huecograbado	pescado	salariado
lado	pirograbado	secretariado
legado	pontificado	senado
lenguado	prado	sotechado
mantecado	prelado	supermercado
marquesado	primado	tablado
matriarcado	principado	teclado
mercado	proconsulado	tejado
noviciado	profesorado	terrado
obispado	proletariado	tinglado
palatinado	protectorado	vado
papado	puñado	venado
párpado	recado	voluntariado
patriarcado	rectorado	campesinado
pecado	reinado	ganado
pensionado	sábado	

#####

SUFIJO -ido

#####

ácido	dióxido	olvido
alarido	ejido	óxido
aminoácido	estallido	péptido
anhídrido	estampido	peróxido
apellido	fosfolípido	pitido
aullido	glúcido	pórfido
balido	glucósido	prótido
berrido	graznido	quejido
bióxido	hidrácido	resoplido
bólido	hidróxido	ronquido
bramido	hipido	ruido
bufido	infrasonido	sacárido
chasquido	ladrido	sarpullido
chiflido	libido	silbido
chillido	malentendido	sonido
chirrido	marido	soplido
colorido	maullido	tronido
contrasentido	monóxido	ultrasonido
crujido	mugido	vagido
descuido	nido	zumbido
despido	nucleótido	

#####

SUFIJO -nte

#####

accidente	aliciente	arconte
agente	almirante	atlante
aguante	ambiente	bacante
aguardiente	apunte	barrunte

bergante	gente	poniente
bisonte	gradiente	punte
bogavante	guante	recipiente
cabrestante	horizonte	relente
cante	iguanodonte	remanente
cociente	inconveniente	remonte
comandante	ingrediente	repente
contingente	levante	rinoceronte
contracorriente	lobagante	rocinante
cosecante	lubigante	semblante
cotangente	mastodonte	serpiente
desmante	mente	sextante
desplante	montante	simiente
detergente	monte	sobrestante
diamante	nepente	somonte
diente	occidente	subafluente
elefante	oriente	subconsciente
ente	oropimente	talante
estante	pescante	tinte
excipiente	pespunte	torrente
expediente	pedemonte	trasplante
frente	plante	tridente
fuelle	polizonte	turbante

#####

SUFIJO -ustre -estre

#####

aligustre
 balaustre
 bimestre
 burgomaestre
 contraestra
 lustre
 maestre
 semestre
 trimestre

#####

Centro de Investigación en Computación - IPN

Autor : Sabino Miranda

Archivo: HUEMASC1.txt

Creado : 10-Dic-2001

Descripción:

##

Este archivo contiene la lista de palabras que son excepciones al paradigma propuesto para

identificar el género masculino tales como: flor, foto, polio, mano, rima, prima, pluma, etc.

##

ANEXO III. ARCHIVOS AUXILIARES

GÉNERO MASCULINO Excepciones FEMENINAS

Considerado: MASCULINO

Excepcion: FEMENINAS

##

#####

#####

SUFIJO -ma

#####

aguagoma	doma	merma
alarma	dracma	mojama
alharma	enjalma	morisma
alhucema	epifonema	norma
alisma	escama	palma
alma	esgrima	paloma
ama	espuma	pamema
amalgama	estima	pantomima
ánima	estratagema	plataforma
antefirma	fama	pluma
apostema	firma	pócima
apotema	flama	poma
arma	flema	postema
asma	forma	prima
bizma	gama	proclama
broma	gamma	quema
bruma	gema	rama
caguama	goma	redoma
calma	gomaespuma	reforma
cama	grama	resma
carcoma	grima	retama
cataplasma	horma	rima
catagorema	lágrima	saloma
chusma	lama	séptima
cima	lástima	sigma
coima	legítima	sima
coma	lima	soflama
contrarreforma	llama	suma
crema	loma	tarima
crisma	lúcuma	toma
cuaresma	luma	trama
dama	mama	víctima
décima	marisma	yema
derrama	maroma	zalema
diadema	máxima	zeugma

