



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

---

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE FILOSOFÍA Y LETRAS

DIVISIÓN DE ESTUDIOS DE POSGRADO

ESTRUCTURACIÓN SEMÁNTICO - PRAGMÁTICA DEL LÉXICO EN  
DOMINIOS RESTRINGIDOS PARA SISTEMAS DE RECUPERACIÓN  
DE INFORMACIÓN

TESIS QUE PARA OBTENER EL GRADO DE  
MAESTRO EN LINGÜÍSTICA HISPÁNICA  
PRESENTA

ANTONIO REYES PÉREZ

ASESOR: DR. GERARDO SIERRA MARTÍNEZ

CIUDAD UNIVERSITARIA 2006



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**A LA MEMORIA DE MI PAPÁ  
Y A LA FUERZA Y APOYO DE MI MAMÁ, MI FAMILIA Y MI CHAVA**

## AGRADECIMIENTOS

Mis más sinceras gracias a todos los que de una u otra forma han participado en este proyecto. A Gerardo por ser, además de un tutor, un amigo. A los miembros del GIL por todo el tiempo que hemos compartido. A los doctores Chantal Melis, Sergio Ibáñez, Carmen Curcó y Alfonso Medina por aceptar participar y mejorar este trabajo con sus observaciones. A mi familia que siempre ha estado conmigo y, en especial, a esa persona que siempre estuvo a mi lado confiando y apoyando. Gracias.

Gracias también al patrocinio de CONACYT.

# ÍNDICE GENERAL

---

## **INTRODUCCIÓN** **9**

---

<b>0.1 ANTECEDENTES</b>	<b>9</b>
0.1.1 EL PUNTO DE PARTIDA	10
0.1.2 DELIMITACIÓN	11
<b>0.2 OBJETIVO GENERAL</b>	<b>12</b>
0.2.1 OBJETIVOS PARTICULARES	13
<b>0.3 HIPÓTESIS</b>	<b>13</b>
<b>0.4 METODOLOGÍA</b>	<b>14</b>
0.4.1 JUSTIFICACIÓN DEL LÉXICO DE LA INVESTIGACIÓN	14
0.4.2 MARCO TEÓRICO	15
<b>0.5 ORGANIZACIÓN</b>	<b>16</b>

## **SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN** **18**

---

<b>1.1 EL PROBLEMA DE LA INMENSIDAD DE LA INFORMACIÓN TEXTUAL</b>	<b>18</b>
1.1.1 QUÉ ES UN SRI	19
1.1.2 IMPORTANCIA DEL ANÁLISIS LINGÜÍSTICO EN LOS SRI	20
1.1.3 EL MÓDULO LINGÜÍSTICO EN LOS SRI	21
<b>1.2 BASES DE CONOCIMIENTO LÉXICO</b>	<b>22</b>
1.2.1 EJEMPLO DE UNA BCL	23
1.2.2 LA BCL DEL PROYECTO	24
<b>1.3 EL PROYECTO DEBO</b>	<b>24</b>
1.3.1 LA BCL DEL DEBO	25
1.3.2 ESTRUCTURA LÉXICA DE LA BCL DEL DEBO	26
<b>1.4 ALGORITMO DE CLUSTERING</b>	<b>27</b>
1.4.1 UNA HERRAMIENTA PARA GENERAR PARES SEMÁNTICOS DE FORMA AUTOMÁTICA	27

## **EL CORPUS DE ANÁLISIS** **30**

---

<b>2.1 GENERACIÓN AUTOMÁTICA DE PARES SEMÁNTICOS PARA EL CORPUS DE ANÁLISIS</b>	<b>30</b>
2.1.1 DEPURACIÓN DEL CA	32
2.1.1.1 PARES RELACIONADOS LÉXICAMENTE	33

2.1.1.2	PARES SIN RELACIÓN LÉXICA	35
<b>2.2</b>	<b>PROBLEMÁTICA</b>	<b>36</b>
<b>2.3</b>	<b>ANÁLISIS EN DEFINICIONES</b>	<b>37</b>
2.3.1	EQUIVALENCIA CONTEXTUAL	38
2.3.2	MÁS ALLÁ DEL CONTEXTO ESPECÍFICO	39
<b>2.4</b>	<b>CORPUS PRUEBA</b>	<b>40</b>
2.4.1	BWANANET	40
2.4.2	RECUPERACIÓN DE CONTEXTOS	41
2.4.3	SELECCIÓN DE CONSTRUCCIONES VÁLIDAS	42
<b>2.5</b>	<b>EJERCICIO CONTEXTUAL</b>	<b>42</b>
2.5.1	ANÁLISIS DE CONTEXTOS	43
2.5.2	VALORACIÓN CONTEXTUAL	44
<b>2.6</b>	<b>HIPÓTESIS COGNOSCITIVA</b>	<b>46</b>

---

## **ANÁLISIS SEMÁNTICO** **47**

<b>3.1</b>	<b>LA GRAMÁTICA COGNOSCITIVA</b>	<b>47</b>
3.1.1	AXIOMAS	48
<b>3.2</b>	<b>EL MODELO APLICADO AL TRABAJO</b>	<b>48</b>
3.2.1	CONSTRUCTION GRAMMAR	49
3.2.1.1	CONSTRUCCIONES	50
3.2.2	FRAME SEMANTICS	50
3.2.2.1	FRAMES	51
<b>3.3</b>	<b>PREÁMBULO Y DISCUSIÓN</b>	<b>52</b>
<b>3.4</b>	<b>ANÁLISIS</b>	<b>53</b>
<b>3.5</b>	<b>CONCLUSIÓN DEL ANÁLISIS</b>	<b>74</b>
<b>3.6</b>	<b>LA PRAGMÁTICA</b>	<b>75</b>

---

## **EL FACTOR PRAGMÁTICO** **77**

<b>4.1</b>	<b>MÁS ALLÁ DE LA GRAMÁTICA</b>	<b>77</b>
<b>4.2</b>	<b>LA PRAGMÁTICA Y EL PAPEL DEL OYENTE</b>	<b>78</b>
<b>4.3</b>	<b>LA LENGUA EN USO</b>	<b>81</b>
<b>4.4</b>	<b>EL RETO DEL ESTUDIO PRAGMÁTICO EN NUESTRO PROYECTO</b>	<b>82</b>
<b>4.5</b>	<b>LA INTERACCIÓN HUMANO - COMPUTADORA COMO COMUNICACIÓN</b>	<b>83</b>
4.5.1	PROBLEMA	83
4.5.2	LOS HECHOS	84
<b>4.6</b>	<b>TEORÍA DE LA RELEVANCIA</b>	<b>86</b>
<b>4.7</b>	<b>NUESTRA PROPUESTA</b>	<b>88</b>
4.7.1	UN MODELO DE DESCRIPTORES A PARTIR DE TR	89
4.7.2	EL <i>DESCRIPTOR</i>	89
4.7.3	IDENTIFICACIÓN DE DESCRIPTORES	91

4.7.4	EL CONTEXTO	94
4.7.5	RELEVANCIA ÓPTIMA	95
4.8	POSIBLES RESULTADOS	96

---

<b>CONCLUSIONES</b>	<b>100</b>
---------------------	------------

<b>5.1</b>	<b>CONCLUSIONES PARTICULARES</b>	<b>100</b>
5.1.1	CAPÍTULO 1	100
5.1.2	CAPÍTULO 2	101
5.1.3	CAPÍTULO 3	102
5.1.4	CAPÍTULO 4	103
<b>5.2</b>	<b>CONCLUSIONES GENERALES</b>	<b>104</b>
<b>5.3</b>	<b>TRABAJO FUTURO</b>	<b>107</b>

---

<b>BIBLIOGRAFÍA</b>	<b>111</b>
---------------------	------------

<b>APÉNDICE A:</b>	<b>FUNCIONAMIENTO GENERAL DEL ALGCL</b>	<b>118</b>
<b>APÉNDICE B:</b>	<b>PARÁMETROS DE CADA CORRIDA DEL ALGCL</b>	<b>120</b>
<b>APÉNDICE C:</b>	<b>51 PARES DE VERBOS</b>	<b>122</b>
<b>APÉNDICE D:</b>	<b>PARES RELACIONADOS LÉXICAMENTE</b>	<b>123</b>
<b>APÉNDICE E:</b>	<b>DEFINICIONES ALINEADAS</b>	<b>124</b>
<b>APÉNDICE F:</b>	<b>BWANANET</b>	<b>130</b>
<b>APÉNDICE G:</b>	<b>CRITERIOS DE BÚSQUEDA EN EL BWANANET</b>	<b>133</b>
<b>APÉNDICE H:</b>	<b>EXTRACTO DE CONSTRUCCIONES VÁLIDAS</b>	<b>136</b>

<b>TABLA DE ABREVIATURAS</b>	<b>140</b>
------------------------------	------------

## ÍNDICE DE TABLAS

---

TABLA 1: ALINEAMIENTO DE DEFINICIONES PARA EL TÉRMINO <i>CASO ERGATIVO</i> .	28
TABLA 2: TOTAL DE PARES GENERADOS CON EL ALGCL SIN IMPORTAR CATEGORÍA GRAMATICAL.	32
TABLA 3: RESULTADOS DE LA PRIMERA DEPURACIÓN DEL CORPUS EXTRAÍDO DEL ALGCL.	33
TABLA 4: ESTADÍSTICA FINAL DE AGRUPAMIENTOS Y FORMA VERBALES.	33
TABLA 5: ANÁLISIS DE LOS VERBOS <i>SUPLIR</i> Y <i>SUSTITUIR</i> EN TRES MEDIOS.	35
TABLA 6: EL CA CON LOS 28 PARES DE VERBOS.	36
TABLA 7: ESTADÍSTICAS DE LA BASE LÉXICA QUE INTEGRA EL CA.	38
TABLA 8: ESTADÍSTICAS DEL CP.	42
TABLA 9: LA EVALUACIÓN DE CONTEXTOS EN NÚMEROS.	44
TABLA 10: EJEMPLO DE LA GENERACIÓN DE PARES SEMÁNTICOS	120



## ÍNDICE DE FIGURAS

---

FIGURA 1: EJEMPLO DEL PROCESO DE BÚSQUEDA ONOMASIOLOGICA	12
FIGURA 2: ARQUITECTURA DEL SISTEMA DE BÚSQUEDA ONOMASIOLOGICA.	26
FIGURA 3: PLANTILLA DEL <i>FRAME</i> PONER	73
FIGURA 4: ORGANIZACIÓN DE LOS <i>FRAMES</i> DEL CA	73
FIGURA 5: EJEMPLO DE UN DESCRIPTOR Y DE LOS NODOS DE INFORMACIÓN A LOS QUE PUEDE LLEGAR.	91
FIGURA 7: ESQUEMA DEL DESCRIPTOR <i>CAUSAR</i>	98
FIGURA 8: BÚSQUEDA EXPANDIDA CON FRAMES Y DESCRIPTORES	98
FIGURA 9: MATRIZ DE FORMAS LÉXICAS QUE INTEGRAN LAS DEFINICIONES QUE SE ALINEARÁN	118
FIGURA 10: MATRIZ GENERAL DE RESULTADOS PARA EL ALINEAMIENTO	118
FIGURA 11: RUTA DE FORMAS DEL ALGORITMO DE WAGNER & FISHER	119
FIGURA 12: INTERFAZ DE ELECCIÓN DE DOMINIOS Y SUBDOMINIOS	133
FIGURA 13: INTERFAZ DE TIPO DE CONSULTA	134
FIGURA 14: INTERFAZ DE ELECCIÓN DE CONSULTA	135

# INTRODUCCIÓN

---

## 0.1 Antecedentes

Sin lugar a dudas, una de las características de nuestra época es la enorme cantidad de información que se transmite por medio de la lengua, desde aquella que está presente en una simple conversación hasta aquella que se almacena en las bases de datos de sistemas informáticos y que constituye grandes corpus textuales y orales. Tal cantidad de información que por doquier nos inunda no es sino una extensión de la función que tiene la lengua como medio de comunicación e interacción y que, en sus varias formas de expresión, ofrece al estudioso una gama interminable de fenómenos que involucran los distintos niveles de análisis lingüístico.

Dicho cúmulo de información, sin embargo, lejos de pertenecer puramente a la investigación lingüística, representa para otras disciplinas una fuente valiosa de explotar. Las ciencias de la computación se han encargado, de un tiempo a la fecha, en crear recursos que puedan aprovechar esta información como materia prima para proveer a la gente de desarrollos y tecnologías que representen un beneficio en cualquier ámbito de su vida.

En este sentido, tanto la lingüística como las ciencias de la computación han convergido en las últimas décadas en la investigación, análisis y desarrollo de tecnologías que explotan el conocimiento inherente a la lengua para modelar, a través de técnicas computacionales, sistemas de interacción humano – computadora. Y es esta sinergia de disciplinas e ideas la que da sentido a este trabajo.

Los sistemas de recuperación de información, que no son más que programas que, dado un estímulo por parte de un usuario, por lo general a través de una interacción textual, responden proporcionando información que se encuentra en sus bases de datos<sup>1</sup> (Hjørland 1998), son uno de los muchos puntos de convergencia entre estas disciplinas.

---

<sup>1</sup> Baste como ejemplo de lo que es un sistema de recuperación de información los buscadores de internet: Google, Yahoo, Altavista, etc., los cuales, a partir de un estímulo generado textualmente, devuelven

Tales sistemas han supeditado por mucho tiempo su desempeño sólo a los métodos y técnicas computacionales y estadísticos. La lingüística, de a poco, ha ido incorporándose en el diseño y planeación de éstos a través de la investigación sobre los distintos fenómenos de la lengua que ven salida en el uso de ésta. En este sentido, uno de los principales intereses de la investigación lingüística se ha enfocado en la forma de estructurar el léxico en las bases de los sistemas de recuperación de información. A medida en que tal estructura pueda representar, de mejor forma, la lengua de hablantes y oyentes y no tanto de gramáticas, estos sistemas podrán explotar con mayor intensidad la infraestructura computacional y, sobre todo, de información, que redunde en beneficios para los usuarios de los sistemas.

### 0.1.1 El punto de partida

En el Grupo de Ingeniería Lingüística (GIL) de la UNAM hemos trabajado, a lo largo de cinco años, en la integración de un conjunto interdisciplinario de personas que se involucren en la investigación lingüística y computacional para el desarrollo de sistemas de cómputo capaces de *reconocer, entender, interpretar y generar lenguaje humano en todas sus formas*<sup>2</sup>. Los retos han sido varios, al igual que los avances y tropiezos. No obstante, gracias al trabajo en conjunto y a la convicción en lo que el área puede aportar a la investigación, tanto lingüística como computacional, así como al público en general, en el GIL no paramos, desde nuestra perspectiva y nuestros enfoques, de abordar los problemas que atañen a estas áreas<sup>3</sup>. Uno de estos problemas lo constituyen los sistemas de recuperación de información, en concreto, un sistema que, desde un enfoque onomasiológico y a través de lengua natural, pueda proporcionar al usuario la información que éste precisa.

---

información que pretende responder a tal estímulo. La respuesta puede variar, desde una palabra o término específico, hasta ligas a documentos en donde se encuentra el estímulo. Cabe destacar que los resultados dependen, en cierta parte, de la manera en la que se formula la búsqueda, pero también de la manera en la que esté estructurada la información en sus bases. Y éste es el punto, *grosso modo*, que abordamos en este trabajo: la forma de estructurar el léxico en estas bases.

<sup>2</sup> Para un mejor entendimiento de lo que es la ingeniería lingüística consultar las siguientes ligas electrónicas: <http://sunsite.dcc.uchile.cl/~abassi/WWW/Lengua/ingenieria.html>  
<http://iling.torreingenieria.unam.mx/>

<sup>3</sup> Para mayor información sobre el trabajo que se ha desarrollado en el GIL así como la infraestructura académica y estudiantil, consultar: <http://iling.torreingenieria.unam.mx/>

Tal sistema de búsquedas onomasiológicas (Sierra 1999) entraña en sí una gama de problemas que se deben resolver para lograr el propósito que con él se busca: recuperar un término a partir de una descripción, vía lengua natural, de lo que un usuario conceptualiza como características intrínsecas del término. Para ello debemos contar con una estructura que primero, a nivel léxico, y luego, a nivel computacional, pueda garantizar que se recupere el término que el usuario busca. En este sentido, abordamos aquí la problemática que representa la estructuración del léxico en la base de conocimiento de este sistema.

### **0.1.2 Delimitación**

El problema que acometemos en este trabajo consiste en proporcionar patrones, sustentados en análisis lingüísticos de un determinado léxico, para que la estructura del léxico en la base de conocimiento del sistema pueda asegurar que se recupere la información que el usuario necesita. A manera de ejemplo de lo que debe hacer el sistema, véase la Figura 1.

Lo que podemos observar en la figura es el proceso simplificado que debe realizar el motor de inferencias de este sistema para recuperar un término. Dado un grupo de palabras que definen el término X, el sistema debe asociarlas con un conjunto mucho mayor de palabras que representan conceptos relacionados para así expandir la búsqueda. El resultado es la recuperación del término X.

La recuperación de este término supone para el sistema que debe existir una base léxica compuesta, además de un léxico controlado que defina los términos<sup>4</sup> (Kang, et al. 2004) y de un conjunto de paradigmas que sirva para expandirlo (Sierra 1999), de una estructura de información tal que permita representar la estructura conceptual de un área determinada así como la integración de léxico que caracterice dicha estructura.

---

<sup>4</sup> Está visto, como señalan (Kang, et al. 2004), que si bien un léxico controlado representa un punto de partida pertinente para la construcción de la base de conocimiento, no es posible conformarla sólo con esta información, puesto que este léxico sólo representaría un subconjunto de un conjunto mucho mayor.

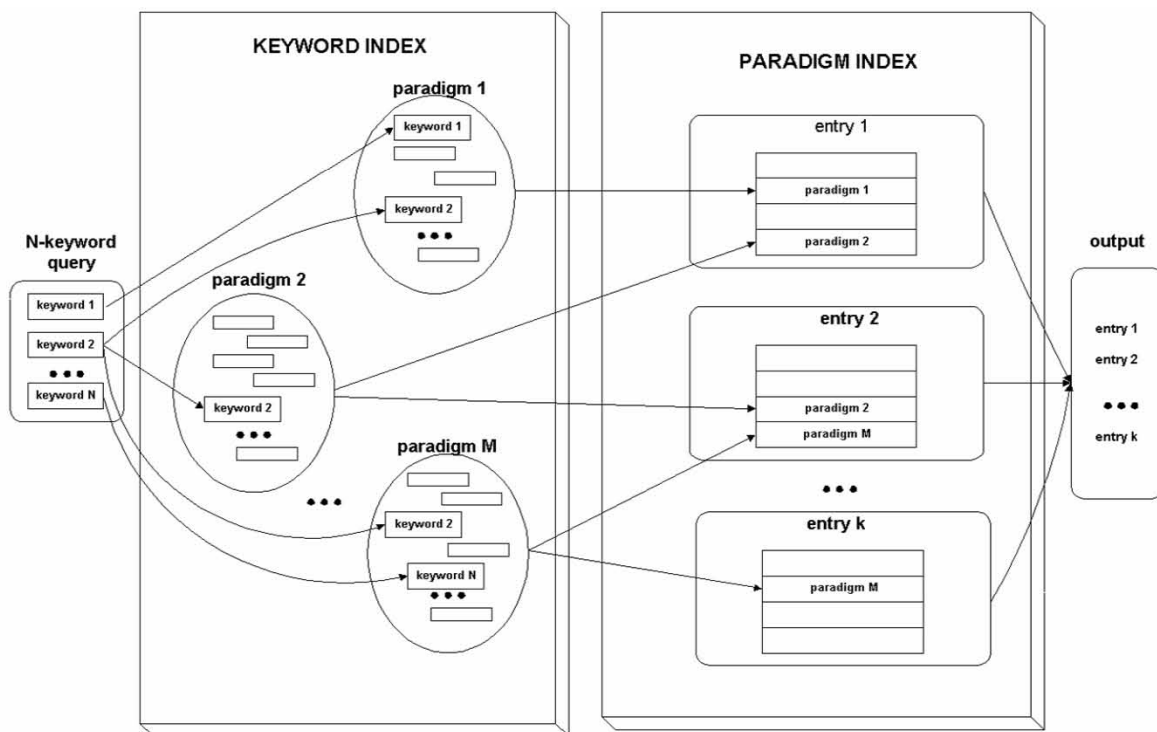


Figura 1: Ejemplo del proceso de búsqueda onomasiológica<sup>5</sup>

En este sentido, nuestro trabajo se suscribe a este sistema de recuperación de información y, en específico, al problema que representa la constitución de su base de conocimiento. Particularmente, nos enfocaremos a analizar el comportamiento de un grupo específico de verbos que, sin ser sinónimos, se comportan como tales. A partir del estudio de estos elementos, cuyo especial comportamiento se da al interior de un corpus terminológico integrado por definiciones, pretendemos sugerir patrones que permitan crear una estructura léxica sustentada en la forma en que los hablantes usan la lengua y no sólo en la gramática.

## 0.2 Objetivo general

Diseñar un esquema para estructurar la información, desde un plano semántico y pragmático, en la base de conocimiento léxico de un sistema de recuperación de información específico.

<sup>5</sup> Figura tomada de Sierra (1999).

Con este esquema que parte del análisis de un grupo de verbos que, sin ser sinónimos se desempeñan como tales en contextos específicos, buscamos sentar las bases para construir un sistema de recuperación de información que, además de estar sustentado en una estructura que pretende involucrar la cognición humana a través del léxico, pueda tener un desempeño más apegado a la comunicación de los humanos.

### 0.2.1 Objetivos particulares

- a) Argumentar semánticamente, bajo los criterios de la gramática cognoscitiva, el comportamiento particular de un grupo de verbos que sin tener relación léxica, dentro de determinados contextos, se comportan como si la tuvieran.
- b) Esto implica analizar qué fenómenos, más allá de la estructura sintáctica, validan que al igual que *X*, funcionen *Y* y *Z* como sinónimos, dado que:
  - *X*, *Y*, *Z* tienen un significado léxico propio y distinto de cada uno;
  - *X* se usa como sinónimo de *Y* en un contexto específico;
  - tanto *Y* como *Z*, pueden aparecer en lugar de *X* en estos mismos contextos sin modificar el significado total de la construcción.
- c) Justificar la inclusión de este grupo de verbos en la base de conocimiento léxico, respaldando con ello los resultados del sistema computacional que los produjo: el algoritmo de clustering.
- d) Proponer un esquema de *frames* que pueda representar la gama de significados que los verbos analizados codifican.
- e) Subrayar la importancia de los procesos de codificación, interpretación, inferencia, reanálisis, decodificación, etc., y su repercusión en la comunicación de los humanos.
- f) Suscribir una propuesta para la integración de información pragmática en la base de conocimiento léxico del sistema de recuperación de información.

### 0.3 Hipótesis

El estudio y análisis de la lengua como filtro de la cognición podrá permitirnos, a través del análisis semántico y pragmático, incorporar información, propia de la

comunicación humana<sup>6</sup>, en la base de conocimiento léxico de un sistema de recuperación específico.

Si esto es verdad, la base de conocimiento de este sistema podrá representar, de mejor manera y de forma más apegada a lo que es la comunicación humana, información que proviene de una esfera que rebasa la gramatical y que, de cierta manera, involucra los procesos de cognición.

## **0.4 Metodología**

El objetivo de la investigación supone que ésta tienda hacia el empirismo. El análisis y la argumentación servirán como bases para proponer una organización del léxico en donde sea posible tomar mucho más en cuenta a los hablantes y sus entornos vivos de comunicación. Esto implica que la base de conocimiento del sistema pueda integrar información que vaya más allá del nivel gramatical, logrando con ello un incremento en los niveles de búsqueda y recuperación de información que estén en función de la forma en que los humanos conceptualizamos.

### **0.4.1 Justificación del léxico de la investigación**

Sentado el objetivo que pretendemos con esta investigación es menester justificar el por qué de trabajar con un léxico integrado sólo por verbos.

En primer lugar, y de acuerdo con Fillmore (1968), Eumeridou, Nkwenti-Azeh, y McNaught (2004), consideramos que el verbo es el elemento central sobre el que gira el entorno conceptual de una construcción.

Creemos, en segundo lugar, al igual que Giulia, Bencini y Goldberg (2000) que, a nivel de una realidad psicolingüística, el verbo representa un marco que es capaz de fungir

---

<sup>6</sup> Entendemos ésta como la serie de procesos dinámicos que parten de la gramática y se reconstruyen en una esfera superior a ella en donde se recupera, además de información léxica, un fondo de conocimiento tal que involucra la cognición humana y su representación a través de la lengua. En el Capítulo 4 profundizamos en el concepto de comunicación.

como constrictor del significado de todo el discurso de una construcción y, por ende, funcionar como el punto al que se pueden anclar significados particulares y complejos.

Por último, dados los puntos precedentes, estamos seguros de que, como indica McCarthy (2001), el verbo es el elemento capaz de funcionar como el pivote de una estructura léxica que puede ser manipulable, de manera más profusa y profunda, por un sistema computacional

#### **0.4.2 Marco teórico**

Las bases teóricas sobre las que gira la investigación provienen de la gramática cognoscitiva. Optamos por este marco puesto que a partir del enfoque con el que aborda la investigación lingüística se pretenden explicar los fenómenos lingüísticos no sólo a través de la propia lengua sino a partir de una base cognoscitiva que funge como catalizadora entre la lengua y el mundo físico, psicológico y neurológico de hablante y oyente.

De aquí que nos inclinemos por el estudio lingüístico por medio de un marco cognoscitivo que nos permita expandir el nivel de análisis y que, de a poco, nos dé pie para integrar, en las bases de conocimiento léxico de sistemas computacionales, procesos más apegados a la realidad, tanto cognoscitiva como comunicativa, de hablante y oyente.

Bajo esta perspectiva, el análisis semántico lo realizaremos a partir de los fundamentos teóricos del modelo *Construction Grammar* (Goldberg 1997) y del marco semántico *Frame Semantics* (Fillmore 1982). Con base en los principios conceptuales de estos modelos buscamos involucrar, de manera más consistente, la realidad lingüística del hablante para poder representarla en la estructura lingüística de la base de conocimiento.

Por otro lado, el aspecto pragmático lo pretendemos analizar desde la plataforma conceptual que nos proporciona el modelo de *Teoría de la Relevancia* (Sperber y Wilson 1995). Cabe destacar que de este modelo recuperamos sólo el concepto general que le subyace y que involucra la realidad lingüística del oyente, puesto que en la investigación no se hace un análisis pragmático como tal, lo que planteamos es sólo la idea de que, por



medio de un estudio profundo, podríamos incorporar información pragmática en la estructura de la base de conocimiento.

## 0.5 Organización

La arquitectura de la investigación está organizada en cuatro capítulos. En el primero describimos el marco general del cual parte la investigación. Con el segundo delimitamos el corpus de trabajo y detallamos la integración del grupo de verbos sobre el que gira el análisis. Dentro del tercer capítulo hablamos del análisis y argumentación semántica de los verbos en cuestión y del esquema que funcionaría como estructura general del léxico. En el cuarto cuestionamos la necesidad de la pragmática para una descripción más abarcadora de la lengua y esbozamos una propuesta para integrar información de esta índole dentro de una estructura más compleja. Al final, daremos las conclusiones a que hemos llegado.

A continuación detallamos lo que presentaremos en cada capítulo.

En el Capítulo 1 hablaremos de lo que son los sistemas de recuperación de información. Centraremos la atención en el problema que representa el tratamiento masivo de la información de carácter textual. Asimismo, delimitaremos el punto que nos interesa de estos sistemas, es decir, las bases de conocimiento. Señalaremos la importancia de la investigación lingüística para la constitución de éstas y ejemplificaremos con la base de conocimiento del proyecto de búsquedas onomasiológicas. Por último, hablaremos de una herramienta que nos ha funcionado para generar parte del léxico que constituiría la base, sobre todo focalizaremos la atención en los resultados que esta herramienta arroja y que, dado el comportamiento del léxico en tales resultados, son de interés para la lingüística.

En el Capítulo 2 hablaremos de cómo conformamos el corpus de verbos con los que trabajamos<sup>7</sup>. Describiremos el comportamiento de los verbos señalando por qué

---

<sup>7</sup> Cabe destacar que estos verbos no están relacionados con una clase particular, es decir, el grupo no está compuesto sólo por verbos estativos, de actividad o de logro, por ejemplo, más bien, el grupo está integrado

creemos que tal comportamiento merece especial atención. Asimismo, señalaremos la problemática que nos plantean estos elementos. De igual forma, especificaremos el porqué construimos un corpus de prueba o *lexical chain*<sup>8</sup>, como lo denominan Kang et al. (2004), y detallaremos el ejercicio que realizamos con él y la posterior evaluación de los resultados de éste. Por último, sentaremos nuestra postura respecto a este comportamiento señalando por qué creemos que se da y en función de qué.

En el Capítulo 3 analizaremos, a partir de las bases de la gramática cognoscitiva, el corpus de verbos que nos ocupan. Con base en este análisis damos una argumentación al porqué de la aparición de estos verbos en distintos contextos. Como cierre del análisis hacemos patente que la argumentación por medio de *frames* nos permitiría crear una estructura de léxico en donde se puedan recuperar, a través de características semánticas generales, significados más complejos y abstractos a partir de uno más particular. La última parte del capítulo incorpora un cuestionamiento sobre la pertinencia e importancia que un análisis pragmático aportaría a los objetivos de la investigación.

En el Capítulo 4 desarrollaremos este cuestionamiento abordando el problema que entraña la lengua más allá de las fronteras de una gramática. Definimos lo que entendemos por análisis pragmático y el reto que representa un análisis de este corte dentro de la investigación. Ejemplificamos la problemática que implica la comunicación humana y el trabajo que conllevaría tratar que una máquina la simulara. Para finalizar, hacemos una propuesta de un esquema en el que tratamos de incorporar la información que obtendríamos de un análisis pragmático en la base de conocimiento de nuestro sistema de recuperación de información.

Finalmente, para cerrar la investigación, dedicamos un módulo en el que vertteremos nuestras conclusiones sobre lo expuesto a lo largo del trabajo.

---

por verbos extraídos de manera aleatoria por un programa computacional cuyo funcionamiento se detalla en el Apéndice A.

<sup>8</sup> En este sentido, adoptamos la postura que plantean Kang et al. (2004) de corpus, es decir, para comprobar que los conceptos de un dominio están determinados por un concepto mucho más general que da cabida a muchos más elementos.

# SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN

---

En este capítulo planteamos la problemática que encierra el manejo y explotación de grandes cantidades de información textual con medios electrónicos. Señalamos que existen sistemas especialmente diseñados para la manipulación de tales volúmenes de información y, sobre todo, puntualizamos que estos sistemas deben incorporar módulos lingüísticos que reflejen de mejor manera la forma en que la lengua, en su aspecto escrito, es utilizada por los hablantes.

Desarrollamos la problemática que lo anterior supone, suscribiendo la investigación a un proyecto en el que se pretende construir un sistema de recuperación de información que precisa de una base léxica muy específica.

Por último, describimos la herramienta computacional que nos ha permitido la extracción de cierta parte del léxico que integrará la base del mencionado sistema y los resultados que extraemos de ella para analizarlos en la investigación.

## 1.1 El problema de la inmensidad de la información textual

La lengua, en sus distintas formas de expresión, constituye el vehículo ideal para generar y propagar el carácter vivo y particular de la ideología, pensamientos, creencias y rasgos de una sociedad. El resultado que de esta interacción se desprende es una cantidad inmensa de información, ya sea textual u oral, que por doquier nos inunda.

Hoy en día, es esta inmensidad el rasgo más prominente que caracteriza la información, y de forma más precisa, la información textual.<sup>1</sup>

---

<sup>1</sup> De aquí en adelante sólo nos referiremos a la información de índole textual, por tanto, cada vez que aparezca la palabra *información*, se sobreentenderá que está en función de la producida o almacenada en forma textual.

La masificación de información que se genera a lo largo del mundo y la aparición de nuevas tecnologías para el almacenamiento y explotación de ésta hacen que día con día los volúmenes de información aumenten de forma incalculable. La consecuencia lógica de tal crecimiento es la imposibilidad de manejarla y, por tanto, perdernos en un sinfín de caracteres escritos irrelevantes.

En el aspecto práctico, esto se vuelve un problema aún mayor al darnos cuenta de lo difícil que resulta encontrar resultados satisfactorios en medio de esta infinidad de información. Baste como ejemplo Internet: búsquedas que sólo arrojan cantidades inmensas de información muy pocas veces relevante. En este sentido, ya no basta sólo con tener grandes bases de datos que almacenen y provean de resultados a los usuarios, sino desarrollar sistemas que permitan una mejor interacción hombre – máquina, de la forma más natural posible, ofreciendo resultados óptimos en cuanto a calidad, cantidad y tiempo de procesamiento.

Si no lográramos realizar esta clase de sistemas, ¿de qué nos serviría tener tanta información convertida sólo en un montón de caracteres incapaces de producir conocimiento?

### **1.1.1 Qué es un SRI**

Una de las tecnologías que desde mediados del siglo pasado ha logrado un desarrollo importante con relación a la explotación de información textual es aquella denominada *recuperación de información*. Tal tecnología proviene de las llamadas ciencias de la información. El objetivo primordial de estas ciencias, según Hjørland (1998), es la de buscar y recuperar textos o alguna otra información dentro de bases de datos o Internet.

La investigación en esta clase de tecnologías trajo consigo el crecimiento de estas ciencias, lo cual se tradujo en el desarrollo de distintas técnicas para recuperar información de bases de datos. La paulatina aplicación de tales técnicas en el medio computacional dio

como resultado la aparición de sistemas informáticos cuya función era la de recuperar documentos a petición de una o varias palabras, frases u oraciones de un usuario.

A esta clase de sistemas se les denominó *sistemas de recuperación de información* (SRI). Lo que éstos hacen, sin entrar en complicaciones computacionales, es lo que Hjørland (1998) denomina como “the process in which users put questions to information systems and consequently get some answers”. Esto es, a un estímulo textual por parte de un usuario, el sistema responde devolviendo documentos o ligas a documentos que contienen el estímulo del usuario.

Este proceso que parece tan sencillo de ejecutar para una computadora precisa de técnicas y métodos<sup>2</sup> que, a lo largo de los años, se han ido refinando, y en donde el dominio matemático parecía exclusivo para el funcionamiento de los SRI. Sin embargo, en las últimas décadas, una disciplina que ha ido incorporando los resultados de su investigación a las ciencias de la información es la lingüística.

Con el conocimiento que ésta ha generado sobre la lengua se busca desarrollar mejores sistemas que logren explotar de mejor manera la información textual, la cual, en última instancia, es una proyección de todo sistema lingüístico<sup>3</sup>.

### **1.1.2 Importancia del análisis lingüístico en los SRI**

Como en todo sistema computacional, el motor principal que genera los resultados opera básicamente con técnicas matemáticas, no obstante, con el paso de los años los SRI han evolucionado con la incorporación de nuevas técnicas que mejoran su desempeño. La interdisciplinariedad ha jugado un papel muy importante en esta evolución; ahora, los SRI

---

<sup>2</sup> Existen varias técnicas aplicadas a los sistemas computacionales que buscan ofrecer mejores resultados a los usuarios; algunas de ellas se basan en métodos estadísticos, algunas otras en funciones algorítmicas, o combinación de métodos. Para este trabajo no es indispensable entrar en detalle sobre la forma en que operan tales métodos, sin embargo, quien esté interesado puede adentrarse en estos tópicos consultando Hjørland 1998, Kang, Dae-Won y Sang-Jo 2004, Sierra 1999, Arotariteia 2004, Lance y Williams 1967, Fukumoto, y Yoshimi 1996, Krovetz 1991.

<sup>3</sup> Al respecto Montgomery 1972, en su artículo “Linguistics and Information Science”, hace un estudio bastante amplio de cómo han sido explotadas estas disciplinas para generar sistemas computacionales que pretenden imitar la comunicación humana en su forma oral y textual.

cuentan con módulos interactivos en los que la estadística o los algoritmos no son los únicos motores que generan resultados. Conocimiento morfológico, sintáctico, semántico e incluso, pragmático y discursivo, han revolucionado la forma en la que los SRI operan.

Pero, ¿en dónde radica el beneficio de que los SRI interactúen con técnicas lingüísticas, puesto que, con o sin éstas, los SRI operan devolviendo resultados sean estos buenos o malos?

El principal beneficio lo obtendrían los usuarios, en específico, al momento de interactuar con el sistema<sup>4</sup>. Sierra (1999) apunta que “a good design for a system which responds to user needs should take into account the profile of the people who will interact with it”. En consecuencia, creemos que a mayor investigación sobre el cómo es y cómo se comporta determinada lengua en su aspecto vivo, esto es, a través de los hablantes y no de las gramáticas prescriptivas, mejor será el diseño de la estructura, tanto computacional como lingüística de los SRI.

Con este conocimiento sobre el perfil del usuario del que Sierra habla —que no puede obtenerse sino de investigación lingüística que especifique cómo es que los usuarios interactuarán, en su propia lengua, con el sistema—; se tendrán pautas para crear modelos de lengua que reflejen con mayor fineza las características lingüísticas y comunicativas de una lengua.

### **1.1.3 El módulo lingüístico en los SRI**

En el apartado anterior señalamos que resulta necesario incluir el conocimiento referente a la lengua en los SRI para robustecer la arquitectura de estos y mejorar la interacción humano - computadora.

---

<sup>4</sup> En este trabajo nos enfocamos en un sistema específico del que se hablará en el apartado 1.3. Este SRI interactúa con el usuario a partir de lengua natural escrita, de ahí que damos por sentado que cada que hablemos de SRI generalizamos la interacción humano – computadora como aquella dada a través de lengua natural en forma textual.

Para poder permear este conocimiento y, en consecuencia, reflejarlo, los SRI cuentan con uno o varios módulos que contienen toda esta información. A estos módulos se les llama *bases de conocimiento léxico* (BCL). En estas BCL se almacena toda la información, estructurada o no, relativa a la lengua. La investigación de los distintos niveles de análisis lingüístico encuentra un formalismo o una estructura en estos módulos.

## 1.2 Bases de conocimiento léxico

Una BCL parte de la idea de obtener, organizar y dinamizar el conocimiento relativo a la lengua. Tal conocimiento se genera a partir del análisis lingüístico de cualquier forma de expresión, sea ésta una conversación o todo el material de una biblioteca. Dicha información lingüística debe proporcionar los elementos mínimos necesarios para estructurar el léxico en la base de conocimiento, convirtiendo al léxico en algo más que una serie de entradas estáticas, sino, como señala Domenig (1988), en información con “*capacidades dinámicas*”.

Con una BCL, los SRI tienen no sólo un grupo de formas aisladas que llenan una base de datos sino:

- i. un banco con la información precisa y suficiente que requiere el sistema;
- ii. una estructura léxica jerarquizada, que se sustenta en análisis lingüísticos, de la información que constituye la BCL;
- iii. una serie de reglas que junto con una base de hechos simulan pensamientos lógicos que determinan la validez de un problema;
- iv. un SRI más potente, con módulos que interactúan ente sí, capaz de inferir la clase de información que requiere el usuario<sup>5</sup>.

En consecuencia, en la medida en que los SRI cuenten con una mejor BCL que les permita interactuar de forma más dinámica y natural, tanto con las entradas léxicas que la constituyen como con el usuario, los resultados que se le presenten a este último serán más satisfactorios y de mayor relevancia.

---

<sup>5</sup> Esto se logra a partir de la descripción lingüística del léxico que conforma la BCL, el cual está en función de un objetivo específico del SRI.

### 1.2.1 Ejemplo de una BCL

Para esquematizar como trabaja una BCL, tómesese en cuenta que ésta interactúa de manera constante con dos elementos:

- i. un motor de inferencias que se encarga de interactuar con la información que está almacenada para la resolución de un problema<sup>6</sup>;
- ii. una base de hechos<sup>7</sup>, en constante aprendizaje, capaz de deducir sucesos válidos a partir de nuevos datos aún sin clasificar.

El proceso que realizan estos dos elementos que interactúan con la BCL se detalla en el siguiente ejemplo<sup>8</sup>:

- i. se designa un conjunto de reglas y una serie de hechos tales como X escribe Y;
- ii. se extrae información relativa al problema que se pretende solucionar, es decir, información específica, en este caso, sobre entes humanos (hecho);
- iii. se asocia a estos hechos un atributo específico (ser escritor), de forma que tenemos:
  - o Cervantes (escritor),
  - o Shakespeare (escritor);
- iv. se extrae información sobre obras asociadas a un concepto, como lo es libro:
  - o Don Quijote de la Mancha (libro),
  - o Hamlet (libro);
- v. se determina una regla que asocie hechos y atributos:
  - o escribir (autor [x], libro [y]);
- vi. se generan relaciones del tipo:
  - o escribir (Cervantes [X], Don Quijote de la Mancha [Y]);
- vii. con esto, el usuario podrá hacer consultas del tipo:
  - o ¿Quién escribió Don Quijote de la Mancha?;
- viii. el SRI podrá devolver al usuario la información que precisa:
  - o Cervantes.

---

<sup>6</sup> Para lograr esto, se emplea un conjunto de reglas previamente establecidas que le permiten determinar si el hecho que va a analizar, según los datos que maneja, es válido o no.

<sup>7</sup> Esta base de hechos está en relación con las reglas que determinan el funcionamiento del motor de inferencias.

<sup>8</sup> Ejemplo expuesto en el *Segundo Seminario de Ingeniería Lingüística* en la conferencia “Bases de Conocimiento Léxico” impartida por Reyes, Aguilar y Morett. Septiembre 10, 2004.



### **1.2.2 La BCL del proyecto**

Uno de los objetivos de esta investigación es el de reforzar la BCL que se integrará en el SRI que desarrollamos y del que más adelante se hablará. Para constituir esta BCL<sup>9</sup>, en el GIL se han desarrollado algunas herramientas computacionales que cumplen con la tarea de obtener el léxico necesario para constituir la mencionada base, de la misma forma en que se ha trabajado en la obtención de un corpus de términos y de definiciones, así como de un corpus especializado para garantizar los recursos léxicos que la constituyen.

La BCL del SRI que desarrollamos, así como el proyecto en el que se engloba, se describe en el siguiente apartado.

### **1.3 El proyecto DEBO**

El SRI que está en desarrollo en el GIL, de carácter lexicográfico, tiene por objetivo que el usuario realice una búsqueda, describiendo con sus propias palabras, un término. El sistema le deberá devolver el término que se asocia con la definición escrita por el usuario.

La arquitectura que plantea este SRI, que parte del concepto y va hacia el término, se denomina onomasiológica. De ahí sus siglas: DEBO, diccionario electrónico de búsquedas onomasiológicas.

En contraposición con el común de sistemas que sólo permiten búsquedas basadas en arquitecturas semasiológicas, esto es, que parten del término y van al concepto, el DEBO permitirá al usuario, según lo plantea en su tesis de doctorado su diseñador, “to use natural language to describe an idea or concept in order to search for terms which he does not know how to designate correctly” (Sierra, 1999).

---

<sup>9</sup> Es preciso señalar que esta base de que hablamos está en función de un dominio específico, en nuestro caso, las terminologías de distintas áreas científicas, tales como lingüística, física, desastres, sexualidad, ingeniería, etc., son nuestros dominios.

El DEBO busca cumplir con lo que se espera de todo SRI y que Schank, Kolodner and DeJong (1981) destacan a la perfección:

An information retrieval system should understand questions that people pose to it in the language most convenient for them to use - the one they speak. People should be able to formulate questions in natural language, and the system should be able to analyse the questions, extract their meanings, automatically create search keys from the natural-language input, and retrieve and generate appropriate answers.

### 1.3.1 La BCL del DEBO

Como arriba mencionamos, el DEBO tendrá una interfaz de consulta en lengua natural, esto supone que el léxico, y la estructura que éste tendrá en la base, deberán estar en función de una simulación, lo más cercano posible, de los procesos de comunicación oral humana.

Lo anterior conlleva a cuestionarse si es factible el hecho de que una máquina pueda interactuar con un humano de manera *inteligente*. La respuesta que nosotros proponemos y que nos tomamos muy en serio, es que, dados los elementos necesarios a nivel lingüístico y computacional, una máquina será capaz de simular, con limitaciones claro está, un proceso de comunicación. En nuestro caso, el de interpretar y codificar el estímulo textual del usuario para poder devolverle el término que busca.

Para lograr esto, la BCL del DEBO deberá tener una estructura léxica flexible capaz de dar el soporte necesario a los diferentes tipos de descripciones que los usuarios hagan de un término. Es por ello que en esta investigación argumentamos lingüísticamente los elementos que constituyen parte de la BCL y proponemos, a partir de análisis semánticos y pragmáticos, una posible estructura del léxico que permitirá devolverle al usuario el término que requiere, independientemente del léxico y del tipo de definición empleados en su búsqueda.

### 1.3.2 Estructura léxica de la BCL del DEBO

La estructura léxica que compone la BCL del DEBO está organizada en cinco módulos que componen la arquitectura del sistema.

- i. términos;
- ii. definiciones;
- iii. fuentes de consulta;
- iv. palabras clave;
- v. paradigmas.

Bajo esta organización, representada en la Figura 2, la BCL cuenta con la información suficiente para garantizar y soportar las búsquedas que los usuarios realicen<sup>10</sup>.

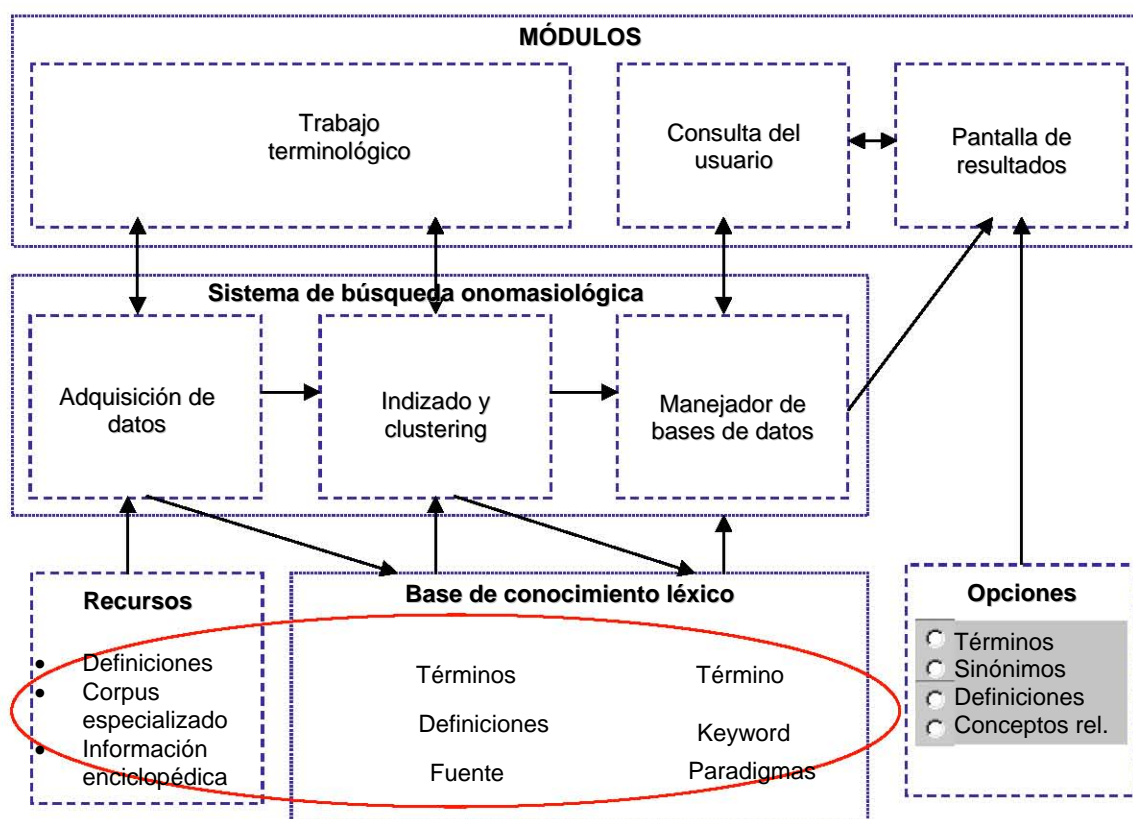


Figura 2: Arquitectura del sistema de búsqueda onomasiológica<sup>11</sup>.

<sup>10</sup> Para obtener detalles más puntuales y una explicación más amplia de cómo está constituida la BCL del DEBO, consultar: Sierra, Reyes y Aguilar (2005).

<sup>11</sup> Figura tomada de Sierra (1999).

## 1.4 Algoritmo de Clustering

Una de las herramientas que han generado parte del léxico que constituye la BCL del DEBO es el denominado *algoritmo de clustering* (AlgCl). Dicho algoritmo fue desarrollado por Sierra (1999) para su tesis doctoral, y modificado posteriormente en su versión para el inglés, por Sierra y McNaught (2000, 2002), y para el español, junto con Castillo (2002, 2003).

El papel que juega el AlgCl en el proyecto es el de extraer elementos que puedan ser intercambiables en contextos específicos. El beneficio que esto representa es el de “to collect together into clusters a set of elements associated by some common characteristic” (Sierra y McNaught 2002), en este caso, la característica de *sinonimia*<sup>12</sup>.

### 1.4.1 Una herramienta para generar pares semánticos de forma automática

Los elementos extraídos con el AlgCl han sido denominados *pares semánticos* (Sierra y McNaught 2000) dada la supuesta característica de sinonimia que tienen, la cual nos permite analizar el significado o uso que estos elementos adoptan tal y como si estuvieran relacionados semánticamente, aunque desde un punto de vista formal, no posean dicha relación (Sierra, McNaught y Castillo 2003).

La manera en que el AlgCl extrae los elementos que propone como pares semánticos es la siguiente:

- i. se toma un conjunto de términos pertenecientes a una misma área con sus definiciones;
- ii. se alinean las definiciones, provenientes de distintas fuentes, de un mismo término (ver Tabla 1);

---

<sup>12</sup> Destacamos el concepto sinonimia dado que en un análisis riguroso, los grupos de formas que el AlgCl propone y con las cuales trabajamos, no son sinónimos léxicos sino formas que, dado un marco semántico y/o pragmático prototipo (y esto es lo que desarrollamos en la presente tesis), adquieren características que les permiten funcionar como si en verdad fueran sinónimos.

- iii. se establecen, mediante reglas como la *distancia de Levenshtein* y *algoritmo de Wagner & Fisher*<sup>13</sup>, parejas de palabras que pueden sustituirse unas por otras y cuya sustitución no cambia el significado de las definiciones;
- iv. los grupos de palabras, que según las reglas que constituyen el AlgCl pueden ser intercambiables, son los pares semánticos o *bindings*<sup>14</sup>.

En algunas ocasiones, los pares semánticos que el AlgCl propone están relacionados a manera de sinónimos; sin embargo, hay veces en las que los elementos que conforman el par no comparten dicha relación. Esto lo podemos ver en la Tabla 1. El que aparezca *marcar* o *distinguir*, que no son sinónimos, no cambia el significado total de la definición.

Def. 1	CASO	ERGATIVO	caso	que	marca	al	agente		de	los	verbos	transitivos
Def. 2	CASO	ERGATIVO	caso	que	distingue	al	sujeto	activo	de	un	verbo	transitivo
Costos	0	0	0	0	1	1	2	3	3	4	4	4
Posible binding	NO	NO	NO	NO	SI	NO	SI	NO	NO	SI	NO	NO
Tipo	equal	equal	equal	equal	matched	equal	matched	null	equal	matched	equal	equal

**Tabla 1: Alineamiento de definiciones para el término *caso ergativo*.**

Esta clase de elementos que extraemos con el AlgCl —y en específico, los verbos— y que muestran la peculiaridad de que aún sin compartir la relación de sinonimia puedan comportarse como sinónimos en contextos específicos, son la base léxica que analizamos en esta tesis para argumentar desde el plano semántico, y en lo posible, pragmático, que:

- i. hay razones para señalar que hay esquemas comunicativos básicos que permiten codificar al oyente la información que el hablante transmite, independientemente de si ésta se suscribe sólo a su aspecto léxico o lo rebasa;
- ii. es posible estructurar en las BCL de cualquier SRI esta clase de información que a los hablantes, en el contexto apropiado, no le representa mayor dificultad para

<sup>13</sup> Ver Sierra (1999), Sierra y McNaught (2000, 2002), Sierra, McNaught y Castillo (2002), Castillo (2002) y Sierra y Castillo (2003). Para un resumen de cómo es el funcionamiento general del AlgCl ver el Apéndice A.

<sup>14</sup> “We introduce the term binding to represent a candidate cluster, i.e. two words that may be used in the same context without changing the meaning of a definition” (Sierra y McNaught 2000).

codificar y que en los medios especializados, como diccionarios y ontologías, no se da a priori o no se registra.

Por tanto, en el siguiente capítulo detallaremos cómo conformamos el corpus de pares que no comparten relación léxica alguna, así como los procesos que seguimos para comparar y evaluar si este comportamiento es privativo de un dominio terminológico o se puede extender a otros dominios y horizontes comunicativos.

## **EL CORPUS DE ANÁLISIS**

---

En este capítulo describimos el proceso de conformación y depuración del corpus de análisis; cómo llegamos a obtener un total de 28 pares de verbos que, según tres medios especializados, no tienen relación léxica de sinonimia pero que funcionan como si la tuvieran.

Detallamos el diseño y construcción de un corpus de comparación a partir de la búsqueda de contextos en el corpus del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra, denominado BwanaNet.

Por último, relatamos el proceso de intercambio de verbos para evaluar la pertinencia de éstos en diversos dominios comunicativos y lanzamos la hipótesis que señala que hay estructuras subyacentes que permiten codificar los eventos independientemente del verbo que aparezca en la definición.

### **2.1 Generación automática de pares semánticos para el corpus de análisis**

Como se señala en el capítulo anterior, los grupos de formas que interesan en la investigación son aquellos verbos que no comparten una relación léxica de sinonimia y que, sin embargo, pueden funcionar como sinónimos en el contexto del cual son extraídos. Para obtener este léxico corrimos el AlgCl sobre cuatro bases terminológicas que integran un banco terminológico mucho más amplio<sup>1</sup>. De los resultados de este proceso conformamos lo que denominamos el corpus de análisis (CA)<sup>2</sup>.

---

<sup>1</sup> La extracción del léxico fue hecha con la versión para el español del AlgCl (Castillo 2002), la cual corre sobre todas las bases que integran el banco terminológico del GIL, cuya consulta se puede realizar en: <http://iling.torreingenieria.unam.mx/diccionarios/>

<sup>2</sup> Cabe hacer énfasis en el hecho de que las bases que integran el banco terminológico están integradas sólo por definiciones, de forma que el CA es un corpus cuyos elementos fueron extraídos exclusivamente de definiciones, las cuales pueden consultarse en el Apéndice E.