#####

SUFIJO -o

#####

mano

foto
mano
moto
polio

#####

SUFIJO -e

#####

aeronave	efeméride	madre
aguachirle	efigie	mansedumbre
aguanieve	elipse	masacre
altiplanicie	elite	medianoche
anafase	esfínges	ménade
apócope	especie	mente
argemone	estirpe	metafase
ascáride	falange	miente
astronave	faringe	mole
ave	fase	molicie
bacante	fe	muchedumbre
barbarie	fiebre	muerte
base	frase	mugre
birreme	frente	nave
bocacalle	fuelle	náyade
boutique	galliforme	nieve
calígine	gente	noche
calle	gradiente	nube
calvicie	gripe	nurse
canicie	hambre	paragoge
cariátide	hecatombe	parte
cariópside	hélice	pepónide
carne	hemorroide	pesadumbre
catástrofe	herrumbre	peste
cédride	higiene	pirámide
chinche	hipérbole	planicie
cicloide	hueste	plebe
clámide	incertidumbre	pléyade
clase	índole	podredumbre
clave	inglete	pose
cohorte	intemperie	pringue
comadre	landre	probóscide
contracorriente	laringe	profase
cosecante	leche	progenie
costumbre	legumbre	prole
cotangente	libidine	psique
cumbre	liebre	quejumbre
cúspide	liendre	raigambre
diástole	llave	reciedumbre
duramadre	lumbre	roulotte

ANEXO III. ARCHIVOS AUXILIARES

sangre	sinécdoque	telofase
sede	sinusoide	tilde
semiconsonante	siringe	toilette
serie	sístole	torre
serpiente	subespecie	tournée
servidumbre	suerte	trirreme
sierpe	suite	ubre
sílfide	superficie	urbe
silice	tarde	urdimbre
simiente	techumbre	varice

SUFIJO -or

coliflor
flor
labor
prensor
sor

SUFIJO -os

tos

SUFIJO -l

aguamiel coniferal semifinal
bacanal diagonal semivocal
besamel espiral señal
cal factorial sucursal
cañamiel hiel versal
capital miel vertical
cárcel multinacional vestal
catedral piel vocal
col sal central

Centro de Investigación en Computación - IPN

Autor : Sabino Miranda

Archivo: HUNUMCOM.txt

Creado : 10-Dic-2001

Descripción:

Este archivo contiene la lista de palabras que son de número común tales como: análisis,

isósceles, judas, lunes, páncreas, tenis, etc.

##

alias análisis cascarabias

corta cabezas
 déficit
 guardacostas
 guardaespaldas
 guardagujas
 jueves
 lanzagranadas
 lanzallamas
 lavafrutas
 limpiabotas
 limpiaparabrisas
 lunes
 manitas
 martes
 matamoscas
 mataratas

mecenas
 mesías
 miércoles
 milhojas
 montacargas
 páncreas
 parabrisas
 paracaídas
 paraguas
 pasamontañas
 perdonavidas
 pesacartas
 portálamparas
 portaminas
 portamonedas
 quitamanchas

rabioles
 ras
 recogepelotas
 rompecabezas
 rompeolas
 rubiales
 sacapuntas
 salvavidas
 tenis
 trabalenguas
 viejales
 viernes
 vivales
 beige
 bíceps

-os
 #####

albatros
 alisios
 amnios
 añicos
 argos
 bártulos
 bentos
 besamanos
 burdeos
 buscapleitos
 buscarruidos
 cachelos
 calabobos
 calzonazos
 caos
 catavinos
 comecocos
 comicios
 correcaminos
 cortacircuitos
 cosmos
 cuentakilómetros
 cumpleaños
 engañabobos
 equiúridos
 escurreplatos

espantapájaros
 etnos
 extramuros
 friegaplatos
 friegasuelos
 guardabarros
 guarda frenos
 intramuros
 lanzatorpedos
 lavamanos
 lavaplatos
 lejos
 matamoros
 matasanos
 matasellos
 metemuertos
 nemertinos
 pararrayos
 pelagatos
 pertrechos
 picamaderos
 picapleitos
 pinchadiscos
 polisacáridos
 portafolios
 portahelicópteros

portalibros
 portaobjetos
 portarretratos
 posavasos
 pronaos
 quebrantahuesos
 quevedos
 quitamiedos
 rascacielos
 rompehielos
 sacabocados
 sacacorchos
 sacadineros
 sacaromicetos
 sacatrapos
 salvabarros
 secamanos
 soplamocos
 tapacubos
 taparrabos
 tétanos
 tocadiscos
 trepatroncos
 trotaconventos
 trotamundos
 zampabollos