El proceso realizado con el AlgCl para obtener los pares semánticos que conforman el CA se esquematiza a continuación:

- i. ingreso a <http://tabasco.torreingenieria.unam.mx/scripts/clusters.exe/home> desde donde se ejecuta el AlgCl ,
- ii. selección del algoritmo de alineamiento semántico,
- iii. elección de la base terminológica sobre la que correrá el AlgCl; en este caso:
  - o lingüística;
  - o física;
  - o sexualidad;
  - o desastres;
- iv. establecimiento de los parámetros de operación del programa<sup>3</sup>,
- v. selección de resultados que se pide que el AlgCl muestre,
- vi. ejecución del AlgCl sobre cada una de las bases.

Una vez que realizamos los pasos anteriores, el AlgCl ejecuta varios procesos<sup>4</sup> para agrupar elementos que, según los parámetros solicitados y las características propias del mismo algoritmo, pueden ser intercambiables<sup>5</sup>. Véanse las definiciones de a) y b) como ejemplo:

- a) *adjetivo numeral partitivo*: adjetivo numeral que **expresa** división (Seco, Manual de gramática española, 1985)<sup>6</sup>.
- b) *adjetivo numeral partitivo*: adjetivo numeral que **significa** división (Bello, Gramática de la lengua castellana destinada al uso de los americanos, 1981).

Los verbos *expresar* y *significar* son formas que según el AlgCl pueden considerarse como sinónimos, aún cuando semánticamente no compartan este rasgo (Castillo 2002), de forma que según el AlgCl ambos verbos conforman un par semántico.

---

<sup>3</sup> Para consultar la información precisa sobre a qué se refiere cada valor y la forma en la que opera esta versión del AlgCl , ver Castillo (2002).

<sup>4</sup> Ver Apéndice A.

<sup>5</sup> Los parámetros utilizados en cada corrida son los mismos para cada base y se pueden consultar en el Apéndice B.

<sup>6</sup> Todas las referencias de las definiciones se pueden consultar en la liga *Fuentes* de la página <http://iling.torreingenieria.unam.mx/diccionarios/>



### 2.1.1 Depuración del CA

Las corridas del AlgCl arrojaron una cantidad bastante grande de agrupamientos<sup>7</sup> en donde había grupos que estaban conformados por categorías distintas a la verbal o por grupos en donde los verbos tenían una relación sinonímica registrada.

Las estadísticas generales de los agrupamientos obtenidos con el AlgCl en las cuatro corridas realizadas se enumeran en la Tabla 2.

Bases	Total de agrupamientos <sup>8</sup>
lingüística	251
física	191
sexualidad	36
desastres	180

**Tabla 2: Total de pares generados con el AlgCl sin importar categoría gramatical.**

Una vez con estos agrupamientos seguimos un proceso para depurar cada par semántico y conservar sólo aquéllos que estuvieran integrados por verbos y, en específico, por verbos que no tuvieran una relación sinonímica ya registrada.

Bajo esta premisa, al conjunto inicial de pares le realizamos una primera depuración que consistió en la selección de los grupos que contenían sólo verbos, eliminando todo aquel grupo en donde aparecieran formas de otra categoría gramatical.

El resultado de esta primera depuración se puede observar en la Tabla 3; la estadística final en la Tabla 4.

---

<sup>7</sup> Por agrupamiento se debe entender al conjunto de palabras, sean de una misma categoría gramatical o no, que el AlgCl propone como relacionadas semánticamente.

<sup>8</sup> Las formas que en esta etapa arroja el sistema son independientes de la categoría gramatical que estamos analizando, esto es, el AlgCl devolvió pares semánticos que se integraban por verbos, sustantivos y algunos adjetivos y adverbios.

Bases	Número de agrupamientos	Total de verbos <sup>9</sup>
lingüística	21	80
física	18	51
sexualidad	6	15
desastres	6	14

Tabla 3: Resultados de la primera depuración del corpus extraído del AlgCl.

Total de agrupamientos	Total de formas verbales
51	158

Tabla 4: Estadística final de agrupamientos y forma verbales.

Con esta primera depuración redujimos de un total de 658 agrupamientos que contenían varias categorías gramaticales, a uno menor de 51 pares<sup>10</sup> que sólo estaba integrado por verbos.

A estos 51 pares restantes se les aplicó una segunda depuración que consistió en eliminar todo agrupamiento en el que los verbos que lo constituían estuvieran relacionados léxicamente.

### 2.1.1.1 Pares relacionados léxicamente

Para determinar qué verbos de los que integraban cada uno de los 51 agrupamientos tenían alguna relación léxica, analizamos en tres medios especializados si alguno de los verbos del par, en alguna de las acepciones de cualquier medio, estaba registrado como sinónimo de su contraparte.

<sup>9</sup> Aquí damos la estadística de formas localizadas con base en la distinción *type/token*. Los números de esta estadística corresponden a los *tokens*.

<sup>10</sup> Los 51 pares integrados por verbos se pueden consultar en el Apéndice C.

Los medios<sup>11</sup> en los cuales realizamos tal análisis son:

- Diccionario Larousse (Larousse Planeta, edición electrónica en CD ROM, 1996),
- Diccionario Anaya (base de datos en formato electrónico),
- Euro WordNet (Interfaz 0.1, en línea en: <http://nipadio.lsi.upc.es/cgi-bin/public/wei1.consult.perl>).

El proceso que efectuamos en cada medio fue el siguiente:

- i. consultamos la forma en infinitivo para cada verbo,
- ii. verificamos si en alguna acepción del primer verbo del par había alguna relación de sinonimia, o en su defecto, de antonimia, con el segundo verbo del par, como se observa en la Tabla 5, en donde los verbos *suplir* – *sustituir* (en versales) están relacionados,
- iii. eliminamos todo par que en alguna acepción estuviera registrado, en cualquiera de los tres medios, como sinónimo o antónimo.

Con este análisis eliminamos todos los pares que, dada su característica ya registrada de sinonimia, no interesaban al desarrollo de la investigación. Cabe destacar que de los 51 pares, hubo dos que tenían relación de antonimia<sup>12</sup>, los cuales también se eliminaron del CA.

Los pares restantes que luego de la depuración realizada no se registraban en estos tres medios como sinónimos, son los que constituyen el CA de la investigación.

---

<sup>11</sup> El uso de estos materiales, a excepción de Euro WordNet, fue motivado porque (1) son diccionarios de lengua reconocidos y, (2) estaban a disposición en formato electrónico, motivo principal para manipularlos. Respecto a EWN, la causa primordial para usarlo fue el hecho de que es una ontología desarrollada con criterios psicolingüistas que pretende mostrar cómo está estructurada la información léxica, y por tanto, el conocimiento, en nuestras mentes; además de que da información referente a las características semánticas de cada elemento que está en su base léxica, motivo que en la investigación resulta de mucha importancia.

<sup>12</sup> El total de pares que tenían relación de sinonimia y los dos que la tenían de antonimia se puede consultar en el Apéndice D.

	<b>SUPLIR</b>	
<b>Larousse</b>	<b>Anaya</b>	<b>Euro WordNet</b>
Reemplazar	E: Completar lo que falta en una cosa o remediar la carencia de ella.	possession lock 0 <i>suplir_1</i>
<b>SUSTITUIR</b>	E: Encubrir los defectos de otro.	social lock 1 <i>suplir_2</i> <b>SUSTITUIR_4</b>
	E: Reemplazar.	
	E: Sobreentender lo que está implícito en una oración o frase.	
	SIN. <B>1.</B> <b>SUSTITUIR</b> , reemplazar. <B>2.</B> Representar, suplantar, sustituir.	
	<b>SUSTITUIR</b>	
<b>Larousse</b>	<b>Anaya</b>	<b>Euro WordNet</b>
Reemplazar	E: Cambiar a una persona o cosa por otra.	change lock 5 reemplazar_1 <i>sustituir_1</i> substituir_1 reemplazar_1
	ANT. Quedarse, permanecer.	possession lock 4 reemplazar_3 <i>sustituir_2</i> substituir_2
	SIN. Reemplazar, relevar, <b>SUPLIR</b> , representar.	social 0 reemplazar_4 <i>sustituir_3</i>
		social lock 1 <b>SUPLIR_2</b> <i>sustituir_4</i>

**Tabla 5: Análisis de los verbos *suplir* y *sustituir* en tres medios.**

### 2.1.1.2 Pares sin relación léxica

Según las bases léxicas de los tres medios arriba mencionados, el CA está conformado por 56 verbos que no comparten ninguna relación léxica formalmente registrada. Tales verbos están agrupados en 28 pares semánticos que se muestran en la Tabla 6.

Estos 56 elementos que a nivel contexto y, probablemente a nivel discurso, llegan a funcionar como sinónimos, son las unidades que analizamos con el fin de argumentar la pertinencia de dichos verbos como constituyentes factibles de estructurar en la BCL del SRI que desarrollamos<sup>13</sup>.

<sup>13</sup> Para un mejor entendimiento sobre la forma en que estas unidades se comportan dentro del contexto del cual fueron extraídas, ver el Apéndice E.

registrar – representar	afectar - caracterizar	atravesar - experimentar	denotar - implicar
utilizar – hablar	considerar - importar	representar - designar	emplear - invertir
determinar – modificar	impresionar - afectar	entender - referir	continuar - preservar
recibir – realizar	describir - barrer	significar - expresar	existir – ocurrir
generalizar – usar	explicar - relacionar	señalar - expresar	propagar – viajar
desempeñar – expresar	obtener - liberar	especificar - expresar	suponer - considerar
marcar – distinguir	emplear - seguir	decir - predicar	reproducir - producir

**Tabla 6: El CA con los 28 pares de verbos.**

## 2.2 Problemática

La problemática que en el fondo analizamos con los 56 verbos que conforman el CA es la de implementar la complejidad, primero léxica y luego cognoscitiva, que representa la comunicación oral humana en la BCL del sistema que desarrollamos. Tal problemática supone en la investigación:

- A. analizar los 28 pares que, según las fuentes consultadas, no tienen relación léxica pero que a nivel contexto específico funcionan como si tuvieran una relación léxica de sinonimia;
- B. evaluar los contextos en los que aparece cada elemento del par para analizar la pertinencia y el no cambio de significado en las definiciones de las que son extraídos;
- C. constituir un corpus de prueba a partir de la extracción de contextos científicos en los que sean utilizados los verbos que analizamos;
- D. intercambiar en cada contexto, de forma manual, los 56 verbos a manera del AlgCl y evaluar si el intercambio de verbos en los contextos del corpus de prueba no produce cambio en el significado de los contextos;
- E. extraer los contextos en los que la presencia de uno u otro verbo no cambie el significado para discutir y argumentar semánticamente, en el capítulo siguiente, qué factores producen que el significado, independiente del verbo que aparezca, no cambie.

El desarrollo de estos puntos sienta las bases para argumentar unos de los objetivos específicos estipulados en la introducción: argumentar semánticamente los resultados del AlgCl.

### 2.3 Análisis en definiciones

Una vez con los resultados depurados del AlgCl reunidos en el CA, procedimos a analizar si los pares que este algoritmo había identificado, mediante reglas matemáticas<sup>14</sup> como posibles sinónimos, tenían una correlación en los contextos de los cuales fueron extraídos.

El análisis que realizamos consistió en verificar, para cada par de verbos, si al intercambiar un verbo por otro en las definiciones de las que provenía cada elemento, la codificación de la definición no sufría un cambio de significado. En a) y b) ejemplificamos el proceso que realizamos en las definiciones.

- a) Adjetivo numeral<sup>15</sup> que **significa (expresa)** multiplicación (Lázaro Carreter, Diccionario de términos filológicos, 1990).
- b) Adjetivo numeral que **expresa (significa)** multiplicación por la serie natural numérica (Alonso, Gramática del español contemporáneo, 1974).

Tanto *significar* como *expresar*, aún sin tener una relación sinonímica registrada, pueden funcionar como si la tuvieran en este contexto. Así con el resto de los pares del CA.

Un fenómeno interesante que depuramos en este análisis fue el de la puntuación. El AlgCl, tanto en su versión para el inglés como para la del español, no considera los signos de puntuación (Sierra y McNaught 2000, 2002; Castillo 2002; Sierra, McNaught y Castillo 2002). Este aspecto que no reconoce el AlgCl y que Castillo (2002) propone como una mejora para el sistema, dada la coherencia textual que implica cada signo y que no es tomada en cuenta en el momento de proponer los pares, lo realizamos de forma manual

---

<sup>14</sup> Toda la información sobre cómo opera el AlgCl se puede consultar de manera detallada en el Apéndice A.

<sup>15</sup> El término a que hacen referencia a) y b) es *adjetivo numeral múltiplo*

para eliminar las definiciones, no necesariamente los pares<sup>16</sup>, del análisis de correlación de significados. De esta forma, consideramos sólo agrupamientos de construcciones sintácticas equivalentes.

### 2.3.1 Equivalencia contextual

Los verbos del CA fueron examinados en las definiciones en las que se comportaban como sinónimos. El total de éstas integra una base léxica constituida por 28 grupos que engloban 62 definiciones<sup>17</sup>, casi una por par, en las que el significado total de las mismas no cambia sino, tal vez, por matices. En la Tabla 7 se muestran las estadísticas de la base léxica que integra el CA.

pares	28
definiciones	62
types	574
tokens	1653

Tabla 7: Estadísticas de la base léxica que integra el CA.

El resultado, luego de observar en cada definición que cualquiera de los dos verbos de los que constituyen el par semántico funcionaba como sinónimo<sup>18</sup> bajo ese contexto particular, dio pautas para destacar que ninguno de los verbos del par se puede concebir como un elemento aislado e independiente de su contexto. Por el contrario, como puntualizan Eumeridou, Nkwenti-Azeh y McNaught (2004), “they depend on context to convey their specialized meaning”.

Esta supuesta *sinonimia contextual*, sin embargo, no concluía nada más allá de señalar que dos construcciones sintácticas equivalentes pueden aceptar uno u otro verbo

---

<sup>16</sup> Recalcamos el hecho de que no se eliminaron los pares, aunque sí algunas definiciones, por el hecho de que para cada término existe desde una hasta N definiciones. Por tanto, si en alguna de ellas los pares no concordaban por la puntuación, en algunas otras no se presentaba el problema y el par funcionaba.

<sup>17</sup> En el Apéndice E se enlistan las definiciones para cada par.

<sup>18</sup> Ver 1) del apartado de 2.3

para referirse a un término particular sin modificar el significado global del concepto que describen.

### **2.3.2 Más allá del contexto específico**

El hecho que, dentro de contextos terminológicos, tanto un verbo como el otro pudieran semejar una función sinonímica, dada el mínimo o nulo cambio de significado en las definiciones de las cuales eran extraídos, no era argumento contundente y suficiente para señalar que los verbos del CA, en este caso particular, pudieran ser tomados en cuenta para incluirlos en una estructura jerárquica como elementos pertinentes en la BCL del proyecto.

Si estos elementos funcionaban como sinónimos cabía la posibilidad de que tal característica estuviese en relación con el simple hecho de que los verbos estaban dentro de un contexto bastante delimitado y definido. Es decir, lo que en la investigación denominamos *contexto específico*.

Lo llamamos así dadas dos características. La primera está en relación con que la supuesta función sinonímica de los verbos del CA se registraba sólo en las definiciones del banco terminológico. La segunda característica es que dichas definiciones están en función de terminologías bien definidas y concretas, las cuales constituyen un marco conceptual perfectamente constituido (Eumeridou, Nkwenti-Azeh y McNaught, 2004) que en sí mismo alberga una estructura léxica definida.

El punto en esta parte era aclarar que el comportamiento de estos verbos no se debía sólo al contexto específico. Si estas unidades llegaran a operar en otros dominios comunicativos de una manera semejante a la operada en los contextos extraídos de las definiciones del banco terminológico, podríamos eliminar el supuesto de que tal comportamiento respondía sólo al contexto específico y, en consecuencia, postular una hipótesis que diera cuenta de qué sucedía con estos elementos.



Para probar lo anterior, construimos un corpus prueba (CP) formado por fragmentos de textos de especialidad en el que analizamos el comportamiento del CA en otras fuentes más allá de sólo definiciones.

## **2.4 Corpus Prueba**

Este corpus lo diseñamos bajo la premisa de comprobar que la sinonimia de los pares de verbos que constituían el CA podía extenderse hacia otros contextos que no fueran únicamente definiciones<sup>19</sup>.

Si los verbos se comportaban en otros dominios comunicativos de forma similar a la mostrada en las definiciones era una muestra para señalar que, lejos de justificar el comportamiento de estos verbos en discursos especializados como algo dado por el contexto, existen plantillas subyacentes comunes a hablantes y oyentes que permiten la codificación de significados sin importar si está presente uno u otro verbo.

Además, es preciso señalar que la codificación de las definiciones no se limita sólo a la presencia de ciertos elementos o argumentos. El significado de una construcción, de acuerdo con Giulia, Bencini y Goldberg (2000), no es evidente si sólo atendemos a elementos aislados, en este caso específico, “the verb in isolation”, sino a través del conjunto; de aquí que lo que pretendemos con el CP es encontrar pautas de comportamiento en otros contextos con elementos léxicos distintos a los que privan en las definiciones.

### **2.4.1 BwanaNet**

Para conformar el CP recurrimos a un corpus ya construido por el Institut Universitari de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra (UPF) en

---

<sup>19</sup> En este sentido, es importante señalar que no por el hecho de enfrentarnos a un dominio como lo es el terminológico, en el cual por ende, hay presencia de definiciones, el comportamiento que los verbos llegaran a adquirir se supeditaba a este carácter terminológico, el cual, si bien determina en buena medida el comportamiento del léxico, también representa, como señala Fillmore (1982), un dominio de vocabulario cuyos elementos, de algún modo, presuponen una esquematización de la categorización y conceptualización humana que encuentra una representación lo suficientemente variada en el léxico presente en las terminologías.

Barcelona, España. Este corpus denominado BwanaNet es un repositorio de información especializada referente a varios dominios terminológicos, entre los que destacan el genoma humano, derecho, medicina, etc.

La información que constituye al BwanaNet es seleccionada por especialistas de cada área y, a diferencia de la información del banco terminológico, no proviene sólo de medios especializados como diccionarios y enciclopedias sino de textos de divulgación, artículos científicos, materiales de investigación, etc., lo que hace que los contextos que recuperamos para probar la pertinencia de los verbos del CA no se supedita sólo a definiciones, sino a contextos que, si bien caen dentro de un dominio de especialidad, son de un carácter mucho más amplio en cuanto a estructuras sintácticas y léxico empleado que los que privan en las definiciones.

El fin que perseguimos al utilizar el BwanaNet como fuente para recopilar el CP fue, en primer lugar, el de aprovechar la información que éste contiene y que, como arriba se señala, representaba un campo *ad hoc* para contrastar el comportamiento de los verbos del CA. En segundo lugar, beneficiarnos con las herramientas de consulta y explotación con que cuenta, dado que tales herramientas permiten buscar con precisión estructuras particulares y, en el caso de este trabajo, recuperar contextos a partir de un lema, a diferencia de las búsquedas que permite el CREA, por lo que los procesos de búsqueda y recuperación se hacen de manera más rica, rápida y efectiva.

El BwanaNet se encuentra disponible en <http://bwananet.iula.upf.edu/>. En el Apéndice F se puede consultar un poco más de información respecto a este corpus.

#### **2.4.2 Recuperación de contextos**

Dadas las herramientas del BwanaNet extrajimos para cada verbo de los 56 que constituyen el CA un total de 50 contextos. La recuperación de éstos se hizo con base en los parámetros que se detallan en el Apéndice G.

Los números totales de contextos que conforman el CP se muestran en la Tabla 8:

contextos	2800
types	7180
tokens	57783

**Tabla 8: Estadísticas del CP.**

### 2.4.3 Selección de construcciones válidas

De los 2800 contextos recuperados seleccionamos sólo las construcciones que nos permitieran hacer una comparación equitativa, esto es, en donde el contexto fuera lo suficientemente entendible como para inferir el significado de la construcción. Obsérvense los contextos de c) y d) para ejemplificar la acción realizada.

- c) Bucéfalo sustancia segunda: correspondería con lo que se **dice** de esos sujetos (sería la accidental).
- d) A pesar de ser un factor que **desempeña** un papel en la determinación de los precios relativos,...

Eliminamos construcciones como la que aparece en c), en donde no había posibilidad de codificar el significado y que, dada la escasez de contexto, no importaba qué elemento apareciera. Las construcciones como la de d) se mantuvieron para el ejercicio de análisis contextual.

## 2.5 Ejercicio contextual

Una vez que tuvimos el CP realizamos un análisis con el que pretendíamos observar el comportamiento de los elementos del par al reemplazar un verbo por otro en estos nuevos contextos.

El fin principal de este análisis consistió en obtener observaciones empíricas sobre la funcionalidad o no de los verbos en textos que no están escritos bajo un patrón

lexicográfico pero que, al igual que las definiciones, representan un conocimiento especializado.

### 2.5.1 Análisis de contextos

Una vez realizado el intercambio de verbos<sup>20</sup> en cada uno de los contextos válidos del CP procedimos a evaluar si, con la presencia de la contraparte del par, el significado podía seguir codificándose sin alteración.

La evaluación consistió en tratar de recuperar el significado original de cada construcción dada la presencia del otro verbo que constituía el par. Si la presencia indistinta de cualquiera de los dos elementos no causaba modificación en el significado global de la construcción, se calificaba como *acceptable*.

En los casos en los que había duda o el significado resultaba ambiguo, la construcción se calificaba como en *duda* y se integraba en un módulo aparte. Las construcciones valoradas como *inacceptables* fueron desechadas del análisis.

Considérense las construcciones de e), f) y g) como ejemplos del análisis realizado.

- e) ponía a cargo de la filial española a un directivo **utilizando** algo de castellano. En ambos casos...(construcción *acceptable*, verbo original *hablar*)
- f) Diversos autores (por ejemplo Lavers, 1972 ) han **hablado** (de)<sup>21</sup> la programación lineal “inversa” para investigar este problema...(construcción en *duda*, verbo original *utilizar*)
- g) produce menos de ambos outputs que el hospital C y **habla** igual cantidad de input. Con múltiples outputs, la...(construcción *inacceptable*, verbo original *utilizar*)

---

<sup>20</sup> Véase el ejemplo 1) del apartado 2.3

<sup>21</sup> Si insertamos en la construcción la preposición **de**, el significado no sufre modificación; sin embargo, dado que tuvimos que agregar esa preposición, clasificamos la construcción como en *duda*.

Realizado el análisis de los contextos del CP procedimos a evaluar cuantitativamente los resultados para proponer un análisis cualitativo bajo los términos de la gramática cognoscitiva.

## 2.5.2 Valoración contextual

La estadística de la evaluación de construcciones *acceptables* y construcciones en *duda* son las siguientes:

- Construcciones calificadas como *acceptables*: 95
- Construcciones calificadas como en *duda*: 272
- Total de construcciones: 367

En la Tabla 9 se muestran los números del análisis por par<sup>22</sup>.

	acceptables	en duda		acceptables	en duda
registrar - representar	21	20	atravesar - experimentar	5	2
utilzar - hablar	1	6	representar - designar		9
determinar - modificar	8	21	entender - referir		2
recibir - realizar	1	8	significar - expresar		4
Generalizar - usar	1	15	señalar - expresar	13	2
desempeñar - expresar		7	especificar - expresar	8	6
marcar - distinguir	3	9	decir - predicar		6
afectar - caracterizar	1	11	denotar - implicar		13
considerar - importar		5	emplear - invertir	10	12
impresionar - afectar	1	18	continuar - preservar		8
describir - barrer	8	6	existir - ocurrir		13
explicar - relacionar	2	12	propagar - viajar		12
obtener - liberar	8	11	suponer - considerar		9
emplear - seguir	3	8	reproducir - producir	1	17

**Tabla 9: La evaluación de contextos en números.**

Es interesante notar que, contra la expectativa de que los verbos del CA no operaran en otros contextos fuera del de las definiciones, las construcciones valoradas

<sup>22</sup> En el Apéndice H se presenta un extracto de las construcciones calificadas como válidas.

como *acceptables* y en *duda* tienen un nivel razonable de incidencia. En la Tabla 9, por ejemplo, se observa que a excepción de 11 de los 28 pares, en el resto hubo al menos una construcción que aceptaba ambos verbos sin alterar el significado; en algunas calificadas como en *duda*, la construcción carecía de una preposición que le permitiera al verbo insertado operar como si fuera el original<sup>23</sup>.

Un punto importante que desprendimos de esto es que no importa si el contexto permite que cualquier verbo aparezca en la construcción de forma indistinta, sino que, a partir de los números podemos apuntar que hay un indicio para pensar que algo sucede en la mente de hablantes y oyentes que permite que, independientemente de cualquiera de los dos verbos que aparezca en la construcción, el significado permanezca intacto como para poder codificar el evento.

En este sentido, extraemos dos puntos importantes del análisis cuantitativo:

- i. el verbo no determina la codificación exacta del significado aunque sea el eje de la construcción y, según algunas posturas, motive la sintaxis y la semántica de la construcción dado que no hablamos mediante unidades aisladas sino en contextos movibles en donde prevalece un significado de construcción (Giulia, Bencini y Goldberg 2000). Además, el verbo en una construcción no garantiza un comportamiento léxico pleno, más bien, aporta matices y variantes de significado e incluso significados nuevos a la comunicación;
- ii. tanto hablante como oyente tienen en mente un patrón recurrente y eficiente de comunicación, el cual extienden hacia otras construcciones a partir de esquemas, resultando construcciones a semejanza del primer patrón, logrando así la codificación del significado.

---

<sup>23</sup> Véase el ejemplo d) del apartado 2.5.1

## 2.6 Hipótesis cognoscitiva

Conforme a los resultados del intercambio de verbos en el CP y el análisis realizado en las definiciones de donde vienen los pares del CA, nos cuestionamos los puntos del apartado anterior y sugerimos la siguiente hipótesis:

No es suficiente argüir que la codificación de significados en los distintos contextos en donde se intercambiaron los verbos, tanto del CA y el CP, se deba sólo al contexto, el cual permite que aparezca uno u otro elemento indistintamente. Tanto en el hablante como en el oyente existen *patrones cognoscitivos*, desprendidos de la gramática y de la comunicación, entendida ésta como un ente social, cultural y heterogéneo, que determinan el significado real de la lengua en situaciones específicas. El estudio y análisis empírico de la lengua como filtro de la *cognición humana* podrá permitirnos, a través de la semántica, señalar que lo que focaliza el hablante y el oyente no son significados léxicos ni contextuales sino significados decodificables bajo *marcos cognoscitivos abstractos y generalizables*.

En este sentido, en el Capítulo 3 analizamos los pares de verbos del CA para señalar que, más allá de una estructura del léxico diseñada a priori sobre la base de información extraída sólo de diccionarios u ontologías, se debe tomar en cuenta el hecho de que hay elementos y procesos que a nivel discurso forman parte de una organización comunicativa y cognoscitiva más compleja, común a hablantes y oyentes, y no sólo a lexicógrafos, terminólogos o especialistas.

## **ANÁLISIS SEMÁNTICO**

---

La problemática que abordamos en este capítulo consiste en analizar, semánticamente, la presencia indistinta de cualquier verbo del par en una construcción. Para ello, nos suscribimos a una perspectiva proveniente de la gramática cognoscitiva, y en específico, a la derivada del modelo denominado Construction Grammar (CxG). Bajo esta visión cognoscitiva argumentamos, mediante ejemplos del CA y del CP, la hipótesis establecida en el capítulo anterior.

El desarrollo del capítulo consta, en su primera parte, de puntualizar los ejes principales de la gramática cognoscitiva, centrando la atención en el modelo de CxG.

Una segunda parte da una explicación, desde un plano cognoscitivo, sobre los fenómenos que están sucediendo para que tanto hablante como oyente puedan codificar, independientemente del verbo que aparezca en la construcción, el evento puesto en foco.

Por último, señalamos la pertinencia de análisis de corte pragmático para dar cuenta de forma más completa de la lengua y robustecer así la BCL del diccionario.

### **3.1 La gramática cognoscitiva**

Esta gramática pretende describir y dar cuenta de la sistematicidad de la lengua, de su estructura y sus funciones y de cómo estas funciones son realizadas, desde el plano de la cognición, por el sistema lingüístico. La tesis principal de los cognoscitivistas es que la lengua refleja patrones sobre cómo las capacidades lingüísticas se fundamentan en la cognición.

Bajo los términos de esta gramática se concibe que, a través de la lengua, se puede estudiar la cognición humana, puesto que la lengua refleja la estructura y organización que los pensamientos e ideas tienen en la mente. Por tanto, la lengua, vista desde esta



perspectiva, debe reflejar propiedades y características de la mente humana y de cómo funciona.

De igual manera, la gramática cognoscitiva establece que tanto pensamiento como conceptualización son parte fundamental e intrínseca de la cognición humana y, como tal, no están fuera del sistema lingüístico, al contrario, son parte de él. Asimismo, procesos como metáfora y metonimia no se conciben como entidades externas a la lengua sino como procesos creativos dependientes del sistema cognoscitivo humano.

Por último, según los lineamientos de esta gramática, conocimiento y experiencia permean en todo instante la lengua. Aquello que está en el mundo y es familiar a hablantes y oyentes se extiende hacia aquello que es menos familiar o tiene un alto grado de abstracción de forma que sea posible conceptualizarlo y categorizarlo.

### **3.1.1 Axiomas**

La lengua proporciona una imagen convencional de una realidad y, según Langacker (1987), debe ser:

- i. simbólica en su naturaleza;
- ii. simbólica más allá del aspecto léxico y hasta la gramática;
- iii. la gramática, a su vez, es un continuo de estructuras simbólicas dispuestas en jerarquías que se traslapan y convergen, las cuales difieren en diversos parámetros que sólo pueden fraccionarse arbitrariamente.

### **3.2 El modelo aplicado al trabajo**

Bajo la premisa de hallar patrones acerca de la forma en que la lengua refleja elementos propios de la cognición, diversos lingüistas han abordado, desde su propia perspectiva, el estudio de la lengua, enriqueciéndola y diversificándola según los objetivos y la visión que cada uno tiene de ella. Sin embargo, para esta investigación, no obstante la diversidad de teorías y modelos, sólo sustraemos las bases conceptuales derivadas del

marco que plantea el modelo desprendido de la gramática cognoscitiva, CxG; y principalmente, los conceptos de Fillmore (1982).

### 3.2.1 Construction Grammar

CxG es un modelo lingüístico desprendido de la gramática cognoscitiva que estudia la lengua y sus fenómenos incorporando los fundamentos cognoscitivos e interaccionales de la lengua<sup>1</sup> junto con elementos extraídos del conocimiento de los hablantes (Goldberg 1997).

De acuerdo con Goldberg (1997), el nivel mínimo de análisis y el punto medular para dar cuenta de la lengua es la construcción. Bajo este marco, la lengua es vista como una suma de patrones (construcciones) en donde tanto forma como significado se integran de manera convencional.

Una palabra (forma) es más que sólo sintaxis; aspectos morfológicos, fonológicos, prosódicos o de entonación están involucrados en ella. El significado de ésta, por su parte, se entiende en un sentido en el que se incluye tanto la semántica léxica como el significado pragmático y el discursivo.

Vista desde la perspectiva de CxG, la gramática es un compendio de redes y patrones, basados en los mismos principios que aquellos de las categorías conceptuales de la lingüística cognoscitiva, tales como la herencia, la prototypicalidad, las extensiones, etc., que permiten la codificación y decodificación de las expresiones lingüísticas<sup>2</sup>.

En este sentido, los principios de CxG que recuperamos para resolver la problemática planteada por los verbos del CA, son los argumentados por Goldberg (1997) que señalan que la interpretación del significado no se da en función de unidades aisladas e independientes, en este caso, los verbos, sino en función de construcciones que integran

---

<sup>1</sup> Cfr. con el sitio oficial de CxG en <http://www.constructiongrammar.org/>

<sup>2</sup> Ídem.

elementos que transgreden su esfera léxica (Nerlich y Clarke 2000) y representan significados más complejos.

### **3.2.1.1 Construcciones**

Según Goldberg, una construcción se entiende como un conjunto constituido por forma y significado, tal que algunos aspectos de la forma o algunos aspectos del significado no son previsible de las partes que la componen o de otras construcciones que ya están establecidas en la lengua. Toda construcción, señala, está relacionada con otras construcciones y, por tanto, no es arbitraria.

De igual forma establece que las construcciones no constituyen un grupo sin estructura sino un sistema integrado, basado en principios generales de categorización.

Por último, acorde con el modelo de CxG, cada construcción se trata como un todo en el cual cada miembro de la estructura es parte integral de la misma. Las palabras, en este sentido, son consideradas como construcciones dado que su forma no es previsible de su significado o de su uso<sup>3</sup>.

### **3.2.2 Frame Semantics**

En el modelo de CxG se liga un marco semántico denominado Frame Semantics, desarrollado principalmente por Fillmore (1982), el cual describe una manera de estructurar y representar el significado enfatizando la continuidad entre lengua y experiencia (Petrucci 1996).

Este marco ofrece una manera de analizar los significados de las palabras tanto como los principios que caracterizan la creación de nuevas palabras y frases (Fillmore 1982). Bajo este modelo, una palabra representa una categoría de experiencia (Petrucci 1995, 1996) puesto que toda palabra, dado un significado, evoca la esquematización de un

---

<sup>3</sup> Cfr. Goldberg 1997.

evento particular analizable a partir de marcos conceptuales referenciales o *frames*. (Fillmore, Wooters, y Baker 2001).

El significado, al igual que la conceptualización<sup>4</sup>, es parte de un proceso cognoscitivo que no reside en una realidad objetiva (Langacker 1991) sino en esquematizaciones basadas en las experiencias de los hablantes (Petrucci 1995), de forma que para entender el significado de las palabras debe existir una estructura conceptual, a partir de *frames*, que suministre un fondo de conocimiento sobre el cual el significado pueda codificarse y decodificarse.

### 3.2.2.1 Frames

El concepto de *frame* no es privativo de la lingüística; de acuerdo con Nerlich y Clarke (2000), en áreas como la psicología o la inteligencia artificial se ha adoptado este término para referirse al principio organizativo que subyace a la representación del conocimiento. Dentro del modelo de Frame Semantics se le ha utilizado para representar un sistema de conceptos estructurados y relacionados entre sí (Fillmore 1982) que dan cuenta del significado semántico de la lengua. Esto es, un *frame* es un dominio de vocabulario, cuyos elementos presuponen la esquematización de nociones del juicio y comportamiento humano (Fillmore 1982).

Según el documento de Fillmore de 1982, para poder codificar este sistema de conceptos es necesario entender X no como un elemento aislado sino como parte de un conjunto integrado desde A hasta Z cuyos elementos siempre están disponibles en contextos específicos. Es decir, para entender un concepto es preciso entender la estructura total en la cual aparece bajo el contexto apropiado<sup>5</sup>.

---

<sup>4</sup> En este sentido, el término conceptualización se entiende tal y como Langacker (1991) señala. Esto es, como un dominio cognoscitivo.

<sup>5</sup> Esto es, *frame* debe ser entendido como un Modelo Cognoscitivo Idealizado (ICM por sus siglas en inglés), un escenario, un esquema, etc., y en donde hablar de un elemento de un *frame* es hablar de  $\eta$  número de elementos relacionados que se activan y se vuelven disponibles para suplir tal elemento en contextos específicos en el mismo *frame*.

La suposición que subyace a esto es que nuestra estructura conceptual está integrada a partir de *frames* y que éstos no constituyen cadenas aisladas de conocimiento, sino una red compleja de conceptos estructurados y referenciales analizables desde el plano de su semántica<sup>6</sup>. Bajo esta perspectiva, es claro que todo *frame* representa la conceptualización de una realidad particular que está en relación con situaciones periféricas ancladas a esa misma realidad (Langacker, 1987).

### 3.3 Preámbulo y discusión

Las definiciones en las que aparecen los verbos del CA así como los contextos del CP no son evidencia formal, pero tampoco casual, de que los verbos que conforman el par puedan funcionar como sinónimos sin que lo sean.

Es cierto, como apunta Fillmore (1982), que el contexto es una parte fundamental para codificar el significado de cualquier construcción, pero justificar y argumentar los elementos del CA en función de él, significa reducir el fenómeno a una cuestión mínima y trivial. Creemos que este fenómeno va más allá del simple contexto.

Los verbos del CA aparecen en un rango aceptable de construcciones, desempeñando una relación sinonímica que al parecer de tres fuentes especializadas no existe, gracias a la capacidad cognoscitiva de proyectar y extender plantillas que codifican información léxica que, aunada a la información extralingüística propia de cada hablante, permite la decodificación del significado sin importar qué elementos aparezcan en una construcción.

La discusión, entonces, es que este fenómeno subyace en la cognición.

Si esto es cierto, el conocimiento que podamos desprender de analizar esta clase de fenómenos nos representaría una gran utilidad en el trabajo de modelado lingüístico

---

<sup>6</sup> Para una explicación más detallada tanto del modelo como de su aplicación para el español, consultar Cristóbal 2001 y Bretones, Cristóbal e Ibarretxe 2001.

puesto que nos permitirá incorporar información, como la que codifican los verbos del CA, en la BCL de nuestro SRI.

### 3.4 Análisis

Establecido el punto que queremos abordar con nuestra discusión y determinado el marco teórico, presentamos el análisis realizado a cada uno de los 28 pares que conforman el CA.

La metodología que seguimos para efectuar el análisis se fundamenta en dos estratos. El primero se basa en extraer, para las definiciones y/o contextos del CP, el posible significado que priva en la construcción. En este punto queremos aclarar que por significado entendemos aquella estructura que, de acuerdo con Langacker (1991), no reside en una realidad objetiva sino en un dominio cognoscitivo que, a la hora de un acto comunicativo, evoca estructuras conceptuales que enmarcan y dan sentido a las palabras y construcciones.

El segundo estrato propone, en función de cada uno de los pares de verbos, un *frame* que permita capturar el significado que prevalece en la construcción. Con estos *frames*, que estarán etiquetados acorde con la categoría que estamos analizando, buscamos establecer plantillas que representen el significado que existe *independently of particular verbs* (Goldberg 2003a).

Cabe destacar que, para nuestros intereses, pretendemos contar con pocas categorías; por tal motivo, buscamos que cada *frame* sea lo más general posible de forma que nos permita, a manera de hiperónimo de la lengua, jerarquizar la información a través del léxico<sup>7</sup>.

---

<sup>7</sup> Es preciso anotar que en muchos de los ejemplos que más abajo analizamos, retomamos el estudio que se ha hecho en distintas investigaciones en el Grupo de Ingeniería Lingüística; por tal motivo, tanto análisis como *frames* y significados que proponemos para los verbos del CA deben entenderse bajo los resultados de tales investigaciones.

Estos resultados pueden consultarse en los trabajos de Sierra y Alarcón (2002); Aguilar y Sierra (2003); Sierra et al. (2004) y Aguilar (2004) en liga *Publicaciones* de la página: <http://iling.torreingenieria.unam.mx/>

Dicho lo anterior, presentamos a continuación nuestro análisis.

1) *atlas lingüístico*<sup>8</sup>

- a) Conjunto de mapas que **registran** la distribución de las varias lenguas en el ámbito de una determinada área, una región, una nación, etc.
- b) Conjunto de mapas que **representan** la distribución de las lenguas y dialectos para un área geográfica dada.<sup>9</sup>

En 1) podemos desprender un significado abstracto de *ubicación*. Este significado está dado si concebimos que las lenguas se mueven, abstractamente, de un plano de codificación como entidades externas, únicas y delimitadas a uno en el que se ubican como constituyentes de un conjunto nuevo, englobado por el término de 1).

Atendiendo a este significado, los pares de 1a) y 1b) involucran un *frame* que denominamos de PONER. Este *frame* codifica un evento como el expresado en 1c).

- c) [X (lengua) se PONE en [Y (mapas) resultando [Z (atlas lingüístico)]]]

Visto bajo la perspectiva de este *frame*, el par proporciona conocimiento más accesible para el aparato cognoscitivo, esto, aunado a la plantilla básica de significado que permea las construcciones, maximiza la codificación del significado del término.

El mismo proceso puede aplicarse en 1d). Para codificar de forma más rápida y segura la información proporcionada podemos recurrir a un *frame* que organice la información que se está transmitiendo.

- d) A pesar de que en la matriz solamente se representan dos países, el resultado del juego es perfectamente claro

2) *bilingüismo*

---

<sup>8</sup> Para consultar las fuentes de donde provienen las definiciones de los términos, consultar el Apéndice E. Información más detallada sobre el material terminológico puede consultarse en la dirección: <http://iling.torreingenieria.unam.mx/diccionarios/>

<sup>9</sup> En todos los casos los incisos a) y b) pertenecen a las definiciones de los términos. La aparición de otros incisos dentro del mismo ejemplo corresponde a construcciones extraídas del CP.

- a) Capacidad de un individuo para **hablar** indistintamente dos lenguas con igual dominio
- b) Aptitud de un hablante para **utilizar**<sup>10</sup> indistintamente dos lenguas

El caso de 2) proyecta un significado de *transmisión*. Esta interpretación se facilita si conceptualizamos que el término focaliza una noción comunicativa.

En este sentido, este significado está motivado por un *frame* denominado de COMUNICAR. Los verbos del par pueden aparecer en la construcción indistintamente gracias a que tanto a hablar como a utilizar subyace un sentido más abstracto como el que posee un verbo como comunicar. Ambos verbos son una representación de un evento en el que se focalizan características vinculadas con una acción más básica, la cual adoptan para propiciar un significado más complejo que el que propician de forma aislada.

En 2c) esquematizamos el significado básico que transmite el término *bilingüismo*:

- c) [X (hablante, individuo) COMUNICA (habla, utiliza) [Y (dos lenguas)]]].

### 3) *adjetivo atributivo*

- a) Adjetivo que **determina** a un sustantivo sin palabra de unión
- b) Adjetivo que **modifica** a un sustantivo sin ninguna palabra de unión

Para este caso, el significado básico que podemos extraer es el de *descripción*. La información que codificamos a partir de las construcciones de 3a) y 3b) es sólo la representación de ciertas características que constituyen el término.

La presencia de los elementos del par corresponde a un *frame* llamado de CARACTERIZAR. Lo que este *frame* codifica es un evento en donde se describe X a partir de

---

<sup>10</sup> En algunas definiciones el significado está dado por un conjunto del tipo V + O (*comer gallo* o *perder la cabeza*) y, por tanto, el emparejamiento de significados está en función de concebir la construcción como un conjunto y no como un solo elemento. Considérense los ejemplos de 2): *utilizar* + *lengua* y más adelante el de 10): *impresionar/afectar* + *ojo* como muestra.



sus propias características. En este sentido, éste y el resto de *frames* no señalan que haya un significado nuevo que pueda codificar cada verbo en aislado sino, como apunta Goldberg (1996), lo que hay es un sentido innovador que adquiere representación a partir de la construcción en la que aparecen tales elementos.

- c) [X (adjetivo atributivo) CARACTERIZA (determina, modifica) [X (adjetivo atributivo) dadas la presencia de [Y (sustantivo) y [Z (sin palabra de unión)]]]]

#### 4) *sujeto*

- a) Elemento o sintagma nominal que **recibe** la acción expresada por el verbo
- b) Posición sintáctica del actante que **realiza** la acción expresada por el verbo

En las definiciones de 4a) y 4b) sucede algo interesante. La información proporcionada para definir el término *sujeto* proviene de ángulos distintos. Cada característica en aislado constituye un rasgo pertinente del término, no obstante poner al mismo nivel *sintagma nominal* con *posición sintáctica del actante* significa mezclar enfoques y niveles.

Ahora bien, el punto que aquí se trata va más allá de justificar una relación para este par de verbos. El fondo es que, si bien es cierto que cognoscitivamente podemos crear significados y extenderlos, siempre dentro de cierto margen, hacia una gama considerable de elementos de la realidad, una máquina no podrá semejar esta capacidad y, en consecuencia, será incapaz de inferir que, para este caso, se trata de dos verbos que están complementando un significado vía dos características.

Por tanto, aunque a nivel construcción 4a) y 4b) contribuyan a un significado común, el querer postular un *frame* que haga referencia a un evento común de este par, sería forzar y obligar a una argumentación que, en sí, no tiene cabida.

#### 5) *laísmo*

- a) Tendencia a suprimir la diferencia de funciones entre objeto directo e indirecto en beneficio de la distinción de género que consiste en **generalizar** *la* y *las* para las referencias a sustantivos femeninos en la función de objeto indirecto

- b) Construcción vulgar e incorrecta que consiste en **usar** *la* y *las* del pronombre ella tanto para el acusativo como para el dativo

El argumento para 5) consiste en conceptualizar un significado de *aplicación* que da como consecuencia un resultado. Las definiciones de 5a) y 5b) favorecen una lectura en la que se focaliza, a partir de una aplicación, el resultado de las construcciones, es decir, el término *laísmo*. De manera más sencilla, el significado de 5a) y 5b) se resume en que *laísmo* es el resultado del *uso* de *la* y *las* en cierto tipo de construcciones.

Ahora bien, el *frame* que coadyuva a este significado y que atañe al par, es el de PONER. El esquema de este *frame* aplica también para este par, *generalizar* y *usar* son parte de una estructura más básica como la que se propone en 5c).

- c) [X (construcción) PONE (<sub>usa, generaliza</sub>) [Y (la, las) en [Z (dativo, acusativo)]]]

El resultado de esta construcción es el término *laísmo*.

6) *predicado verbal*

- a) Predicado que se **desempeña** por un verbo  
b) Predicado que se **expresa** por un verbo

El significado que proponemos para 6a) y 6b) es el de *descripción*. Los elementos léxicos de ambas construcciones propician una lectura en la que el foco es concebir un término a partir de la descripción de rasgos pertinentes de éste. Claro está que cada construcción varía en grado de especificidad y nivel de descripción (Bretones, Cristóbal e Ibarretxe 2001) que va de la mano con los elementos que integran la definición. En este caso, *desempeñar* y *expresar* son las unidades que motivan dicha gradación.

Por otro lado, creemos que *desempeñar* y *expresar* están en función de un *frame* que llamamos de CARACTERIZAR. El par hace alusión a un sentido menos específico que el que posee cada uno. Esto es, el marco aúna información que se vincula a rasgos propios de una acción que léxicamente y en aislado cada verbo no transmite. El hecho de que ambos

verbos en una construcción codifiquen el significado arriba mencionado está en relación con la información que el *frame* puede englobar y que permite que aparezcan rebasando sus fronteras de significado.

En 6c) ejemplificamos el *frame*.

- c) [X (predicado) se CARACTERIZA (desempeña, expresa) por [Y (verbo) resultando [Z (predicado verbal)]]]

7) *caso ergativo*

- a) Caso que **marca** al agente de los verbos transitivos
- b) Caso que **distingue** al sujeto activo de un verbo transitivo del sujeto de un verbo intransitivo en las lenguas ergativas

El significado subyacente a 7a) y 7b) es el de *descripción*. Por un lado la primera construcción enfatiza una idea de señalar ciertas características, en tanto que la segunda, focaliza rasgos para diferenciar. El *frame* que incide en este significado es el de CARACTERIZAR. Ambas definiciones describen el término en función de lo que arriba señalábamos, enfatizar y diferenciar rasgos:

- c) [X (caso ergativo) CARACTERIZA (marca, distingue) [Y (agente)]]

La misma constante puede verse en 7d): *distinguir* y *marcar* están en función de un *frame* como el de CARACTERIZAR; la presencia de éste facilita la codificación de una misma escena evocada a partir de diferentes elementos (Fillmore 1982) puestos en perfil.

- d) utilizar (los recursos) [X] es el objetivo económico inaplazable que CARACTERIZA (distingue, marca) a los (países subdesarrollados) [Y] de los que no lo son

8) *base de articulación*

- a) Conjunto de hábitos articulatorios y de características acústicas que **afectan** a todos los elementos del plano de expresión de una lengua
- b) Conjunto de hábitos articulatorios que **caracteriza** a todas los sonidos de una lengua

Hay un significado básico de *posición* en las construcciones.

Los *hábitos articulatorios* y las *características acústicas* son elementos que, puestos en perfil, definen *base de articulación*; los sonidos permanecen en fondo.

El *frame* que aplica a este par es el de PONER. De esta forma tenemos que tanto los elementos puestos en perfil como los que están en fondo, crean el marco sobre el cual actúa el *frame* de PONER, favoreciendo la codificación del significado base. Cabe destacar que en este *frame*, como apunta Fillmore (1982), lo que sucede es una extensión de un esquema básico a uno complejo al que se le transfieren rasgos que diferencian y caracterizan un significado.

9) *cinemática*

- a) Parte de la mecánica que estudia el movimiento **sin considerar** las causas que lo producen
- b) Es la parte de la mecánica que estudia el movimiento de los cuerpos **sin importar** las causas que lo producen

En 9a) y 9b) plantemos un significado primario de *descripción*.

Al igual que en los ejemplos anteriores, en donde el significado proviene de un sentido descriptivo, los elementos presentes en las construcciones esquematizan una estructura integral en donde se privilegia una construcción descriptiva. En este sentido, estamos ante la simple puntualización de un estado. No hay un resultado ni un cambio de estado, sólo la descripción de un evento estático.

Ahora bien, respecto a los verbos de este par<sup>11</sup>, el *frame* que sugerimos es el de CARACTERIZAR. El significado que entraña este *frame* engloba los significados aislados de *considerar* e *importar*; así, la presencia indistinta de cualquiera de ellos en la construcción

---

<sup>11</sup> En este caso el par de elementos analizados no está en función de dos elementos aislados sino de una cadena que involucra la preposición (sin) que califica el sentido de los verbos.

se debe a que el marco actúa habilitando significados más complejos que los de estos elementos en aislado.

En 9c) detallamos el esquema de tal *frame*.

- c) [X (cinemática) estudia [Y (movimiento) [*sin* CARACTERIZAR (considerar, importar) [Z (causas)]]]]

10) *luz*

- a) Energía radiante en forma de ondas de longitud tal que **impresionan** al ojo humano y producen la sensación llamada visión
- b) Podemos definir la luz como aquel fenómeno físico que **afecta** al ojo humano produciendo lo que llamamos la visión

El significado base de las construcciones de 10) es de *percepción*. Dicho significado está motivado por la suma de las partes que interactúan, a nivel construcción, con un *frame* específico. La codificación de este significado estriba en, como explica Fillmore (1982), que uno debe computar, en su imaginación, la información disponible a partir de un esquema pertinente que dé sentido al significado que se desprende de la construcción.

Bajo esta perspectiva tenemos que *energía radiante* y *fenómeno físico*, en perfil, y *ojo humano* en fondo, producen la lectura de un significado integral de percepción. Este mismo significado lo podemos observar en 10c).

- c) los fuegos artificiales japoneses que en verano impresionaron el público durante 40 minutos

Establecido el significado, el *frame* que resuelve los pares es de CAUSAR. Dados los elementos léxicos de las construcciones se codifica un evento como el que se describe en 10d).

- d) [X (energía radiante) CAUSA [Y (visión) en [Z (ojo)]]]

11) *segunda ley de Kepler*

- a) La línea entre un planeta y el sol **describe** áreas iguales en el mismo tiempo
- b) La línea imaginaria que une a un planeta con el sol **barre** áreas iguales en tiempos iguales

El significado de 11) es de *movimiento*. La escena evocada por ambas construcciones se ancla a un proceso en donde *línea*, en perfil, y *planeta* y sol, como puntos de referencia, privilegian una conceptualización de trayectoria, la cual implica movimiento.

Este mismo proceso puede verse en 11c).

- c) Los vientos que barren la sabana africana depositan parte de ese polvo

Dada esta interpretación, el *frame* que la facilita es de RECORRER. Su esquema se presenta en 11d)

- d) [X (línea) RECORRE (*barre, describe*) [Y (área)]]

12) *teoría*

- a) Conjunto organizado de ideas o leyes que sirven para **explicar** determinado orden de fenómenos
- b) Serie de las leyes que sirven para **relacionar** determinado orden de fenómenos

Con las construcciones de 12) se puede desprender un significado de *cognición*. Éste se da si asumimos que las *leyes* o *ideas*, en perfil, y los *fenómenos*, en fondo, proyectan, en conjunto, la concepción del término *teoría*.

Por otro lado, los verbos de este par se resumen en un *frame* de CONCEPTUALIZAR.

- c) [X (leyes) sirven para CONCEPTUALIZAR (*explicar, relacionar*) [Y (fenómenos)]]

13) *energía nuclear*

- a) La **obtenida** por la fusión o fisión de núcleos atómicos
- b) La **liberada** por la desintegración de los núcleos de los átomos

El significado base que aplicamos a 13a) y 13b) es de *movimiento*. Si atendemos a una lectura en donde los *átomos* se mueven de un estado inicial a uno meta para producir *energía*, podemos argumentar que el movimiento es el punto focal de las construcciones.

Este significado se codifica gracias a un *frame* de CAUSAR. En 13c) se señala el esquema.

- c) [X (fusión atómica) CAUSA (libera, obtiene) [Y (energía nuclear)]]

Dentro de estos contextos, *liberar* y *obtener* representan un concepto más cohesionado de información que se puede codificar mediante el *frame* propuesto.

Otro ejemplo del *frame* que está operando para el par en cuestión se registra en la construcción 13d).

- d) Esta es precisamente la que se CAUSA (libera, obtiene) cuando se quema carbón, madera o petróleo

#### 14) *anticonceptivo*

- a) Métodos con productos o con procesos que se **emplean** para evitar un embarazo  
b) Métodos y procesos que se **siguen** para evitar la concepción de un ser

Lo que sugerimos para las construcciones de 14) es una lectura primaria de *resultado*. Para conceptualizar este significado es preciso concebir que *procesos* y *métodos* son las formas a las cuales se ancla la forma *evitar*. El conjunto establece una relación de significados más complejos que apuntan hacia un evento en donde el resultado es no causar *embarazo* o *concepción*.

El *frame* es de CAUSAR, sólo que con un modificador negativo. Tanto *emplear* como *seguir* se suscriben a un significado más primario en donde se privilegia un evento causativo. La plantilla se representa en 14c).

- c) [X (procesos) ~CAUSAN [Y (embarazo)]]

Bajo la perspectiva causativa, el tercer argumento que supone el *frame* está elidido a nivel superficial, no así en la estructura profunda<sup>12</sup>.