#####

ANEXO III. ARCHIVOS AUXILIARES

```
## Centro de Investigación en Computación - IPN
## Archivo: HUNUMCOM.txt
## Creado : 10-Dic-2001
## Descripción:
## Este archivo contiene la lista de palabras que son de número común para adjetivos tales
## como: isósceles, judas, perdonavidas, etc.
#####
cascarabias
corta cabezas
guardacostas
guardaespaldas
guardagujas
isósceles
judas
mataratas
mecenas
milhojas
narizotas
páncreas
perdonavidas
viejales
vivales
ultravioleta
clave
beige
```

```
#####
## Centro de Investigación en Computación - IPN
## Autor : Sabino Miranda
## Archivo: HUEGENCOM1.txt
## Creado : 10-Dic-2001
## Descripción:
## Este archivo contiene la lista de palabras femeninas que son excepciones al paradigma
## propuesto para identificar el género común tales como: vista, pista, lista, etc.
##
#####
## SUFIJO -ista femeninos
#####vista
pista
lista
#####
## Centro de Investigación en Computación - IPN
## Autor : Sabino Miranda
## Archivo: HUEGENCOM2.txt
## Creado : 10-Dic-2001
## Descripción:
## Este archivo contiene la lista de palabras masculinas que son excepciones al paradigma
## propuesto para identificar el género común tales como: cliente, accidente, etc.
##
#####
## SUFIJO -ente masculinos
#####
cliente
accidente

#####
## Centro de Investigación en Computación - IPN
## Autor : Sabino Miranda
## Archivo: HUADJ.txt
## Creado : 10-Dic-2001
## Descripción:
## Este archivo contiene los adjetivos demostrativos, posesivos, etc.
##
#####
mío                nuestro                ese
mía                nuestra                esa
tuyo                vuestro                aquel
tuya                vuestra                aquella
suyo                este
suya                esta
```

Glosario de términos

Artículo del diccionario	Conjunto de palabras (una palabra clave y múltiples palabras relacionadas) que se encuentran en un archivo de texto.
Base primaria o archivo primario	Conjunto de artículos del diccionario en un archivo de texto.
Base secundaria o archivo secundario	Conjunto de artículos del diccionario formateados en un archivo de texto.
Caracterización morfológica	Proceso que obtiene los rasgos morfológicos y semánticos de los componentes de las colocaciones.
Clase de inclinación	Conjunto de posibles desinencias de una palabra.
Colocación	Combinación de dos palabras adyacentes, o a través de una palabra funcional, que aparecen en los textos con alta frecuencia.
Colocado	Palabra que se combina con otra considerada como la palabra clave.
Corpus	Colección de piezas del lenguaje que son seleccionadas y ordenadas de acuerdo a criterios explícitos lingüísticos para utilizarse como muestras del lenguaje.
DEC	Diccionario Explicativo Combinatorio.
Formateo	Proceso que transforma un archivo de entrada con colocaciones a una estructura secundaria con las posiciones morfológicas y semánticas correspondientes de cada componente de la colocación.
MTT	Meaning \leftrightarrow Text Theory (Teoría Significado \leftrightarrow Texto).
MB	Unidad de medida (Mega Bytes).
Palabra base	Véase palabra clave.
Palabra clave	Palabra considerada como el encabezado del artículo del diccionario.
Palabra relacionada	Véase colocado.
RAM	Memoria de acceso aleatorio.

Tesoro	Diccionario muy grande jerarquizado.
Usuario	Persona que opera el sistema.
WordNet	Base de datos muy grande con relaciones semánticas entre palabras.
Utilería	Programa que realiza un tarea específica.