15) *adolescente*

- a) Persona que **atraviesa** por la edad del cambio entre ser niño y adulto
- b) Persona que **experimenta** los cambios correspondientes a la adolescencia

A las construcciones que aparecen en 15) subyace un significado de *cambio de estado*. El evento focalizado por las construcciones se puede parafrasear señalando que X pasa de un estado A hacia uno B. En el ejemplo de 15c) puede apreciarse este significado subyacente.

- c) los trabajadores deben adaptarse a un mundo que está experimentando profundos cambios.

Lo que codifica 15c) es que los *trabajadores* deben pasar de un estado de *mundo* A hacia uno B. Asimismo, en 15a) y 15b) se debe codificar que el término *adolescente* no es más que la transición de un estado A hacia uno B.

El *frame* es, por tanto, uno de MOVER. La característica principal de este *frame* es la de movimiento abstracto y reflexivo. Al igual que el significado que subyace en las construcciones arriba mencionadas, lo que este par de verbos implica es que X se mueve sobre sí mismo del estado Y al estado Y'.

En 15d) ejemplificamos el esquema del *frame*.

- d) [X (persona) [SE MUEVE de [Y (niñez) a [Y' (adolescencia)]]]]

16) *sistema afectable (sa)*

- a) sa **representa** cualquier sistema integrado por el hombre y sus elementos necesarios para subsistir

---

<sup>12</sup> El tercer argumento que no se expresa en el esquema puede representarse mediante un argumento *dummy* [Z (humano)].



- b) **sa designa** cualquier sistema integrado por el hombre y los elementos que necesita para subsistir: ciudad, regiones productoras de elementos, etc.

La lectura que hemos hecho sobre las construcciones de 16) es de *descripción*. El *frame* que sugerimos para resolver el par es uno de CARACTERIZAR.

Tanto *representar* como *designar*, en conjunto con el resto de los elementos de las construcciones, lo único que hacen es describir un estado. Por tanto, el verbo caracterizar, que aquí lo hemos adoptado como un verbo específico que entraña, por ejemplo, el significado de describir y definir, puede representar de buena manera las cualidades expresadas por los verbos del par.

- c) [X (sa) CARACTERIZA [Y (sistema integrado por el hombre)]]

Lo que la plantilla de 16c) dice es que un *sa* está caracterizado por ciertas características presentes en Y. La misma argumentación puede aplicarse en 16d).

- d) el término gana cierta precisión si hacemos constar que para nosotros [CARACTERIZA designa, representa] aquella clase de planificación

#### 17) *desastre*

- a) Por desastre se **entiende** a los estados mismos de daño
- b) No sólo se **refiere** a los estados mismos del daño sino a todas las consecuencias adversas que se caracterizan por múltiples alteraciones -y hasta rupturas- del orden normal de las relaciones productivas comerciales sociales y políticas de la sociedad

El significado que permea las construcciones de 17) es, según nuestra interpretación, de *cognición*. El término *desastre* es conceptualizado a partir de la suma de significados y características descritas por el léxico de las construcciones.

Partiendo de esta presuposición, el *frame* que proponemos para este par es el de CONCEPTUALIZAR. Tanto *entender* como *referir*, así como el significado global de las

construcciones, apuntan hacia un sentido más elemental como es el de conceptualizar algo dadas ciertas características.

En 17c) diseñamos la plantilla para el *frame* propuesto.

- c) [X (sujeto elidido) CONCEPTUALIZA (entiende, refiere) [Y (desastre) a partir de [Z (características del desastre)]]]

18) *adjetivo numeral múltiplo*

- h) Adjetivo numeral que **significa** multiplicación  
i) Adjetivo numeral que **expresa** multiplicación por la serie natural numérica

*adjetivo numeral partitivo*

- j) Adjetivo numeral que **expresa** división  
k) Adjetivo numeral que **significa** división

*presente*

- l) Tiempo verbal que **significa** la coexistencia del atributo con el momento en que proferimos el verbo  
m) Tiempo verbal que **expresa** la coexistencia de la acción con el acto de la palabra

Independientemente del término a que hacen referencia las construcciones de 18), el significado que se puede extraer de todas ellas es el de *descripción*. El *frame* que facilita este significado y resuelve los pares es el de CARACTERIZAR. Por tanto, el significado primario del par y de las construcciones en las que aparecen es sólo el de describir una entidad mediante la predicación de ciertas características.

En 18g) se puede ver con más claridad esta propuesta.

- n) La noción de progresiva liberalización [CARACTERIZA significa, expresa] el compromiso por parte de los Miembros

19) *grado positivo*

- a) Grado que **señala** la simple cualidad sin indicar intensidad
- b) Grado que **expresa** la cualidad de un objeto tal como es y sirve para la formación de otros grados de comparación

En las construcciones de 19), el significado que aplica también es de *descripción*. En consecuencia, y siguiendo el patrón de nuestra argumentación, el *frame* que envuelve los verbos en cuestión es el de CARACTERIZAR<sup>13</sup>.

La pareja *señalar* y *expresar* se adscribe a un significado más complejo que el que codifica en aislado; por tanto, el que aparezca uno u otro en la definición se debe a que un sentido menos elaborado, representado por el *frame*, aunado al conjunto que representa la construcción, permite la correcta codificación del significado del término. No es que los verbos pierdan sus características sino que adquieren, para este contexto, rasgos que maximizan el proceso de codificación del significado.

## 20) *artículo*

- a) Es un morfema exclusivo del sustantivo que **expresa** el género y el número del mismo
- b) Clase de determinante actualizador del nombre virtual del que **especifica** el género y el número

Igual que en los ejemplos anteriores, el significado primario de 20) es el de *descripción*. En cuanto al par, el *frame* que opera es el de CARACTERIZAR, descrito en 20c).

- c) [X (artículo) CARACTERIZA (<sub>expresa, especifica</sub>) [Y (género, número) de [Z (sustantivo)]]]

Cabe mencionar que en las construcciones 18) a 20) encontramos una constante. El significado básico que debe ser codificado es el de *descripción* de una entidad. El valor que cambia es el de los grupos de pares que pueden ser intercambiables. Es de notar que

---

<sup>13</sup> Es de notar que una buena parte de los significados como de los *frames* que hemos propuesto recaen en un mismo patrón que podríamos llamar definitorio. El hecho de que es un corpus de definiciones dispara que los significados de perfil para estas construcciones tiendan hacia un mismo punto: definir un término. Lo anterior constriñe la gama de posibilidades para señalar *frames* que licencien dichos significados y, como resultado, este sentido de descripción es el que priva en las construcciones.

también dentro de este grupo hay una constante, *expresar*, que también aparece en 6). La contraparte en las construcciones incluye los verbos *significar*, *señalar*, *especificar* y *desempeñar*, respectivamente. Este conjunto de verbos cumple la función de matizar, especificar o dar mayor información que facilite la codificación del significado, y esto se logra gracias al *frame* que maximiza los procesos de integración de significados.

En este sentido, es obvio, como se ha dejado ver en la investigación, que los *frames* que proponemos para explicar estos fenómenos sólo pueden ser aplicables bajo un punto de vista en donde se enfoquen eventos abstractos que den lugar y originen significados más complejos. A fin de cuentas, retomando las palabras de Ghidini y Giunchiglia (2001): “all the contexts are a replication of the same context”, por tanto, podemos establecer que todo significado específico es motivado por uno más general.

#### 21) *predicado*

- a) Conjunto de signos lingüísticos que manifiesta aquello que se **dice** del sujeto de la oración
- b) Lo que se dice -lo que se **predica**- del sujeto

En este ejemplo destacamos que el significado es de *comunicación*. Extraemos tal significado rescatando una lectura en la que el conjunto integrado por los *signos lingüísticos* (elemento de perfil) comunica una característica del *predicado* (elemento de fondo).

El *frame* que da cuenta de este par y fomenta el significado de comunicación es el de EXPRESAR. En 21c) definimos la plantilla para este *frame*.

- c) [X (signos lingüísticos) EXPRESA (dice, predica) características de [Y (sujeto)]]

#### 22) *presente actual*

- a) Presente que **denota** una acción que coincide con el momento de hablar
- b) Presente que **implica** una acción continua ya que dentro de su duración se incluye el momento en que hablamos

El significado de este ejemplo es de *comunicación*. La aparición de uno u otro verbo se debe a un *frame* en el que se focaliza un sentido de EXPRESAR.

La plantilla del *frame* se establece en 22c).

c) [X (presente) EXPRESA (denota, implica) [Y (acción)]]

En este *frame*, *denotar* e *implicar* comprenden un significado que se motiva por la esquematización de significados referenciales que dan lugar a la interpretación del término en cuestión. Un evento como el de expresar algo puede dar cuenta, de forma más abarcadora, de la presencia de cualquiera de estos dos elementos en las definiciones de 22a) y 22b) puesto que no se está focalizando el significado léxico de cada elemento sino la integración de éstos en un significado cohesionado por la suma de las partes. Tal es el caso del ejemplo que referimos en 22d).

d) Ya en mis mismas palabras se [EXPRESA denota, implica] que esto no es así, cosa que debiéramos resolver

23) *velocidad*

- a) Relación entre el espacio recorrido por un móvil y el tiempo **empleado** en recorrerlo
- b) Relación entre la cantidad de movimiento, de trabajo etc., realizados y el tiempo **invertido** en realizarlos

En este ejemplo se repite el fenómeno de 4). Las características puestas en perfil en cada definición no hacen referencia al objeto que está en fondo desde el mismo ángulo. Cada característica focaliza aspectos distintos del término. El significado, insistimos, puede ser general a 23a) y 23b), no así un *frame* que pretenda integrar este par de verbos a un significado común.

Ahora bien, si los pares de 23) y 4) no pueden resolverse mediante una estructura de *frames* dadas las circunstancias en las que aparecen, el objetivo es analizarlos y

explicarlos a través de otro nivel lingüístico, es decir, desde la pragmática. Para ello, retomamos esta problemática y la discutimos en el siguiente capítulo.

#### 24) *primera ley de newton*

- a) Todo cuerpo **continúa** en su estado de reposo o movimiento rectilíneo uniforme a menos que sea impelido a cambiar dicho estado por fuerzas que actúen sobre él
- b) Todos los cuerpos **preservan** su estado de reposo o movimiento uniforme a menos que se le aplique una fuerza exterior

El significado global de las construcciones es el de *describir un estado*. El *frame* que facilita la codificación de este significado es el de POSEER. Si X POSEE Y, entonces X adquiere todas las características de Y. El resultado, por tanto, es que mediante un *frame* en el que se focaliza un sentido básico de posesión, se describa un estado, el cual puede dejar de ser si se aplica Z.

- c) [X (cuerpo) POSEE [Y (estado de reposo o movimiento rectilíneo uniforme) a menos que [Z (fuerza exterior)]]]

#### 25) *fuerza de fricción cinética*

- a) Existe una fuerza de fricción llamada fuerza de fricción cinética cuando **existe** un desplazamiento de un cuerpo sobre otro
- b) Es la fuerza que existe cuando **ocurre** un desplazamiento de un cuerpo sobre otro

Para los pares de 25a) y 25b) hemos considerado un significado que codifica un evento de *movimiento*. La información que se ha puesto en perfil en las definiciones (*fuerza* y *desplazamiento*) manifiesta un concepto que implica un movimiento. El punto de referencia (*cuerpo*) refuerza este significado.

El *frame* que explica que *existir* y *ocurrir* aparezcan en estas construcciones fortaleciendo el significado de movimiento es de CAUSAR.

- c) [X (desplazamiento cuerpo 1) sobre [Y (cuerpo 2) CAUSA [Z (fuerza de fricción cinética)]]]

26) *teoría electromagnética de la luz*

- a) Maxwell desarrolló la teoría del campo electromagnético; demostró que la acción electromagnética se **propaga** en formas de ondas transversales y que las ondas luminosas son de naturaleza electromagnética.
- b) Lo que la teoría electromagnética de la luz quiere demostrar es que la luz **viaja** en forma de ondas electromagnéticas (en sí, son ondas electromagnéticas) y que no modifica en lo absoluto sus características o propiedades el medio.

La lectura que sugerimos para las construcciones de 26) es también de *movimiento*. Lo que subyacentemente se comunica con estas definiciones es un evento abstracto que implica movimiento. Esto es, para codificar el término de 26) el oyente debe activar un esquema elemental de movimiento. En este sentido, el *frame* para *propagar* y *viajar* es de MOVER.

Cabe señalar que si bien, tanto significado como *frame* convergen en un mismo concepto, la diferencia que existe entre éstos es que, mientras el significado de movimiento permea toda la construcción a partir de los elementos que componen la definición, en el *frame* se especifica un concepto de movimiento, digamos reflexivo e intrínseco, aplicable sólo a los verbos *propagar* y *viajar*.

En 26c) exponemos el esquema para este *frame*.

- c) [X (luz) [se MUEVE (<sub>propaga, viaja</sub>) en forma de [Y (ondas)]]]

27) *eje*

- a) Recta alrededor de la cual se **supone** que gira una línea para engendrar una superficie o una superficie para engendrar un sólido
- b) Recta fija alrededor de la cual se **considera** que gira una línea para engendrar una superficie o una superficie para engendrar un cuerpo geométrico

El significado que se toma en este ejemplo es de *movimiento*. Tal sentido proviene de un esquema más general de *estaticidad*. En este sentido, el valor proyectado de

estaticidad, codificado por el elemento (*recta fija*), procede de una expresión que implica movimiento (*línea que gira*). Esto es, el significado básico no es arbitrario puesto que una noción menos particular lo integra y le da sentido.

Ahora bien, en relación con el par, el *frame* de CONCEPTUALIZAR da cuenta de la función que están desempeñando *suponer* y *considerar* en las definiciones. Este sentido más abstracto del *frame* provee la información necesaria que hace que los verbos del par tengan y aporten significado a la construcción.

- c) [X (eje) se CONCEPTUALIZA (*supone, considera*) a partir de [Y (recta) y [Z (línea)]]]

La plantilla delineada en 27c) es la misma que se puede aplicar en 27d).

- d) [CONCEPTUALICEMOS *supongamos, consideremos*] que existen siete centros (A G) tales que...

## 28) *deforestación y desertificación*

- a) Cambio ecológico que despoja a la tierra de su capacidad para sostener y **reproducir** vegetación, actividades agropecuarias y condiciones de habitación humana
- b) La desertificación se define como un cambio ecológico que despoja a la tierra de su capacidad para sostener y **producir** vegetación, actividad agropecuaria y condiciones de habitación humana

El significado básico en 28) es de *transformación*. La información que comunica el léxico de las construcciones proporciona los elementos necesarios para concebir un evento en el que existe un cambio. Esto es, el término se entiende gracias a la información puesta en perfil (*cambio ecológico*) que interactúa con aquella de fondo (*tierra*) más todo el cúmulo de información que proporciona un contenido más amplio de conocimiento.

Ahora, respecto a los verbos de este par hay que señalar que estos elementos se originan de una misma raíz, por lo cual podrían estudiarse como relacionados. Sin embargo, como ya mencionamos, dicha relación no está establecida, al menos, en los medios



revisados. Consecuencia de esto es que argumentamos su comportamiento en las definiciones mediante el *frame* de 28c).

c) [X (cambio ecológico) CAUSA (<sub>produce, reproduce</sub>) [Y (desertificación) en [Z (tierra)]]]

El *frame* descrito en 28c) es de CAUSAR. De manera que tanto *producir* como *reproducir* pueden aparecer indistintamente en estas construcciones, cada uno especializando o reforzando el significado de la definición, a través de un marco más general de CAUSA que encierra los significados particulares de éstos.

Ahora bien, con base en la argumentación que hemos sugerido para los pares del CA, presentamos en la Figura 3 un modelo del esquema léxico que comprendería un *frame*.

Dado un *frame* como el de PONER, la información que éste codificaría en la BCL iría desde la activación de los pares que están relacionados con tal *frame*, *registrar/representar* o *afectar/caracterizar* en este caso, hasta la incorporación y activación de información relacionada con ellos, por ejemplo, información sinonímica.

Por otro lado, además de la plantilla de información que codifica un *frame* (ver Figura 3), es necesario jerarquizar y organizar los *frames* en la BCL. En este sentido, los 9 *frames* que se desprenden del análisis pueden organizarse a manera de una red que agilice y optimice la información que procesará el SRI.

En la Figura 4 mostramos, con base en la ontología Euro WordNet<sup>14</sup>, cómo estos 9 *frames* pueden relacionarse, organizarse y jerarquizarse con el fin de obtener un mejor provecho de la información que contendrá la BCL.

---

<sup>14</sup> Cfr. la interfaz 0.1 en: <http://nipadio.lsi.upc.es/cgi-bin/public/wei1.consult.perl>

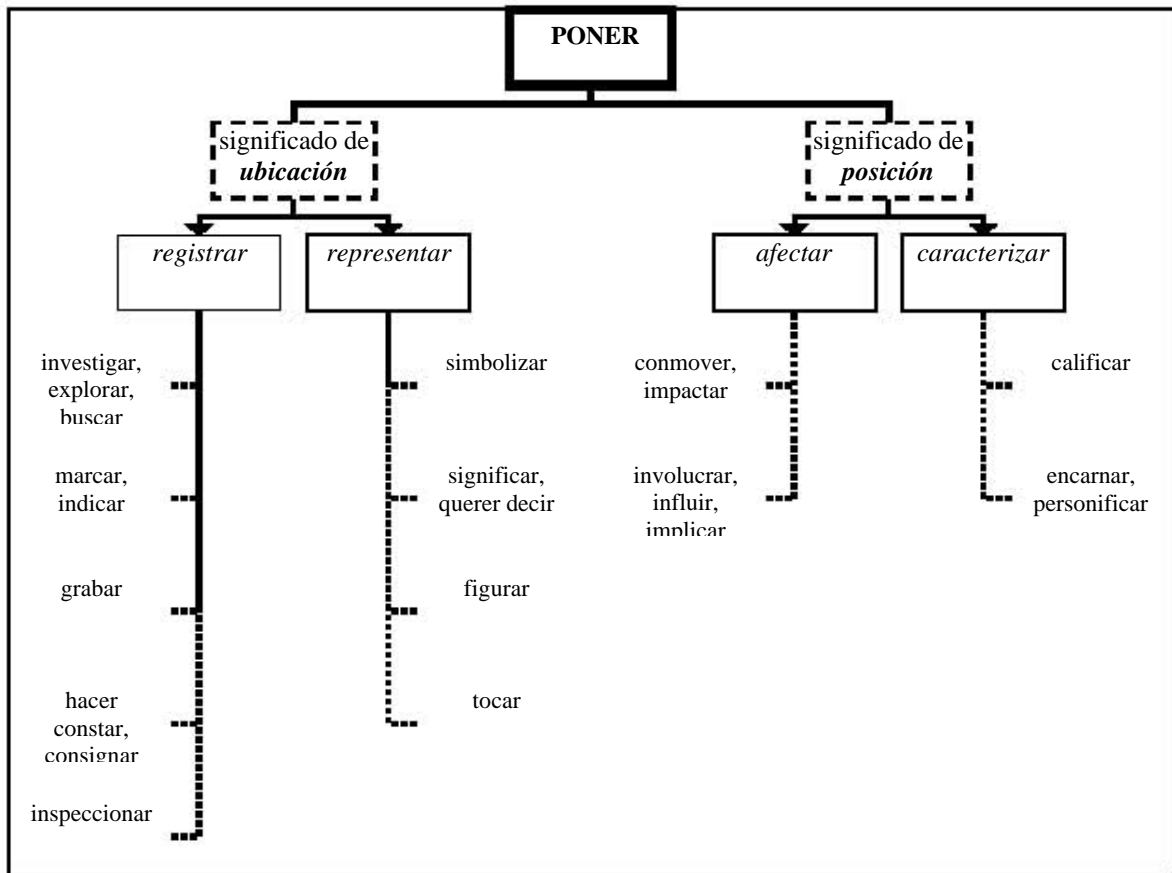


Figura 3: Plantilla del *frame* PONER

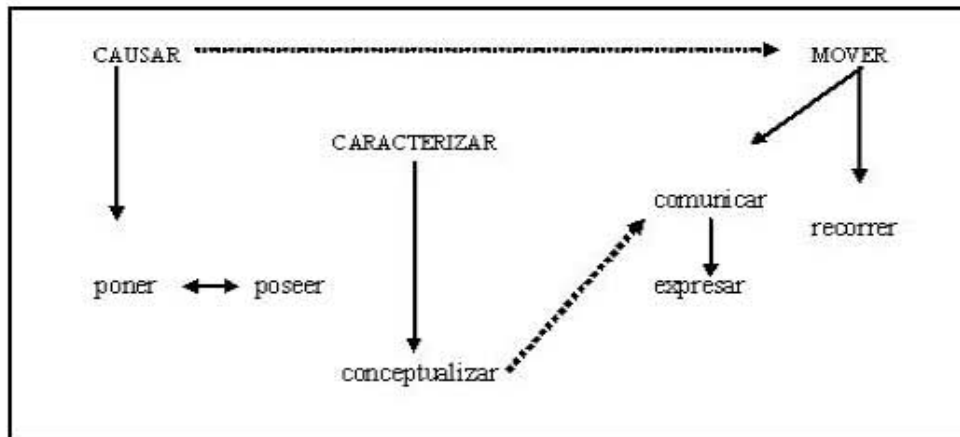


Figura 4: Organización de los *frames* del CA

De acuerdo con las relaciones léxicas de Euro WordNet, los 9 *frames* que propusimos pueden organizarse a partir de relaciones léxicas de hiperonimia, hiponimia, sinonimia, antonimia, etc., como lo muestra la Figura 4. De esta forma, con base en una organización del léxico por medio de patrones psicolingüísticos (ver Fallbaum 1998), la

BCL podrá organizar el léxico semejando la estructura lingüística que tenemos los humanos, resultando con ello un SRI, cuya organización del léxico y cuyo funcionamiento, estén más apegados a la realidad lingüística de hablantes y oyentes.

### 3.5 Conclusión del análisis

Una vez establecida la argumentación que explica el porqué de la aparición indistinta de los elementos del CA en las definiciones de 1) a 28) y, esquematizado el significado que, en función de las construcciones, debería ser codificado para interpretar de forma más rápida y segura el término que describe cada definición, detallamos las conclusiones que hemos extraído de esta etapa.

La interpretación que hemos seguido para argumentar la problemática planteada en este capítulo pretende reflejar, hasta cierto nivel, los procesos mentales que subyacen a la codificación de un significado. Para ello, hemos propuesto un conjunto de *frames* con el que pretendemos representar determinadas categorías que sean capaces de sintetizar una estructura conceptual.

Es cierto, como puntualiza Cristóbal (2001), que los *frames* que sugerimos no describen la información gramatical, sintáctica o semántica que ayudaría a un mejor entendimiento de la lengua y, como consecuencia de esta falta de información, estaríamos discriminando los matices con que cada verbo que integra el *frame* puntualiza el significado. Sin embargo, nuestra idea ha sido la de representar categorías en función de conceptos abstractos y no categorías en función de cosas o variables. En este sentido, adoptamos la postura de Evans y Green (2005) que señalan que la función de la lengua es representar o simbolizar una estructura conceptual, no la de describir cosas.

Bajo esta perspectiva, todo *frame* está constituido por una estructura de conocimiento que existe independientemente de su interacción con el léxico que está presente en una construcción (Fillmore 1982). La construcción permite la integración de las

estructuras conceptuales, establecidas por el *frame*, en un patrón movable que facilita la codificación del significado.

Lo anterior supone que la lengua contiene información sobre la realidad, y que esta información permite establecer relaciones conceptuales entre palabras, agrupándolas en escenarios o marcos determinados, asociados a una expresión lingüística que depende de un contexto, el cual, de acuerdo con Langacker (1987), representa una entidad cognoscitiva que describimos mediante un *frame*. Esto es, en palabras de Nerlich y Clarke (2000): “no word [o concepto] exists in isolation”.

Ahora bien, si nuestra interpretación ha sido acertada, concluimos que a través de esta perspectiva podemos explicar los posibles eventos que deberían ser codificados para procesar, de forma correcta, la información que transmiten, a partir de una construcción, los verbos del CA. Sin embargo, como más arriba señalamos, a pesar de que con una estructura de *frames* se opacan y eliminan los matices y las peculiaridades que cada verbo aporta a la definición, para el fin de proponer una estructura que pueda ser implementada en una BCL e interpretada por el motor de búsquedas de un SRI, tal estructura representa más un beneficio que un daño.

No obstante este fin, es indudable que no basta con este nivel de análisis. Es obvio que si tratamos de ver este fenómeno desde una perspectiva distinta, por ejemplo, desde la comunicación, nos encontraremos con otra serie de fenómenos que no habíamos considerado. Por lo anterior, se vuelve casi imprescindible un acercamiento desde la mayoría de los enfoques posibles. Por ello, dentro del marco de este trabajo, tratamos de acercarnos a esta problemática desde un enfoque que involucre el análisis de corte pragmático. En el siguiente apartado resumimos en qué consiste este acercamiento.

### **3.6 La pragmática**

Dado nuestro interés por dar pautas para que los sistemas informáticos puedan, en la medida de lo posible, representar la lengua tal y como la usan hablantes y oyentes, es preciso tomar en cuenta una gama considerable de análisis y enfoques. Es indudable que la

lengua no sólo se resuelve en un plano semántico o sintáctico, ni siquiera a partir de la suma de éstos. Una respuesta semántica a un problema muy específico como el de los pares del CA es sólo una parte del análisis que se debe hacer para diseñar la BCL de nuestro SRI.

El factor pragmático, considerado en este trabajo como parte esencial de la forma en que los usuarios interactuarían con el sistema, es un punto a cuestionar y a desarrollar para poder crear mejores sistemas<sup>15</sup>. Por tanto, en el capítulo siguiente nos enfocamos a:

- i. señalar la importancia y pertinencia de un análisis de corte pragmático que permita comprender mejor los fenómenos que suceden en la comunicación;
- ii. detallar la problemática que representa la comunicación humana para su tratamiento en sistemas informáticos;
- iii. describir, con ejemplos del CA, algunos de los fenómenos que son de interés para la pragmática y para nuestra investigación.

---

<sup>15</sup> Un ejemplo de esto es que los usuarios, al momento de introducir su propia descripción del término, estarán dando sus valoraciones y percepciones de dicho término. Al hacerlo, focalizarán aspectos, características y cualidades que, si bien son generales al término, variarán en peso y relevancia de acuerdo con cada usuario.

## **EL FACTOR PRAGMÁTICO**

---

A lo largo de este capítulo nos aproximamos a la problemática que encierra la comunicación humana más allá de la gramática que rige a una lengua. Para ello, recurrimos a la pragmática. En primer lugar señalamos qué entendemos por pragmática y cómo ésta puede repercutir a la hora de describir, de forma más abarcadora, la lengua a través de sus procesos comunicativos.

En segundo lugar, nos centramos en cuestionar si es necesario un estudio pragmático para constituir la BCL de un SRI. ¿Qué ventajas conllevaría un análisis de este tipo? ¿Cómo deberíamos entender la comunicación para poder semejar sus procesos a la hora en que un humano interactúe con una computadora? ¿Podríamos, en todo caso, decir que esta interacción puede llamarse comunicación?

Asimismo describimos, con ejemplos del CA, la problemática que entraña establecer patrones que describan los procesos comunicativos, así como el riesgo que representa un modelado computacional de un proceso tan abstracto como el de la comunicación.

Por último, concluimos presentando el esquema de nuestra propuesta para incorporar el estudio de corte pragmático en la arquitectura de la BCL que formará parte del diccionario de búsquedas onomasiológicas.

### **4.1 Más allá de la gramática**

La lengua como sistema dinámico, es decir, como un ente en constante interacción entre las reglas de su propia gramática y el uso que los hablantes hacen de ella, no puede ser caracterizada sólo bajo términos gramaticales. En el análisis y estudio de ésta deben ser consideradas las partes de mayor vitalidad: hablante y oyente. Cada uno de ellos, con el uso

particular y general que hacen de la lengua, crea y recrea un sistema lingüístico que no es perfecto y que está en constante movimiento y cambio.

Aquí, en el uso real que hablante y oyente hacen de la lengua, en donde la actuación, de que habla Chomsky, tiene más peso que la competencia, es donde entra el estudio pragmático. Ahí mismo es donde el significado rebasa la esfera semántica de condiciones de verdad para instalarse en un espacio en el que el comportamiento lingüístico no es homogéneo y la codificación y decodificación de lo que se comunica se transforma en procesos mucho más complejos que las discusiones sobre si se puede o no decir, sin violar la gramaticalidad de la lengua.

Lejos de las reglas y de los oníricos sistemas lingüísticos, cerrados y perfectos, está la parte más viva y creativa del sistema: la lengua de hablantes y oyentes. La pragmática debe dar cuenta de ésta y explicar cómo el sistema lingüístico ha sido rebasado por los procesos vivos de la comunicación.

#### **4.2 La pragmática y el papel del oyente**

Estamos todos de acuerdo en que el sistema lingüístico se regula a partir de sus propias reglas, lo cual nos permite tener un sistema que no sea caótico y que asegure la continuidad de la comunicación; sin embargo, ¿podemos estar igual de seguros que estas mismas reglas morfofonológicas, sintácticas y semánticas sean los ejes inviolables para realizar nuestros actos comunicativos?, o más aún, ¿establecer que aquello que dicta una gramática sea el medio, único e infalible, por el cual regimos nuestros actos comunicativos asegurando la codificación de éstos y, en consecuencia, su inmediata decodificación?

Nuestra respuesta es no. Y no sucede porque la lengua en uso rompe, un gran número de veces, con su significado lingüístico. Esto es, como señala Escandell-Vidal (2004), la lengua no puede ser caracterizada sólo bajo los términos de su gramática.

Es un hecho que no nos comunicamos por medio de oraciones perfectamente gramaticales, de hecho, muchas veces ni siquiera con oraciones. Nuestra comunicación diaria es un continuo que a cada momento se hace y se rehace a partir de inferencias, supuestos, creencias, implicaturas, reanálisis, etc., que constituyen un acto comunicativo mucho más complejo, además de interactivo, en donde el oyente juega un papel de gran importancia que aquél que aprendemos de niños y que pretende explicar la comunicación como un proceso, hasta cierto punto burdo, en donde un emisor transmite un mensaje a un pasivo receptor, quien lo codifica sin mayor problema, y lo devuelve perfectamente decodificado.

Si esto fuera así de sencillo, como lo muestran esos viejos esquemas, ¿por qué sucede, en ocasiones, que cada uno de los protagonistas del acto de habla está comunicando situaciones distintas y cree que su contraparte está codificando y decodificando lo que se le transmite?

Esto es algo que va más allá de nuestra competencia como hablantes. Tiene que ver más, según nuestra opinión, con nuestro papel como oyentes. ¿Hasta qué grado lo que pretendemos transmitir puede variar según la interpretación del oyente? ¿Hasta qué punto lo que decimos de forma explícita es suficiente para lograr una comunicación? ¿Qué hay con aquello que transmitimos implícitamente?

Lo que codificamos y decodificamos no está en función de la realidad que se emite sino de aquélla que se percibe. En este sentido, podemos decir que en los procesos de comunicación el oyente pasa de ser un simple paciente a un agente del acto comunicativo, a la misma altura y jerarquía que el hablante.

En 29) ejemplificamos lo que queremos mostrar.

29) La burra de mi prima me molesta demasiado.

A falta de un contexto mayor, la decodificación de 29) podríamos solucionarla, hasta cierto punto, de forma satisfactoria y desde el plano del hablante, si nos sujetamos a



un análisis sintáctico. Sin embargo, si nos situamos en el plano del oyente, podríamos cuestionarnos qué es lo que se quiso decir, cómo se dijo, qué se puede implicar y qué se puede inferir para decodificar esa construcción y poder emitir 30).

30) Pues sácala de tu casa.

Al igual que 29), podríamos decir que 30) es lo suficientemente ambigua como para tomarnos la molestia de resolverla de otra manera que no fuera desde la sintaxis y la semántica; no obstante, esto es dejar de lado el papel de la comunicación, que atañe a la pragmática, y centrarnos en el papel de las reglas de la gramática.

Y, en este punto, cabe preguntarnos qué perderíamos si optamos por suscribirnos sólo a la gramática y no tratamos de estudiar más allá de ésta, que es donde entra el papel de la pragmática. Si no atendemos a la pragmática como disciplina autónoma y seria, dejaríamos de estudiar y ver fenómenos, como los que señala Curcó (2004) cuando define lo que es el estudio serio y espinoso de la pragmática:

A partir del reconocimiento de la distancia entre el significado lingüístico y el comunicado y del papel de la inferencia en la comunicación, se concibe a la pragmática como una capacidad mental, un sistema de procesamiento de información diseñado para interpretar un fenómeno del mundo real: la conducta comunicativa. Vista de este modo, la pragmática constituye un dominio propio, no es un apéndice de la semántica ni una parcela de la filosofía del lenguaje. Tiene un objeto de estudio bien delimitado –los procesos inferenciales que permiten construir una interpretación, vistos como procesos mentales reales– y su metodología y fuentes de evidencia son diversas. Se trata una disciplina teórica, pero con fundamentos y fuentes de evidencia empíricas. Busca establecer respuestas que sean psicológicamente plausibles y sus criterios de validez no se limitan a los del análisis filosófico.

La pragmática, como disciplina que se encarga del estudio formal del sistema lingüístico en uso y de los procesos de inferencia y decodificación de la información explícita e implícita que integran la comunicación, constituye uno de los retos más importantes para la lingüística computacional. De aquí que en los siguientes apartados detallemos en qué consiste este reto, ejemplificándolo con la problemática que representa la construcción de la BCL del SRI.

### 4.3 La lengua en uso

Toda palabra comunica una gran variedad de significados: la decodificación de lo que se transmite con una palabra no sólo depende del significado léxico que cada palabra tiene, ni siquiera de la suma de significados de todas las palabras que aparecen junto a ella en una construcción. La decodificación de lo que pretendemos transmitir a nuestro interlocutor, sea éste un ente humano o, en nuestro caso, una máquina, va más allá de interpretar significados léxicos o significados en contexto. Nuestra comunicación, como antes apuntamos, es una serie de procesos tan complejos como para resumirlos con un simple esquema, igual al que aprendíamos en nuestros textos de educación primaria, que constreñía la comunicación a un emisor, un mensaje y un receptor.

Este esquema se basa en la idea de que a toda palabra le corresponde un sólo significado y un sólo significante que jamás cambia. Ahora, lo que sucede con este esquema es que está dejando fuera la comunicación. Esto es un grave problema, ya diversos autores han señalado que esto no funciona así. Luhmann (1990) apunta que las palabras, aún en su sentido léxico, “do not carry a meaning”.

Vistas desde el plano comunicativo, las palabras, tal y como señala Curcó (2004), no abarcan ni agotan los significados que se transmiten en la comunicación. En el uso, éstas adquieren un comportamiento distinto al que les corresponde según el sistema lingüístico. Tal comportamiento provoca una *evolución* del significado, el cual, siguiendo al mismo Luhmann (1990), “[is] perceived on the basis of the perceiver’s background experience”.

Este punto de vista que integra la comunicación en el estudio de la lengua entraña un gran problema, al menos, para nuestra investigación. Citamos como ejemplo las ideas de Luhmann. Si las palabras transmiten significados que en algunas ocasiones no están vinculados con el significado que léxicamente les corresponde<sup>1</sup>, ¿cómo es que podemos

---

<sup>1</sup> Claro está que el significado no puede variar arbitraria ni ilimitadamente. Si esto ocurriera, la comunicación no existiría. La comunicación es un proceso social, por tanto, no es posible variar el significado de manera drástica y a antojo. El significado varía y cambia siempre en un contexto social, por tanto, dentro de cierto margen, el cambio de significados o los nuevos significados que adquieren las palabras deben estar sujetos a un entorno social que le dé sentido.

decodificar lo que se nos transmite si nuestro fondo de experiencia, como indica Luhmann, llegara a ser distinto del de nuestro interlocutor?

No vamos a responder ahora esta pregunta. Lo que nos interesa rescatar es lo siguiente: si nosotros que contamos con un fondo de conocimiento y experiencia que, suscrito a un marco social, nos permite interpretar, inferir, reanalizar, implicar, etc., aquello que se nos transmite y, que de cierta forma, asegura que nos comuniquemos, entonces cómo hacer para que parte de estos procesos puedan ser simulados, con base en la información de la BCL, generando como resultado un acto en el que el SRI devuelva, a modo de respuesta comunicativa, el término que busca un usuario. Éste es el reto, a nivel de análisis pragmático, que describimos en el siguiente apartado.

#### **4.4 El reto del estudio pragmático en nuestro proyecto**

Hoy en día, implementar modelos de lengua que se apeguen a lo que es un sistema lingüístico en uso y no sólo a su descripción gramatical, es una tarea necesaria e inherente al avance tecnológico. La comunicación, entendida como una gama de procesos activos con referentes intrínsecos y extrínsecos, es parte fundamental para describir una lengua. Por tal motivo, si con análisis exhaustivos que abarquen todos los niveles de análisis lingüístico, incluyendo el pragmático, podemos extraer patrones del comportamiento real de una lengua, creemos que se puede hacer extensivo el resultado hacia sistemas informáticos con el fin de mejorar el desempeño de éstos en su interacción con los usuarios.

El reto que esto supone es cómo representar el conocimiento, que va más allá de sólo información gramatical, en la BCL de nuestro SRI para poder aproximarnos, en lo posible, a los procesos que se desarrollan en la comunicación humana<sup>2</sup>. Esto es, desde la perspectiva del diccionario onomasiológico, cómo semejar un proceso de comunicación en el que la computadora desempeñaría un papel de vital importancia, el del oyente<sup>3</sup>.

---

<sup>2</sup> En el apartado 4.7 describimos cómo abordamos este reto y cómo, en el marco de nuestra investigación, lo pretendemos resolver.

<sup>3</sup> De hecho, este rol de oyente está en función del proceso que el sistema debería hacer para devolver al usuario el término que éste *comunica* por medio de su propia descripción.

## 4.5 La interacción humano – computadora como comunicación

Si suponemos que el rol principal que ejecutará la computadora será, desde una perspectiva comunicativa, el de oyente, debemos primero cuestionarnos si podríamos llamar comunicación a la interacción entre un humano y una computadora. Cuestión que en sí no es trivial y que nos remite a preguntarnos si la interacción que una madre tiene con su bebé de nueve meses puede ser considerada comunicación o si bien, la interacción entre un hombre y su mascota puede ser vista como un proceso, mínimo si se quiere, de comunicación.

Ahora, volviendo a nuestro cuestionamiento inicial, defendemos que el proceso de devolver al usuario el término que busca a partir de la descripción que hace del mismo, supone, de alguna manera, que pretendemos que el diccionario, entendido éste como el conjunto de módulos que conforman su arquitectura<sup>4</sup>, simule un proceso de comunicación. Pero, ¿podríamos llamarle comunicación?

En efecto, si nos situamos en el plano en el que un usuario introducirá un concepto en la interfaz de un sistema y, sobre todo, en el proceso que el sistema, al igual que un oyente, debe realizar para codificar el concepto del usuario y devolverle, a modo de respuesta, un término, creemos que, aunque de manera elemental o primitiva, estamos hablando de comunicación. Por tanto, nuestra tarea es dar argumentos que permitan formalizar parte de los elementos que tienen que ver con los procesos comunicativos para que el diccionario pueda simularlos y se *comunique* con el usuario de una manera más cercana al entorno comunicativo de los humanos.

### 4.5.1 Problema

El hecho de concebir que la interacción humano – computadora<sup>5</sup> entraña un proceso de comunicación representa, como arriba señalamos, todo un problema para el trabajo que se realiza en lingüística computacional.

---

<sup>4</sup> Ver Figura 1 en el Capítulo 1.

<sup>5</sup> En específico, la interacción que se dará con el diccionario onomasiológico y sus usuarios.

Para ejemplificar esto, partamos de la premisa de que nosotros como hablantes y oyentes no limitamos el conocimiento y el uso que hacemos de las palabras sólo a sus aspectos lingüísticos, al contrario, junto con ellos manipulamos datos exteriores, datos extralingüísticos, como el contexto, la información de mundo, la formación social y cultural, etc., que determinan el uso real y particular que damos a las palabras (Sager 1990), luego entonces, cómo podríamos concebir que una máquina pudiera contar con esta clase de información que a su vez le permitiera realizar procesos de inferencia, reanálisis, interpretación, etc., para establecer una *comunicación* con un humano.

En este sentido, el diccionario electrónico de búsquedas onomasiológicas presupone que la BCL debe proporcionar esta información al motor de búsquedas para que éste realice los procesos de inferencia, reanálisis, interpretación, etc., que permitan devolver al usuario el término que busca<sup>6</sup>.

#### 4.5.2 Los hechos

Ahora bien, el punto que se desprende de este problema es cómo extraer y representar esta información en la BCL, si tanto la información como el léxico con el cual ésta se transmite es tan cambiante y variado como la cantidad de posibles usuarios del sistema. Como ejemplo de ello, baste echar una mirada a las definiciones en las que aparecen los verbos del CA.

El léxico que los expertos utilizan para definir un término, visto desde la categoría que nos ha ocupado durante la investigación, es, en algunos casos, común en las definiciones; sin embargo, hay una gama considerable de unidades que divergen y que, sobre todo, focalizan aspectos distintos y relevantes de aquello que comunican y que varían de acuerdo con el hablante. Véanse las construcciones de 31) para ejemplificar lo anterior.

##### 31) *predicado verbal*

- d) Predicado que se **desempeña** por un verbo
- e) Predicado que se **expresa** por un verbo

---

<sup>6</sup> Véanse los apartados 1.2 y 1.2.1 en el Capítulo 1.

Tanto la construcción de 31a) como la de 31b) pretenden transmitir un mismo significado, la diferencia estriba en que, a partir de los elementos *desempeñar* y *expresar*, se focaliza un tipo específico de información que el hablante perfila para que el oyente decodifique el término en cuestión. Claro está que la presencia de uno u otro verbo puede explicarse mediante los *frames* que hemos sugerido; sin embargo, los matices de información que entraña cada verbo, y que los hablantes perfilan para describir un constructo de información específica (Ghidini y Giunchiglia 2001), quedan eliminados por una estructura tal que los coloca al mismo nivel.

Ahora contrástelese con los ejemplos de 32) y 33).

32) *sujeto*

- a) Elemento o sintagma nominal que **recibe** la acción expresada por el verbo
- b) Posición sintáctica del actante que **realiza** la acción expresada por el verbo

33) *velocidad*

- a) Relación entre el espacio recorrido por un móvil y el tiempo **empleado** en recorrerlo
- b) Relación entre la cantidad de movimiento, de trabajo etc., realizados y el tiempo **invertido** en realizarlos

Si en 31) la diferencia de información que está en foco está en relación con un par de verbos, en los ejemplos de 32) y 33) la información que se perfila está en función de rasgos que, a distinto nivel y desde distintos ángulos, son relevantes para definir el término.

Quien pronuncia 32a) y no 32b),  *Cree* que está proporcionando información que debe asegurar, en quien la recibe, la conceptualización más fácil del término, vía la relevancia de ésta y el menor esfuerzo que representa para procesarla (Sperber y Wilson 1995). El mismo comportamiento se puede extender a las construcciones de 33) y, en un nivel superior, a la comunicación en sí misma. Cada participante de un acto comunicativo perfila su realidad y la proyecta, mediante el léxico que elige, en su contraparte comunicativa.

Bajo esta premisa y resumiendo lo expuesto hasta ahora, lo que pretendemos destacar con estos ejemplos es que para entender una construcción no basta con codificar su estructura sintáctico-semántica (Goldberg 1997) sino en asimilar la información, que no depende de la gramática pero que proporciona una visión parcial y, sobre todo, particular del mundo (Ghidini y Giunchiglia 2001) que nos permita decodificar lo que un hablante transmite a un oyente o, en nuestro caso, a una máquina.

Una vez señalado el problema y ejemplificado con hechos, señalamos que lo que nos interesa es una pragmática que describa la lengua, no que la prescriba<sup>7</sup>, de manera que podamos realizar análisis que, en lo posible, nos provean de los conocimientos suficientes para recrear la comunicación a partir de la información que se estructurará en la BCL de nuestro SRI. Para ello, recurrimos a la *Teoría de la Relevancia* (TR) y, en específico, al concepto que a ésta subyace.

#### **4.6 Teoría de la Relevancia**

La hipótesis principal que subyace a la teoría desarrollada por Sperber y Wilson (1995) es que la cognición humana siempre se encamina hacia la búsqueda de *relevancia*. Todo proceso de comunicación puede ser analizable y decodificable si asumimos la premisa de que nuestros actos comunicativos se rigen por un patrón cognoscitivo, general a todos los humanos, a través del cual la manera en que construimos y reconstruimos nuestra comunicación, sea explícita o implícita, está en función de ser relevantes.

Bajo esta perspectiva, Sperber y Wilson (SW) definen relevancia como un proceso inferencial que, además de involucrar módulos lingüísticos, implica módulos cognoscitivos tales como la memoria, los pensamientos, el conocimiento particular y comunitario, etc., y un contexto que, en conjunto, orientan la comunicación hacia la búsqueda de efectos que produzcan relevancia (Wilson y Sperber 2004, Sperber 2005). Asimismo, SW señalan que toda información que pueda ser procesada debe cumplir con los siguientes postulados para poder generar relevancia:

---

<sup>7</sup> En este sentido, hacemos referencia a los estudios que defienden una descripción de la lengua mediante condiciones, reglas o principios que rara vez son aplicables y verificables en la lengua en uso real.

- I. a mayor número de efectos cognoscitivos, mayor relevancia;
- II. a menor esfuerzo de procesamiento, mayor relevancia.

De igual modo, argumentan que la relevancia no se da independientemente de un contexto, al contrario, hablan de *relevancia en contexto*. Indican que la relevancia está en función de un contexto apropiado. No todos los contextos son adecuados para decodificar lo que se nos comunica, ni nos comunicamos bajo un único contexto. En TR se distingue un contexto que se construye y reconstruye activamente, permitiendo así la generación de mayores efectos cognoscitivos por el menor esfuerzo de procesamiento.

Asimismo, SW apuntan que la relevancia no se da de manera aleatoria sino bajo un principio que rige la comunicación: todo estímulo emitido, señalan, crea la expectativa de su propia relevancia óptima. A este principio lo denominaron *principio comunicativo de relevancia*. Según este principio, tanto oyente como hablante, al comunicarse, están en constante búsqueda de una *relevancia óptima*. En este sentido, un estímulo ostensivo será óptimamente relevante si:

- a) es lo suficientemente relevante para ameritar su procesamiento;
- b) es el más relevante de acuerdo con las capacidades y preferencias de quien emite tal estímulo.

Para aclarar un poco a qué se refieren SW, señalamos como ejemplo un caso del español de la ciudad de México.

- A) Sigue hacia el norte dos calles más y luego das vuelta hacia el noroeste.
- B) Sigue derecho dos calles más y luego das vuelta hacia tu mano izquierda.

Situándonos en un contexto en el que tratamos de averiguar cómo llegar a cierto lugar, vemos que el enunciado de A) da información que, para ser relevante, según el concepto de SW, necesita un costo de procesamiento alto, mientras que el de B) da la misma información sólo que con otro tipo de formas. Lo que obtenemos con A) y B) son efectos cognoscitivos, tal vez igual de importantes, sólo que con la diferencia de que en B), el costo de procesamiento es mucho menor. Es mucho más fácil de decodificar cuál es el camino a seguir si nos situamos espacialmente a partir de nuestro cuerpo, por ejemplo, la



mano izquierda, que decodificarlo en términos de puntos cardinales como el noroeste. Aunado a esto y de acuerdo con la noción de relevancia óptima, esperamos como respuesta un estímulo lo suficientemente relevante para procesarlo, confiando en que nuestra contraparte cooperará devolviéndonos su mejor estímulo; en este sentido, B) ofrece un mejor estímulo puesto que aporta la información solicitada además de que quien lo emite está cooperando al devolver un estímulo óptimamente relevante. Por tanto, los efectos cognoscitivos que produce B) así como el bajo costo de procesamiento que representa más el nivel de relevancia que provoca, hacen que dentro de este contexto, B) sea más relevante que A).

#### **4.7 Nuestra propuesta**

Tal y como hemos mencionado a lo largo de la investigación, optar por una estructura de información que no contemple un plano pragmático significa dejar fuera información muy valiosa que nos puede ayudar a realizar mejores búsquedas y con mayor profundidad. Esto es, si la información de la BCL sólo estuviese integrada por una estructura de *frames*, llegaríamos sólo a un nivel de búsqueda: recuperar un término a partir de un léxico más o menos vasto. En cambio, si incorporáramos la información relevante que los hablantes están perfilando, la cual se refleja mediante el léxico que emplean<sup>8</sup>, podríamos reforzar los patrones de búsqueda y dar pautas para que, además de recuperar un término, el motor de inferencias pueda profundizar en los niveles de búsqueda y ofrecer información adyacente o nueva, según las necesidades del usuario y el tipo de búsquedas (Reyes y Solares 2004), robusteciendo así el SRI y volviéndolo más interactivo.

Para ejemplificar lo anterior baste hacer referencia a los buscadores de Internet. La mayoría de estos buscadores lo único que hacen es recuperar, para un búsqueda más o menos compleja, infinidad de documentos que contienen las palabras empleadas en la búsqueda pero con información pocas veces relevante.

---

<sup>8</sup> En este caso sólo hacemos referencia al léxico utilizado puesto que, a diferencia de la comunicación oral, en la escrita no podemos recuperar, a no ser por patrones tipográficos, información desprendida del tono, gestos, lapsus, etc.

Ahora bien, nuestra propuesta consiste en que, al menos como un primer acercamiento, los pares del CA, además de integrar un *frame*, puedan ser descriptores de información. Esto es, que engloben información más allá del léxico, es decir, información pragmática que nos permita proponer, a partir de una búsqueda más profunda, información nueva o complementaria.

En los siguientes apartados explicamos cómo es que pretendemos, a partir de TR, integrar esta estructura de descriptores.

#### 4.7.1 Un modelo de descriptores a partir de TR

Para que la BCL del SRI pueda semejar un proceso de comunicación, nos es preciso sustraer este concepto de relevancia que permea la teoría de SW. Para el diseño del SRI de corte onomasiológico que proyectamos, es de suma importancia tomar en cuenta qué es lo que perfila el usuario, más allá del léxico, para definir el término. En este sentido, el concepto de relevancia nos servirá para analizar el comportamiento del léxico de un campo particular. La identificación de las formas que mayor relevancia puedan tener en un discurso especializado nos permitirá proponer descriptores del dominio.

De esta forma, el modelo que proponemos se construirá a partir de una red de descriptores. Éstos aglomerarán información lingüística y extralingüística que permitirá interactuar con el sistema de una manera más apegada a lo que sería la comunicación. Así, el sistema tendrá la capacidad de inferir el término, a partir de la descripción tan vaga o concisa, que el usuario formule.

#### 4.7.2 El *descriptor*

Definimos *descriptor*<sup>9</sup> a aquella forma que, como parte de un dominio específico, puede rebasar su significado léxico y adoptar, según el uso que tiene en el discurso especializado y en el habla coloquial, significados y usos más complejos capaces de

---

<sup>9</sup> Sobre el tema de descriptores y la forma en cómo se ha adoptado una teoría pragmática para el análisis de la representación del conocimiento, consultar Reyes y Solares (2004).

representar el conocimiento del área. En cierto sentido, un descriptor es una extensión de lo que Fillmore (1982) llama *frame* y que nosotros hemos adoptado en la investigación desde el punto de vista semántico.

Con un descriptor pretendemos crear nodos para que otros significados que en primera instancia no se vincularían directa y, en consecuencia, coherentemente con el discurso, puedan ser activados. Esto nos permitiría traer información más profunda y relevante en función de que el descriptor, como apunta Rieger (1988), acumulará información que podrá disparar porciones de conocimiento que, de otra forma, no podrían activarse. Esto quiere decir que con un descriptor, a diferencia de un *frame*, pretendemos reflejar la información que es propia del discurso y no de la gramática, y que, una vez activado, ayude al sistema a codificar y decodificar la información que el usuario perfila en su descripción del término.

La Figura 5 representa, a grandes rasgos, la estructura que tendría un descriptor<sup>10</sup>. Una forma, en este caso *lingüística*, puede comprimir información que no es estrictamente aquella que encontramos en un diccionario. Creando nodos específicos de enlace se podría recuperar información que rebase los alcances de la simple forma *lingüística*, yendo desde los nodos elementales como el (a) y el (b) hasta nodos como el (n), (x) o (y) en los que se traspasan los límites, digamos léxicos, de la forma, y se profundice hacia puntos en los que el descriptor operaría.

---

<sup>10</sup> Es preciso mencionar que de aquí en adelante nuestros ejemplos estarán vinculados al plano nominal, que es la categoría sobre la cual hemos trabajado con mayor profundidad y que permite un mejor entendimiento de lo que queremos mostrar con el concepto de *descriptor*; no obstante, haremos referencias a los pares del CA como el objetivo hacia el cual nos encaminamos.

## Lingüística

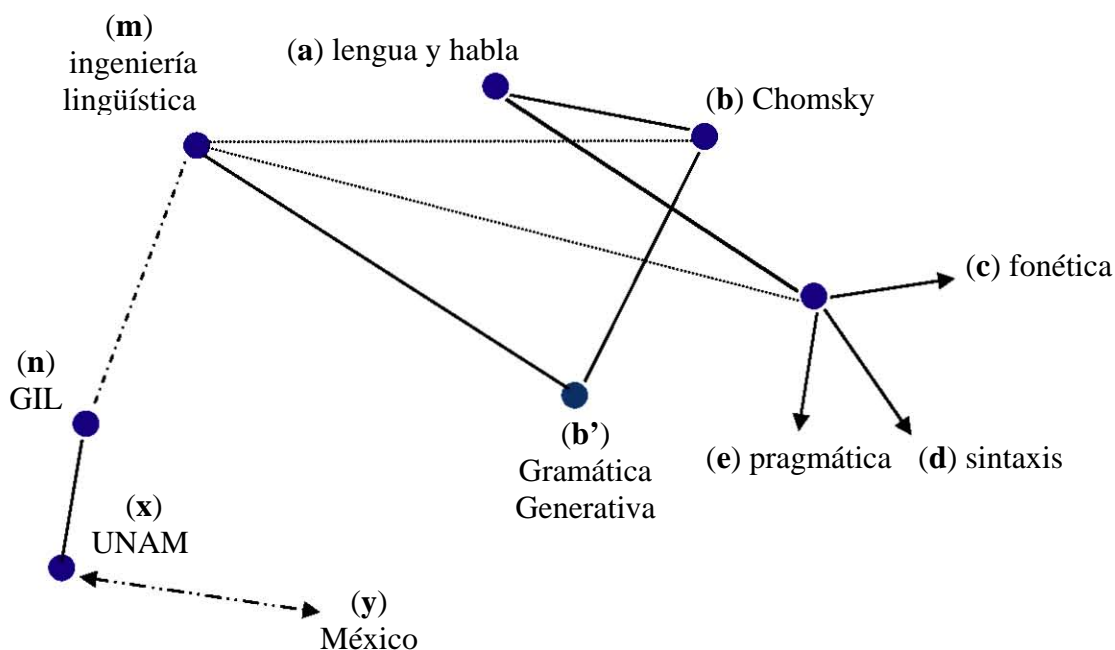


Figura 5: Ejemplo de un descriptor y de los nodos de información a los que puede llegar.

La lingüística no es sólo fonética y sintaxis, sino que se va más allá, hasta nodos en donde se encuentre información del tipo: la ingeniería lingüística, como vertiente aplicada de la lingüística, se desarrolla en México y se hace en el GIL de la UNAM.

### 4.7.3 Identificación de descriptores

Los hablantes tienen a su servicio la lengua y la forma en cómo la utilizan; no están apegados a un modelo de descripción como el aristotélico, por ejemplo. Esto significa que los usuarios del diccionario no coincidirán en la forma en que describen sus términos y, por tanto, habrá una gama considerable, no sólo en cuanto a léxico sino en cuanto a significados y características que perfilarán para los términos y que estarán en relación con quien conceptualiza<sup>11</sup>.

<sup>11</sup> Según Galinsky (1991:243), los términos son “representantes de conceptos claramente definidos”. Tal aseveración resulta cuestionable, al menos, desde el punto de vista onomasiológico. Más adelante veremos que en la práctica, con hablantes que no son expertos pero que son los potenciales usuarios del diccionario, estos conceptos que definen los términos no parecen ser tan claros como señala Galinsky.

En este sentido, y de acuerdo con Langacker (1987), presumimos que lo que hacemos cuando nos comunicamos es construir mundos conceptuales que proyectamos en nuestro interlocutor. Cada conceptualización está en función de un significado (Langacker 1991) y en relación con situaciones periféricas ancladas a una realidad particular, de modo que, aunque es un proceso caracterizado por rasgos comunes, éstos no pueden ser universales<sup>12</sup>.

Así, con base en el patrón que seguimos para comunicarnos y que SW denominan relevancia, buscamos identificar formas que dada su incidencia estadística en discursos especializados y habla coloquial<sup>13</sup> puedan representar puntos de relevancia. Ahora, el concepto que sustraemos de TR no aplica para medir la relevancia con la que nos expresamos, ya sea de forma oral o escrita, sino para dar un mayor peso a las formas que funcionarían como descriptores. Si siempre estamos en busca de mayores efectos cognoscitivos por el menor esfuerzo de procesamiento, podemos entonces tratar de apresar este concepto y así justificar la presencia de un descriptor y del conocimiento que éste representaría en la BCL.

El procedimiento para identificar los descriptores que sugerimos parte de lo cuantitativo para llegar a lo cualitativo. Un primer paso es construir un corpus con discurso especializado: información enciclopédica, de diccionarios, de artículos, etc., son material que serviría para elaborar este corpus. El segundo paso es construir un segundo corpus con información de hablantes, expertos y no expertos, sobre la misma área que el especializado. Comparar ambos corpus y extraer las formas que estadísticamente tengan mayor incidencia<sup>14</sup> es el siguiente paso. Hasta aquí el aspecto cuantitativo; la parte cualitativa consiste en analizar, escrupulosamente, bajo qué circunstancias aparecen tales formas.

---

<sup>12</sup> Estos procesos de conceptualización que reflejan una vista parcial de un mundo (Ghidini y Giunchiglia 2001) son inherentes a la construcción del diccionario onomasiológico y representan un fuerte problema puesto que partimos del hecho de que el usuario debe conceptualizar un término para poder definirlo a través de un concepto.

<sup>13</sup> Claro está que el habla común estará delimitada por un dominio específico de conocimiento.

<sup>14</sup> Las formas podrían ser, en algunos casos, las mismas puesto que hablamos de dos discursos que se expresan de manera distinta pero que, en cuanto al contenido, se desarrollan paralelamente.

¿Tienen función de tópico, de foco, de conector?, para así argumentar, pragmáticamente, el porqué tal forma podría ser etiquetada como un descriptor<sup>15</sup>.

A partir de lo anterior, podríamos extraer las formas que, según nuestra interpretación y adaptación del concepto de relevancia, pudieran funcionar como descriptores; sin embargo, no estamos seguros que sea la manera correcta. Bien podría ser que estemos equivocados y que sólo estemos recurriendo a la estadística como el recurso que nos solvete el problema de identificación de descriptores. A pesar de eso, consideramos que si guiamos el procedimiento con los principios de la TR podemos justificar de mejor modo nuestros descriptores. Por ejemplo, con el procedimiento que argüimos, cumplimos, al menos, con el punto II de la TR que señala que a menor esfuerzo de procesamiento mayor relevancia. Si la forma aparece en varios planos de manera constante, quiere decir que no representa mayor complejidad para decodificarla y, por tanto, tiene un grado alto de relevancia en el discurso. El punto I requiere de un análisis y argumentación mucho mayor. Cómo saber que la sola forma que extraemos puede representar un buen número de efectos cognoscitivos como para que sea relevante. De entrada cumple con el principio comunicativo de relevancia, puesto que toda forma emitida, ya sea en contexto especializado o coloquial, crea la expectativa de su propia relevancia. La pregunta es cómo argumentar la relevancia de esta forma y no de otra que, por ejemplo, tenga menor peso estadístico. Para ello requerimos de un análisis cualitativo más exhaustivo que con su respuesta satisfaga al cien por ciento lo que nos cuestionamos. Sabemos que esto representa un esfuerzo enorme, no obstante, debe realizarse para dar mayor solidez a las formas que sugerimos como descriptores.

Ahora bien, lo que obtendríamos con este procedimiento sería un conjunto de formas que reflejarían estadísticamente y, según el principio de TR, los puntos más relevantes que los hablantes, en discursos especializados y descripciones propias de un área, privilegian.

---

<sup>15</sup> Hasta aquí sólo señalamos que debe haber un análisis cualitativo que comprenda un estudio pragmático. No obstante, no decimos si con TR podríamos abordar dicho estudio. En este sentido, el punto no es imponer o ensalzar una teoría sobre otra, sino dejar abierta la posibilidad para abordar y examinar el fenómeno no sólo desde la pragmática sino desde cualquier nivel de análisis lingüístico.

#### 4.7.4 El contexto

Hasta aquí hemos explicado qué es un descriptor y el procedimiento para extraerlo. Con todo, hemos obviado la noción contexto. Según SW, el contexto es la chispa que hace que la información pueda ser relevante. El contexto, como antes señalamos, no es algo estático, la información debe interactuar con éste para producir efectos cognoscitivos. Argumentan que no toda la información es relevante ni todos los contextos son propicios para transformar esa información en relevante. De ahí que destaquen la noción de *relevancia en contexto*. Para ejemplificar a qué se refieren SW, veamos el ejemplo de C).

C) *fuego...*

La decodificación del enunciado de C) será más accesible y representará mayores efectos cognoscitivos en relación con el contexto bajo el que es emitido. Si nos situamos en un contexto en el que está en juego la vida, por ejemplo un incendio, la relevancia será sumamente mayor que si estamos en un contexto en el que nos preguntan si queremos fuego para encender un cigarro.

Un aspecto importante de subrayar es que, puesto que el contexto es un factor principal para decidir si algo es o no relevante y éste se construye y reconstruye activamente, es lógico pensar que si para alguien, dado un contexto propicio, la emisión de ciertas palabras fue relevante, tal vez esa misma información resulte poco relevante dicha con las mismas palabras y refiriéndose al mismo término, en otra situación. En este sentido, ¿cómo asir una concepción de contexto que pueda ser representada en la BCL del diccionario?

La respuesta que proponemos está en función con lo que señalábamos respecto a si podemos llamar comunicación a la interacción entre un humano y una computadora. Para nosotros es una comunicación elemental, por tanto, no puede ser tan compleja como la de los humanos. Pese a ello, debe simular, en lo posible, la comunicación humana y todo lo que le rodea. Es por eso, y por fines operativos, que restringimos la noción de contexto dinámico a contexto invariable.

Por tal motivo, el dominio sobre el que trabaje el SRI será nuestro contexto. A través del dominio, que siempre será el mismo, se procesará e interpretará la información.

#### 4.7.5 Relevancia óptima

Desde el plano pragmático, la identificación de descriptores y delimitación del contexto son los dos puntos principales para que la BCL cuente con la información necesaria y suficiente para poder ofrecer al usuario una respuesta que sea óptimamente relevante.

Como ya antes mencionamos, extraemos la esencia del concepto que plantean SW y lo adecuamos a nuestras necesidades. No es que la búsqueda del usuario sea poco relevante como para ameritar su procesamiento, ni que sea la más relevante según las propias habilidades del usuario; más bien, lo que señalamos es que, con base tanto en los descriptores como en la información léxica que constituirá la BCL, podremos dotar al sistema de los recursos suficientes para inferir, a partir de una búsqueda como la de 34), que el término que busca el usuario es *pene*.

34) Sirve para hacer *pipí*<sup>16</sup>.

De esta forma tenemos que para el diseño de la BCL no sólo importará que incorporemos información como la de 35) o 36) sino también, y en esto estriba la riqueza que aporta el diccionario al análisis semántico y pragmático, principalmente, aquella de 34) que es más común a los potenciales usuarios del SRI.

35) Miembro viril (DRAE 1995).

36) Órgano genital masculino que permite la cópula y que forma parte del aparato urinario; falo. (Clave 1997).

---

<sup>16</sup> Definición proporcionada por un usuario piloto del diccionario onomasiológico de sexualidad para el término *pene*, en el Segundo Seminario de Ingeniería Lingüística del Grupo de Ingeniería Lingüística de la UNAM.



#### 4.8 Posibles resultados

Como ya expusimos, la pragmática será de mucha importancia para el diseño de la BCL puesto que con la información que desprendamos de análisis de este corte y su posterior reflejo en los descriptores, podremos dotar al sistema de recursos que le permitirán, en lo posible, saber qué privilegia un usuario en su definición, qué presupone que la máquina debería saber o qué debería inferir para poder devolver un resultado relevante<sup>17</sup>. Esto significa que, dada una búsqueda como la de 37), el sistema deberá realizar procesos semejantes a los de la comunicación humana para poder devolver el término correspondiente.

37) en muchas casas se le da el nombre de *Fido*...

Dados los elementos suficientes, el sistema podrá inferir que *Fido* es un nombre común con el que los dueños llaman a sus perros y, en consecuencia, decodificar que lo que el usuario busca es el término *perro*.

La manera en que esto se daría está en función de descriptores. El descriptor debería acumular información que le permita a la BCL relacionar, a manera de hiperónimo e hipónimo, significados lingüísticos y extralingüísticos que permitan al sistema establecer un nodo entre el descriptor y la información con la que se liga. El dominio, que funciona como contexto rico pero estático, eliminará la posibilidad de ambigüedad y el resultado consistirá en, una vez realizados los procesos necesarios de interpretación, codificación, inferencia y decodificación, devolver al usuario el término que busca.

En la Figura 6 detallamos, retomando el ejemplo de 37), cómo se conectaría la información a través del descriptor.

---

<sup>17</sup> En este sentido, como apuntan Grootjen y van der Weide (2002), intentamos dotar al sistema de información que le permita, al igual que en la comunicación humana, hacer que “the participants know each other's language and somehow predict the impact of the words they use”.

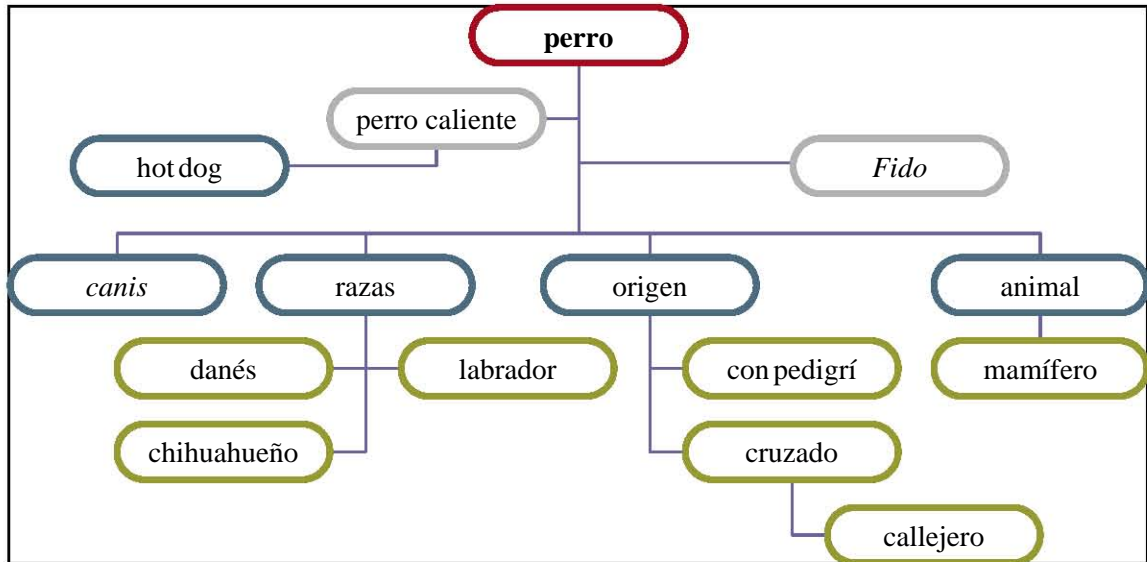


Figura 6: Ejemplo de estructuración de información lingüística y extralingüística en la BCL.

Ahora bien, retomando los pares del CA, nuestro interés es que podamos generar plantillas que puedan, siguiendo la metodología que hemos propuesto, ser los descriptores de un área específica.

Tomando en cuenta la estructura de *frames* que hemos sugerido, en la Figura 7 mostramos que uno de los verbos del CA, en este caso, *causar*, además de codificar información léxica propia del verbo y de integrarse en un conjunto más amplio de significados codificados por el *frame* de *MOVER*, podría comprimir información que sobrepasa aquella que se puede representar mediante tal *frame* y que a nivel comunicativo tiene una fuerte carga de significación, permitiendo así, búsquedas con resultados más relevantes y con mayor profundidad, tal y como se observa en la Figura 8

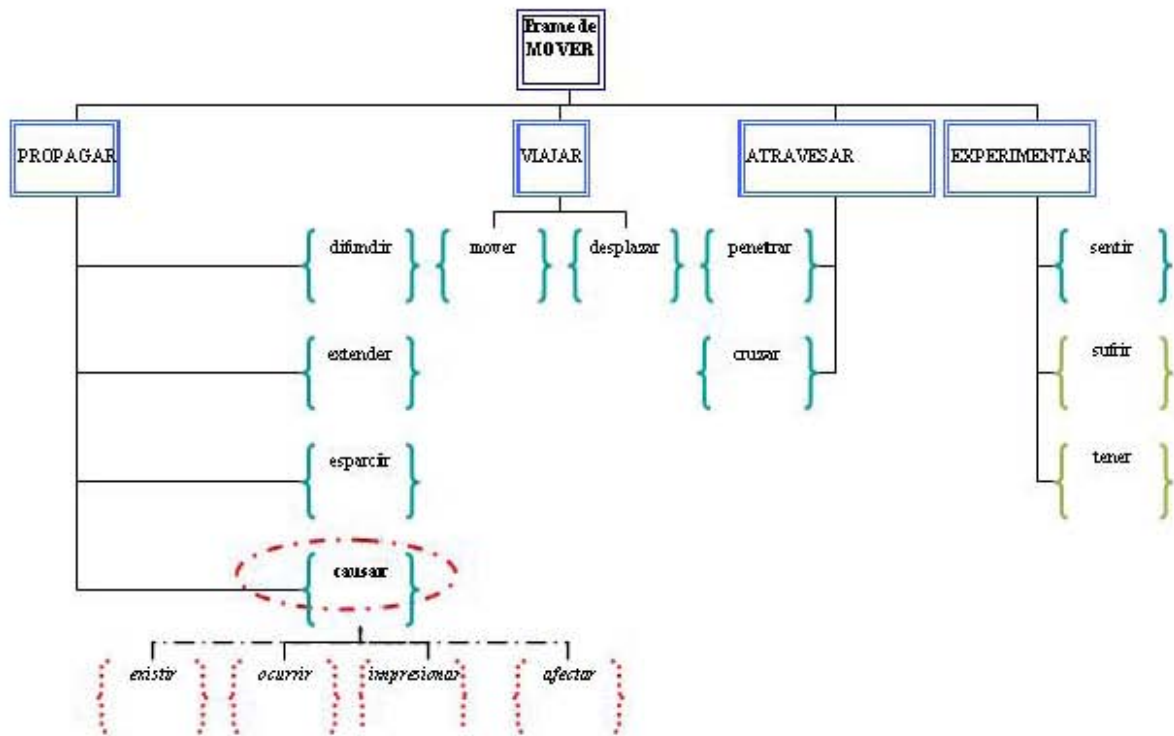


Figura 7: Esquema del descriptor *causar*



Figura 8: Búsqueda expandida con Frames y Descriptores

La estructura de *frames* y descriptores permitirá al diccionario, a partir de los elementos que el usuario introduce en su definición, en este caso retomemos de ejemplo al verbo *causar* (Figura 7), recuperar el término que busca, además de un conjunto de términos que se relacionan<sup>18</sup>, ya sea con el *frame*, con el descriptor o con ambos.

Así, mediante la estructura de la información en la BCL a partir de *frames* y descriptores, podríamos construir un SRI que, además de estar sustentado en una estructura que pretende involucrar la cognición humana a través del léxico, pueda tener un desempeño más apegado a la comunicación de los humanos, representando al usuario una mejora en cuanto a la relevancia y calidad de los resultados.

---

<sup>18</sup> En este caso, esta relación se puede observar en los *frames* que propusimos en el Capítulo 2 y en las construcciones citadas en los ejemplos de ese mismo capítulo.

## CONCLUSIONES

---

### 5.1 Conclusiones particulares

A modo de cierre de la investigación, presentamos en este módulo los puntos más importantes que podemos rescatar de cada capítulo así como las consideraciones finales de cada uno de ellos.

#### 5.1.1 Capítulo 1

En este capítulo hablamos de los sistemas de recuperación de información. Dimos especial énfasis al problema que representa el tratamiento de grandes volúmenes de información y la consecuente necesidad de desarrollar mejores sistemas que puedan manejar estos cúmulos de información textual.

Manifestamos, asimismo, que nuestros objetivos estaban encaminados hacia presentar un prototipo en el que se abordaba el problema de estructurar el léxico en la BCL de un SRI de corte onomasiológico. En este sentido, definimos qué es una BCL y ejemplificamos, a partir de las necesidades que nos plantea el proyecto del diccionario electrónico de búsquedas onomasiológicas, el tipo de información que debe incorporar esta BCL.

Al final del capítulo describimos una de las herramientas que nutre esta BCL de manera automática, el AlgCl. De esta herramienta rescatamos el conjunto de elementos que, a modo de resultado, etiqueta como pares semánticos (Sierra 1999, Castillo 2002, Sierra y McNaught 2002), es decir, como palabras que comparten características semánticas que las hacen comportarse de manera similar.

Señalamos que, aunque algunas veces estas palabras sí comparten características semánticas, en algunas otras no. Así, estas últimas son las que, dado el fenómeno que

implica su compartimiento que las hace funcionar como equivalentes, integran el corpus del que hablamos en el Capítulo 2.

Cabe mencionar que a partir del comportamiento de las formas que analizamos, buscamos proporcionar al SRI una BCL que pueda integrar información que no es específica de un diccionario o una ontología, sino de la realidad lingüística de los hablantes. En este sentido, el punto más importante con el que concluimos el capítulo consistió en establecer el hecho de que para que un SRI, sea del corte que fuere, pueda devolver al usuario la información que éste precisa, no basta sólo con incorporar en la BCL un repertorio léxico común a un dominio particular sino en procurar que ésta concentre, por medio de ítems léxicos, información que represente no sólo a las palabras sino al contenido de éstas (Kang et al. 2004).

### **5.1.2 Capítulo 2**

Una vez que expusimos cuáles eran los elementos que nos interesaban, dado su atípico comportamiento, reprodujimos en el Capítulo 2 los procesos que efectuamos para amalgamar un corpus que denominamos de análisis. Este CA agrupaba un conjunto de verbos que, según tres fuentes especializadas, no compartían un rasgo común de sinonimia y que, sin embargo, a nivel contexto, funcionaban como si lo compartieran.

El paso que luego realizamos fue el de observar el comportamiento que los pares del CA tenían en las definiciones de las que fueron extraídos. Con ello corroboramos que el AlgCl, aún cuando no tomaba en cuenta la semántica de las palabras y, por ende, agrupaba palabras que no compartían características semánticas (Castillo 2002), había propuesto pares que si bien, no podíamos llamar semánticos, sí se comportaban como si compartieran rasgos semánticos que les permitieran actuar como sinónimos.

Ahora bien, dado ese comportamiento, quisimos eliminar la premisa de que ése estaba dado en función únicamente del tipo definitorio de contexto. Para ello, construimos un CP en donde recuperamos contextos de distintas áreas en donde aparecían estos

elementos. El fin fue el de observar el comportamiento de éstos fuera de un marco definitorio.

El resultado que obtuvimos con el CP fue el de corroborar que ese comportamiento no era privativo de un contexto definitorio. La posibilidad de que cualquiera de los dos miembros que constituían el par pudiera aparecer indistintamente sin que se perdiera el significado global de la construcción fue motivo para determinar que el contexto no era la única vía por el que se podía explicar la conducta de esos verbos.

Con base en esa premisa, desarrollamos una hipótesis que llamamos cognoscitiva. Lo que pretendíamos probar con ella era que el comportamiento de esos verbos no se debía a un hecho como el contexto o a una cuestión de estilo, sino a factores más complejos que, suponíamos, involucraban procesos en los que la cognición se reflejaba no sólo a través de relaciones léxicas, sino por relaciones más allá de esta esfera.

### 5.1.3 Capítulo 3

A partir de la hipótesis que planteamos en el capítulo previo, en éste el objetivo fue comprobarla. La forma en la que lo hicimos fue a través de un análisis en el que retomamos las bases de la gramática cognoscitiva.

Con base en los conceptos teóricos de los modelos CxG y Frame Semantics, propusimos una argumentación a la aparición indistinta de los verbos del CA en diversos contextos en los que funcionaban como equivalentes.

El primer paso fue señalar que había un significado de construcción que, retomando un enfoque gestáltico, suponía que las construcciones se perciben como una totalidad, es decir, que *el todo es más que la suma de las partes*. El significado existe y adquiere un sentido específico al interior de un contexto; de manera aislada, el significado no es más que la representación de una entrada de diccionario.

Como segundo paso de la argumentación, propusimos un esquema en el que explicábamos la aparición indistinta de los verbos en función de un *frame* que permitía codificar significados particulares bajo la premisa de que hay significados más generales que licencian los primeros.

En este sentido, lo que pretendimos aprehender con la noción de *frame* es lo que Kang et al. (2004) consideran como *concept cluster*. Esto es, un elemento que pueda representar la estructura conceptual que entrañan las palabras.

Finalmente, luego de la argumentación que hicimos a los verbos del CA, concluimos que, más allá del significado contextual, es evidente que existen patrones cognoscitivos que licencian que un verbo pueda estar en sustitución de otro debido a que, tanto para el hablante como para el oyente, existen marcos que engloban estructuras más generales que, por medio de esquematizaciones, adoptan significados particulares, dando sentido y cohesión a las construcciones en donde esos verbos funcionan como pares semánticos.

Asimismo, asentamos que no bastaba con un plano semántico para dar cuenta, en forma más completa, de lo que sucedía con esos verbos y, en un plano más general, con la lengua. Si bien, para nuestros intereses, la estructura a base de *frames* podía representar un avance en la incorporación en la BCL de léxico que reflejara fenómenos como los vistos en los verbos del CA, no era suficiente para que el SRI pudiera semejar procesos como los que debía realizar para devolver al usuario el término que buscaba.

#### **5.1.4 Capítulo 4**

Como última parte de la investigación, desarrollamos a lo largo del Capítulo 4 el cuestionamiento sobre la necesidad de incorporar un análisis de corte pragmático para robustecer la estructura de la BCL.

Abordamos este cuestionamiento a partir del problema que representaron en todo el trabajo los verbos del CA. Con ellos quisimos señalar, de manera muy acotada, que la



tarea de integrar información propia de la comunicación humana en la BCL es un hecho que, si bien implica un reto enorme, puede también representar un beneficio en el momento de implementar estructuras que reflejen la realidad lingüística de hablante y oyente.

Por último, sugerimos una manera de incorporar información pragmática en la BCL. Con el concepto de descriptor quisimos ejemplificar que esta clase de información puede integrarse como parte de la estructura de *frames*.

La forma en que pretendemos concentrar esa información es a partir de un análisis pragmático con base en el marco teórico de TR. El concepto que subyace a este modelo lo adoptamos para justificar cuáles podrían ser los elementos que representarían, en mayor medida, el conocimiento que queremos representar en la BCL.

El resultado que deberíamos obtener con lo que sugerimos es que el SRI pueda realizar búsquedas con mayor nivel de profundidad y expansión, logrando con ello que el usuario se beneficie con un sistema que, semejando la estructura conceptual y los procesos de comunicación propios de los humanos, le represente una ventaja en cuanto a tiempo y calidad de resultados.

## **5.2 Conclusiones generales**

A lo largo de la investigación hemos desarrollado la hipótesis de que es posible incorporar elementos de la cognición, a través del análisis de la comunicación de los humanos, en la BCL de un SRI específico.

La tesis que sustenta esto se basa en analizar la realidad y el entorno lingüístico tanto de hablantes como oyentes. La forma en la que nos propusimos demostrarlo fue a partir del estudio de un grupo de verbos que, según tres fuentes especializadas, no comparten una relación sinonímica pero que a nivel discurso se comportan como si la tuvieran.

El punto que queríamos tocar con los elementos que el AlgCl proponía como relacionados semánticamente, sin que para ello se hubiera realizado un análisis previo, era tratar de evidenciar que esto no era una mera cuestión de estilo o de contexto sino de factores que residen en un plano cognoscitivo que licencia la codificación del significado.

Bajo esta perspectiva, nuestra hipótesis fue tratar de demostrar que la codificación del significado, independientemente del verbo que aparecía en la construcción, estaba en función de evocar representaciones abstractas, comunes a hablantes y oyentes, que daban sentido a un significado más particular expresado con alguno de los verbos del CA. Así, el análisis que efectuamos a los pares del CA estuvo en función de comprobar tal hipótesis.

De acuerdo con lo anterior, cada *frame* que propusimos pretendía demostrar que, más allá de un significado léxico, lo que codificamos en nuestra comunicación es un significado nuclear que se compone de distintos segmentos de información, entre ellos el léxico. Visto así, no importaba qué elemento aparecía en la construcción puesto que no codificábamos su significado básico sino fragmentos de conocimiento encapsulados en la construcción en la que aparecía ese elemento.

Ahora, otro punto con el que quisimos dar énfasis a la idea que trabajamos en la investigación fue la de señalar que, además de tomar en cuenta una estructura de léxico basada en la semántica, era necesario un trabajo que retomara los problemas que atañen a la pragmática. Bajo esta idea, quisimos demostrar que resultaba muy importante hacer un estudio en el que se incluyera una perspectiva pragmática con el fin de incorporar más y mejores patrones en la BCL; básicamente, pretendíamos hacer constar que la lengua no se construye sólo vía patrones gramaticales sino vía factores que involucran elementos fuera de la lengua que adquieren significado al interior de un sistema lingüístico y sociocultural.

Con los ejemplos que incluimos en el módulo pragmático pretendimos mostrar cuán difícil es tratar de incorporar información de índole pragmático en un sistema computacional; no obstante, ofrecimos una propuesta que pretendía rescatar esa información y, de algún modo, integrarla en la BCL.

El concepto de descriptor fue el medio con el cual quisimos aprehender el factor pragmático y representarlo en una forma en que el SRI lo pudiera interpretar. Claro está que con un concepto como éste no es posible asir toda la información que se transmite en un plano comunicativo, no obstante, parte del significado de ésta, que es privilegiado por hablantes y oyentes, podría simbolizarse mediante un descriptor.

Todo lo anterior, sin embargo, no es sino parte de un acercamiento a una investigación más extensa que debe involucrar un análisis mucho mayor que pueda reflejar a más profundidad los fenómenos lingüísticos y comunicativos para poder incorporarlos a la BCL de cualquier tipo de sistema computacional. Pese a todo, creemos que con lo que desarrollamos en la investigación cumplimos con los objetivos que nos planteamos en un principio.

El objetivo general de presentar un esquema que pudiera organizar el léxico de una BCL lo cumplimos al presentar una estructura a partir de *frames*, por el lado semántico, y de descriptores, por el pragmático. De esta manera, además de cumplir nuestro objetivo, justificando con ello los resultados del AlgCl, dimos testimonio de que es viable aproximarse a los procesos de cognición que se reflejan en la comunicación, y que es posible armar una estructura léxica con base en análisis que tomen en cuenta a hablantes y oyentes y su realidad comunicativa.

Ahora bien, dado nuestro interés por analizar algo tan abstracto y sin resultados contundentes como lo es la cognición, es justo señalar que la argumentación que hicimos puede parecer construida a priori e incluso, frágil. En este sentido, es claro que no podemos declarar que tenemos evidencia formal de procesos de cognición; no obstante, pensamos que también es justo decir que toda la argumentación parece viable y, sobre todo, operativa para los fines que perseguimos. A fin de cuentas, nuestro interés no es explicar lo que sucede en nuestras mentes ni dar cuenta completa de la lengua sino tratar de identificar y extraer patrones, justificados con la lengua en uso, que puedan ser incorporados en nuestra BCL.

Con esta perspectiva, podemos concluir que la hipótesis que señalamos al principio de la investigación tiene, al menos, pauta. Y ésta está justificada en la lengua. En ese sentido, consideramos que el razonamiento que realizamos y que dio pie para los esquemas que planteamos puede servir de base para representar tanto el léxico como el conocimiento en la BCL de nuestro SRI.

### 5.3 Trabajo futuro

Hasta aquí hemos desarrollado el análisis, la argumentación y la propuesta para estructurar el léxico en la BCL del SRI de búsquedas onomasiológicas. Si bien creemos que, al menos dentro de los alcances de nuestro proyecto, la idea de incorporar información lingüística, codificada por el análisis semántico, y extralingüística, codificada por el análisis pragmático, resulta interesante, operativamente pensamos que la idea podría servir como base para un proyecto de mayor alcance. En esencia, queremos decir que a partir de un acercamiento de este tipo, podríamos encaminar nuestras investigaciones hacia horizontes más complejos.

Es claro que con nuestra investigación sólo nos centramos en un objetivo muy específico enmarcado en un proyecto bien delimitado y que de cierta forma resolvimos con la estructura de *frames*, es cierto también que no realizamos un análisis más profundo en el que involucráramos nuestra propuesta pragmática. No obstante, conforme la idea central de la tesis, pensamos que podríamos desarrollarla a profundidad para dirigir nuestros esfuerzos más allá de los SRI. En este sentido, nuestro más fuerte interés es la *minería de textos*.

De acuerdo con Liddy (2000), podemos considerar la minería de textos como aquel proceso de análisis a través del cual se pretende descubrir y capturar información semántica para almacenarla dentro de una *estructura de organización de conocimiento* (KOS), con el fin de que sea utilizada como fuente de conocimiento.

Profundizando nuestras investigaciones pretendemos, a través de un sistema de minería de textos, proporcionar una visión más amplia, a la vez que selectiva, de la información que constituye los inmensos volúmenes de documentos. Con ello intentamos

sobrepasar los límites de los SRI, los de resumen automático, los de traducción automática o los de extracción de información, entre otros, expandiendo los alcances de búsqueda de información para dinamizarla y convertirla en conocimiento.

Para ilustrar hacia donde nos dirigimos, citamos las investigaciones de Swanson y Smalheiser<sup>1</sup> quienes aplicaron un sistema de minería de textos a un corpus sobre publicaciones del área de la biomedicina. Lo que encontraron fue información del tipo:

- El estrés se asocia a las migrañas.
- El estrés genera pérdidas de magnesio.
- Los bloqueadores de calcio previenen las migrañas.
- El magnesio es un bloqueador natural del calcio.
- La difusión de la depresión cortical (SCD) aparece en casos de migraña.
- Altos niveles de magnesio inhiben el SCD.
- Los pacientes con migraña tienen una alta agregación de plaquetas.
- El magnesio puede suprimir la agregación de plaquetas.

El sistema pudo arrojar esta información gracias a la estructura del léxico en la BCL, el cual se sustentaba en análisis morfosintácticos y semánticos principalmente. Ahora, el punto no es que el sistema devolviera esta información sino que ésta, a través de la investigación de Swanson y Smalheiser, reveló una relación de la que nadie se había dado cuenta hasta entonces: *un vínculo entre la falta de magnesio y algunos tipos de migrañas.*

Este ejemplo demuestra la importancia de la estructura del léxico en las BCL's. Si éste no hubiese estado estructurado en la base de conocimiento del sistema de minería de textos que ocuparon Swanson y Smalheiser, la información se habría quedado sólo en las bases y no se habría transformado en conocimiento que no existía explícitamente (Hearst 1999).

---

<sup>1</sup> Tomado del artículo de C. Obeso. *La minería de textos.* [en línea] <http://www.infonos.com/faq/2.5.html>

Si lográramos avanzar en análisis que nos permitieran proyectar un sistema de minería de textos, podríamos llegar hacia horizontes en donde el conocimiento lingüístico, estructurado en una BCL, sería aprovechado en proyectos como los que se muestran a continuación:

- ❖ Inversión de capitales
  - Búsqueda y extracción de información
  - Estimación de relevancia
  - Navegación y resumen automático
  
- ❖ Compañías de investigación o de seguros
  - Categorización de mensajería
  - Compartir información con empleados
  - Extracción de fragmentos estructurados y características
  
- ❖ Dependencias gubernamentales
  - Recuperación de información inteligente
  - Trazado de cadena de eventos
  - Suplir documentos con sus resúmenes
  
- ❖ e-Business
  - Aprender los intereses de los usuarios
  - Corresponder a los intereses de los usuarios con recursos disponibles
  
- ❖ Ingeniería de sistemas
  - Ingeniería de sistemas de búsquedas
  - Portales de internet
  - Sistemas de manejo de conocimiento

Claro está que para realizar proyectos de esta envergadura es necesario realizar análisis más profundos que involucren, además de los planos semántico y pragmático, el sintáctico y el morfológico. Todo ello combinado con análisis sobre un corpus mucho más

rico en donde se extiendan los análisis del plano verbal hacia el nominal y adjetival. Con ello estaríamos perfilando nuestra investigación lingüística hacia una vertiente muy aplicada en donde todos los resultados estarían encaminados a aprovechar y dinamizar la información para convertirla en conocimiento.

## BIBLIOGRAFÍA

---

- Aguilar, C. 2004. "Uso de información sintáctica y semántica para el reconocimiento de definiciones en corpus de especialidad". En *1º Encuentro de Estudiantes de Literatura y Lingüística UNAM*, Fac. de Filosofía y Letras UNAM, México.
- Aguilar, C. y G. Sierra. 2003. "Análisis de predicaciones verbales dentro de contextos definitorios". En *Actas del VII Congreso Nacional de Lingüística*, Universidad de Guadalajara, México.
- Akman, Varol. 2002. *Context in Artificial Intelligence: A Fleeting Overview. English version of "Contesti in intelligenza artificiale: una fugace rassegna"*, La Svolta Contestuale, Carlo Penco. McGraw-Hill, Milano, [en línea] <http://www.cs.bilkent.edu.tr/~akman/papers.html>.
- Arotariteia, D. y S. Mitrab. 2004. *Web Mining: a Survey in the Fuzzy Framework. Information Sciences*. Elsevier. In press. 1 – 15.
- Atkins, B. 1995. "The Role of the Example in a Frame Semantics Dictionary". En *Essays in Semantics and Pragmatics*. Eds. M. Shibatani y S. Thompson. John Benjamins. Amsterdam/Philadelphia. 25 – 42.
- Bor, K. 2002. "The applicability of universal pragmatics in information retrieval interaction: a pilot study". En *Information Processing and Management* 38, 237 – 248.
- Bretones, C., M. Cristóbal e I. Ibarretxe. 2001. *Exiting events in spanish: boundary i-schema and move x-schema*. International Computer Science Institute Technical Reports 01-007.
- BwanaNet*. [en línea]: <http://bwananet.iula.upf.edu/>
- Castillo, G. 2002. *Algoritmo revisado para la extracción automática de agrupamientos semánticos*. Tesis de Maestría. UNAM.
- Copestake, A. 1995. "[The representation of group denoting nouns in a lexical knowledge base](#)". En: P. St. Dizier and E. Viegas (eds), *Computational Lexical Semantics*, Cambridge University Press. *Earlier version in Proceedings of the Second Seminar on Computational Lexical Semantics*, Toulouse, France.
- Corpus de Referencia del Español Actual (CREA). [en línea]: <http://www.rae.es>



- Cristóbal, M. 2001. *Arriving events in English and Spanish: a contrastive analysis in terms of Frame Semantics*. International Computer Science Institute. Technical Reports 01 – 009.
- Curcó, C. *Apuntes de clase*. Semestre 2004 – 1.
- Davies, M. *Corpus del español*. Illinois State University/NEH. [en línea]: <http://www.corpusdelespanol.org>
- Diccionario Anaya* (base de datos en formato electrónico).
- Diccionario de la lengua española*. Espasa Calpe. Versión 21.1.0. 1995.
- Diccionario de uso del español actual (CLAVE)*. Diccionarios SM. 1997.
- Diccionario Larousse*. Larousse Planeta, edición electrónica en CD ROM, 1996.
- Domenig, M. 1988. “Word Manager: A System for the Definition, Access and Maintenance of Lexical Databases”. En *Proceedings of the 12th conference on Computational linguistics*. Budapest, Hungary. August 22 – 27, 154 – 159.
- Escandell-Vidal, V. 2004. “Norms and Principles. Putting Social and Cognitive Pragmatics Together”. To appear in *Current Trends in the Pragmatics of Spanish*. R. Márquez y M.E. Placencia (eds.). Amsterdam, John Benjamins.
- Eumeridou, E., B. Nkwenti-Azeh, y J. McNaught 2004. *An Analysis of Verb Subcategorization Frames in Three Special Language Corpora with a View towards Automatic Term Recognition*. Computers and the Humanities 38: 37–60. Kluwer Academic Publishers. Netherlands.
- Euro WordNet* (Interfaz 0.1) [en línea]: <http://nipadio.lsi.upc.es/cgi-bin/public/wei1.consult.perl>).
- Fallbaum, C. (ed.). 1998. *WordNet: an electronic lexical database*. Cambridge: MIT Press.
- Fillmore, C. 1968. “The Case for Case”. In Bach, E. and Harms, R.T. (eds.), *Universals in Linguistic Theory*, North Holland, New York, pp. 1–88.
- Fillmore, C. 1982. “Frame Semantics”. En *Linguistics in the Morning Calm. Selected Papers from SICOL-1981*. Seoul: Hanshing Publishing Company. 113-137.
- Fillmore, C., C. Wooters, y C. Baker. 2001. “Building a Large Lexical Databank Which Provides Deep Semantics”. En *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. Hong Kong. [en línea]: <http://www.icsi.berkeley.edu/~framenet/>

- Fukumoto, F. y S. Yoshimi. 1996. "An Automatic Clustering of Articles Using Dictionary Definitions". En *The 16<sup>th</sup> International Conference on Computational Linguistics. COLING 1996*, Vol. 1. pp. 406 – 411.
- Galinsky, C. 1991. "Terminología y sociedad moderna: el papel de INFOTERM". *Las industrias de la lengua*. trad. E. Lavín. Madrid: Fundación Germán Sánchez Ruipérez.
- Gates, D. y P. Shell. 1993. "Rule-based Acquisition and Maintenance of Lexical and Semantic Knowledge". En *Proceedings of the Sixth Conference on European Chapter of the Association for Computational Linguistics*. Utrecht, The Netherlands. April 21 – 23, 149 – 157.
- Gemino, A., D. Parker, A. Kutzschan. 2003. "Is Relevance Relevant? Investigating Coherence in Knowledge Sharing Environments". En *Proceedings of the Second Annual Workshop on HCI Research in MIS*, Seattle, WA, December 12-13, [en línea]: [http://cte.rockhurst.edu/sighci/icis\\_2003/HCI03\\_18.pdf](http://cte.rockhurst.edu/sighci/icis_2003/HCI03_18.pdf)
- Gennari, S. y D. Poeppel. 2003. "Processing correlates of lexical semantic complexity". En *Cognition* 89, B27 – B41.
- Ghidini, C. y F. Giunchiglia. 2001. "Local Models Semantics, or contextual reasoning = locality + compatibility". En *Artificial Intelligence* 127, 221–259.
- Giora, R. 1997. "Discourse coherence and theory of relevance: Stumbling blocks in search of a unified theory". En *Journal of Pragmatics* 27, 17 – 34.
- Giora R. 1998. "Discourse coherence is an independent notion: A replay to Deirdre Wilson". En *Journal of Pragmatics* 29, 75 – 86.
- Giulia M., L. Bencini y A. Goldberg. 2000. "The Contribution of Argument Structure Constructions to Sentence Meaning". En *Journal of Memory and Language* 43, 640–651.
- Goldberg, A. 1996. "Making One's Way Through the Data". En *Grammatical Constructions. Their Form and Meaning*. Oxford: Clarendon Press. 29 – 53.
- Goldberg, Adele. 1997. "Construction Grammar". In E.K. Brown & J.E. Miller (eds.), *Concise Encyclopedia of Syntactic Theories*. New York: Elsevier Science Limited.
- Goldberg, A., D. Casenhiser y N. Sethuraman 2003a. "A lexically based proposal of argument structure meaning". *Proceedings of the Annual Chicago Linguistics Society*.
- Goldberg, A., D. Casenhiser y N. Sethuraman. 2003b. "Learning Argument Structure Generalizations". In press. To appear in *Cognitive Linguistics*.
- Grice, P. 1989. "Logic and conversation". En *Studies in the way of words*. Cambridge, Harvard University Press.

- Grootjen, F. y P. van der Weide. 2002. "Conceptual Relevance Feedback". En *Proceedings of the 2002 IEEE International Conference on Systems, Man and Cybernetics, (NLPKE 2002)*, Tunis.
- Grootjen, F. y P. van der Weide. 2003. *Information Retrieval as Semantics Transformation*. Technical Report NIII-R0303, Computing Science Institute, University of Nijmegen, Nijmegen, The Netherlands.
- Hearst, M. 1999. "Untangling Text Data Mining". En *Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics*, University of Maryland, June 20-26.
- Hjørland, B. 1998. "Information Retrieval, Text Composition, and Semantics". En *Knowledge Organization 25 (1/2)*, Royal School of Library and Information Science, Copenhagen. 16 – 31.
- Hobbs, J. 1978. "Why is discourse coherent?". En *SRI International*.
- Johnson, C. y C. Fillmore. 2000. "The FrameNet Tagset for Frame-Semantic and Syntactic Coding of Predicate-Argument Structure". En *The Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000)*, April 29-May 4, 2000, Seattle WA, 56-62. [en línea] <http://www.icsi.berkeley.edu/~framenet/>
- Kang, B., D. Kim y S. Lee. 2004. "Exploiting Concept Clusters for Content-Based Information Retrieval". In press en *Information Sciences*. Elsevier.
- Korhonen, A. 2002. "Assigning Verbs to Semantic Classes via WordNet". En *Proceedings of the COLING Workshop on Building and Using Semantic Networks*. Taipei, Taiwan.
- Korhonen, A. y T. Briscoe. 2004. "Extended Lexical-Semantic Classification of English Verbs". To appear in *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*. Boston, MA.
- Krovetz, R. 1991. "Lexical Acquisition and Information Retrieval". En *Lexical Acquisition: Building the Lexicon Using On-Line Resources*. U. Zernik (Ed.). 45 – 64.
- Lance, G. y W. Williams. 1967. "A general Theory of Classificatory Sorting Strategies". En *Advances in Computer Typesetting: Proceedings of the 1966 International Computer Typesetting Conference*. Ed. W. Jaspert. London: The Institute of Printing. 271 - 277.
- Langacker, R. 1987. *Foundations of Cognitive Grammar. Volume 1*. USA: Stanford University Press.
- Langacker, R. 1991. *Concept, Image and Symbol. The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter. 315 - 342.

- Liddy, E. 2000. "Text Mining". En *Bulletin of the American Society for Information Science* Vol. 27, No.1. [en línea] <http://www.asis.org/Bulletin/Oct-00/liddy.html>
- Luhmann, N., 1990. *Essays on Self-References*. Columbia University Press, New York.
- McCarthy, D. 2001. *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences*. Ph. Thesis. University of Sussex.
- Montgomery, C. 1972. "Linguistics and Information Science". En *Journal of the American Society for Information Science*. 195 – 218.
- Nerlich, B. y D. Clarke. 2000. "Semantic fields and frames: Historical explorations of the interface between language, action, and cognition". En *Journal of Pragmatics* 32, 125 - 150.
- Noppeney, U. y C. Price. 2004. "Retrieval of Visual, Auditory, and Abstract Semantics". En *NeuroImage*, Volume 22, 164 - 170.
- Obeso, C. *La minería de textos*. [en línea] <http://www.infonos.com/faq/2.5.html>
- Petruck, M. 1995. "Frame Semantics and the Lexicon: Nouns and Verbs in the Body Frame". En *Essays in Semantics and Pragmatics*. Eds. M. Shibatani y S. Thompson. John Benjamins. Amsterdam/Philadelphia. 279 – 298.
- Petruck, M. 1996. "Frame Semantics". En J. Verschueren, J. Ostman, J. Blommaert, and C. Bulcaen (eds) *Handbook of Pragmatics*. Amsterdam/Philadelphia: John Benjamins.
- Pinar, A. e I. Ciceklib, 2002. "Pragmatics in human-computer conversations". En *Journal of Pragmatics* 34, 227-258.
- Poznański, V. 1992. *A Relevance-based Utterance Processing System*. Ph. Thesis. University of Cambridge. England.
- Pustejovsky, J. 1998. "The Semantics of Lexical Underspecification". En *Folia Linguistica* 32, 323 – 347.
- Ramos, F. 1998. "Relevance theory and media discourse: A verbal-visual model of communication". En *Poetics* 25, 293-309
- Reyes, A. y G. Sierra. 2003. "Patrones para una estructuración semántica de un diccionario onomasiológico". En prensa en *Lingüística Mexicana*.
- Reyes, A. y E. Solares. 2004. "Estructuración de datos en bancos terminológicos para procesos de representación de conocimiento en sistemas de recuperación de información". En *Proceedings del IX Simposio Iberoamericano de Terminología. Barcelona*.

- Rieger, B. 1988. *Relevance of Meaning, Semantic Dispositions, and Text Coherence. Modelling Reader Expectations from Natural Language Discourse*. Amsterdam. John Benjamins. [en línea] <http://www.ldv.uni-trier.de:8080/ldvpage/rieger/pub/aufsaeetze/urbi/urbi-abstract.html>
- Rodrigues, J. M., et al. 1998. *Galen-In-Use: Using artificial intelligence terminology tools to improve the linguistic coherence of a national coding system for surgical procedures*. [en línea] <http://www.univ-st-etienne.fr/dspim/doc/medinf98.doc>
- Rouchota, V. 1998. “Connectives, Coherence and Relevance”. En *Current Issues in Relevance Theory*. Eds. Villy Rouchota y Andreas H. Jucker. John Benjamins. Amsterdam/Philadelphia.
- Sager, J. C. 1990. *A practical course in terminology processing*. Amsterdam, John Benjamins Publish.
- Schank, R., J. Kolodner, y G. DeJong. 1981. “Conceptual information retrieval”. En *Information Retrieval Research*. R.N. Oddy et al (eds). London: Butterworths.
- Sierra, G. 1999. *Design of a concept-oriented tool for terminology*. Ph. Thesis. Department of Language Engineering, UMIST, England.
- Sierra, G., A. Reyes y C. Aguilar. 2005. “Hacia el desarrollo de una base de conocimiento léxico para diccionarios onomasiológicos”. En *Actas del XIV Congreso de ALFAL*.
- Sierra G., et al. 2004. “Definitional Contexts Extraction from Specialised Texts”. En *Practical Applications in Language and Computers PALC 2003*. Peter Lang, Alemania, 2004. 21-31.
- Sierra, G. y G. Castillo. 2003. “Algoritmo flexibilizado de agrupamiento semántico” En *Estudios de lingüística aplicada*. Num.38, 69 – 80.
- Sierra, G. y J. McNaught. 2002. “Analogy-based method for semantic clustering”. En *Computational Linguistics and Intelligent Text Processing*. Fondo de Cultura Económica, México.
- Sierra, G., J. McNaught y G. Castillo. 2002. “Extracting semantically rich knowledge from alignment of definitions”. En *Computational Linguistics and Intelligent Text Processing*. Num.1, Volumen 1, 245-249.
- Sierra G.y R. Alarcón. 2002. “Identification of Recurrent Patterns to Extract Definitory Contexts”. En *Computational Linguistics and Intelligent Text Processing. Colección en Ciencias de la Computación*, Alexander Gelbuck (ed.), Springer Berlin. 436-438.
- Sinha, C. 1999. “Grounding, Mapping and Acts of Meaning”. En *Cognitive Linguistics: Foundations, Scope and Methodology*. Theo Janssen and Gisela Redeker (eds.) Berlin & New York, Mouton de Gruyter, 223 - 255.

- Sperber, D. y D. Wilson. 1995. *Relevance: Communication and Cognition*. 2ª ed. Oxford: Blackwell.
- Sperber, D. y D. Wilson. 1996. “Fodor’s Frame Problem and Relevance Theory (reply to Chiappe & Kukla)”. En *Behavioral and Brain Sciences* 19:3. [en línea] <http://www.dan.sperber.com/frame.htm> 1996.[consulta: 8 de noviembre de 2003].
- Sperber, D. 2005. “Modularity and Relevance: How can a Massively Modular Mind be Flexible and Context-Sensitive? To appear in *The Innate Mind: Structure and Content*. Edited by Peter Carruthers, Stephen Laurence, & Stephen Stich.
- Trujillo, F. “La teoría de la relevancia como base para una nueva interpretación de la comunicación”. En *Eúphoros* 3, 221-232. Universidad de Granada. [en línea] <http://www.ugr.es/~ftsaez/relevancia.pdf>.
- Wilson, D. 1998. “Discourse coherence and relevance: A replay to Rachel Giora”. En *Journal of Pragmatics* 29. 57 - 74.
- Wilson, D. y T. Matsui. 1998. “Recent approaches to bridging: Truth, coherence, relevance”. En *UCL Working Papers in Linguistics* 10 [en línea] <http://www.phon.ucl.ac.uk/home/PUB/WPL/98papers/wilson.pdf> [consulta: 1 de diciembre de 2003].
- Wilson, D. y D. Sperber. 2004. “Relevance Theory”. En G. Ward and L. Horn (eds) *Handbook of Pragmatics*. Oxford: Blackwell, 607-632.
- Wilson, M. 1993. “Knowledge Engineering”. En *Proceedings SOFSEM '93*. Czech Republic, 3 - 12.

APÉNDICE A: FUNCIONAMIENTO GENERAL DEL ALGCL

La identificación de pares se hace con base en el alineamiento de definiciones; para ello, el programa toma para un mismo término, dos definiciones de dos fuentes distintas. La alineación consiste en tomar todas las formas que constituyen una definición e insertarlas en una matriz de formas; a cada forma que compone la definición le corresponde un valor consecutivo desde 0 hasta  $\mathcal{N}$  (Figura 9).

	--	alkalimeter	an	apparatus	for	determining	the	concentration
--	0	1	2	3	4	5	6	7
alkalimeter	1							
an	2							
instrument	3							
for	4							
ascertaining	5							
the	6							
amount	7							

Figura 9: Matriz de formas léxicas que integran las definiciones que se alinearán

Una vez que se le asigna un valor a cada forma en la matriz, el programa ejecuta el algoritmo de *Levenshtein Distance*. Con este algoritmo se genera una matriz de costos la cual está en función de los movimientos (costos) que son necesarios para que una forma de una definición pueda intercambiarse con la de otra definición sin que el significado se altere (Figura 10).

	--	alkalimeter	an	apparatus	for	determining	the	concentration
--	0	1	2	3	4	5	6	7
alkalimeter	1	0	1	2	3	4	5	6
an	2	1	0	1	2	3	4	5
instrument	3	2	1	1	2	3	4	5
for	4	3	2	2	1	2	3	4
ascertaining	5	4	3	3	2	2	3	4
the	6	5c	4	4	3	3	2	3
amount	7	6	5	5	4	4	3	3

Figura 10: Matriz general de resultados para el alineamiento a partir del algoritmo de *Levenshtein Distance*

El siguiente paso consiste en correr el algoritmo de *Wagner & Fisher* con el cual se genera la posible ruta, determinada por los costos de inserción, que se debe seguir para intercambiar las formas de las definiciones (Figura 11).

	--	alkalimeter	an	apparatus	for	determining	the	Concentration
--	0	1	2	3	4	5	6	7
alkalimeter	1	0	1	2	3	4	5	6
An	2	1	0	1	2	3	4	5
instrument	3	2	1	1	2	3	4	5
For	4	3	2	2	1	2	3	4
ascertaining	5	4	3	3	2	2	3	4
The	6	5	4	4	3	3	2	3
amount	7	6	5	5	4	4	3	3

Figura 11: Ruta de formas del algoritmo de Wagner & Fisher

Conforme estos pasos, el AlgCl propone grupos de palabras que pueden ser intercambiables en las definiciones de las cuales son extraídas.



## APÉNDICE B: PARÁMETROS DE CADA CORRIDA DEL ALGCL

---

1. Parámetros de operación utilizados en cada corrida del AlgCl:
  - número de definiciones a considerar (0 equivale a todas) = **0**
  - lematizador = **activado**
  - idioma = **español**
  - mínimo valor de LCC = **5**
  - costo por inserción = **1**
  - costo por borrado = **1**
  - costo por sustitución = **1**
  - costo por igualdad = **0**
  - alineamiento del algoritmo original = **activado**
  - máxima profundidad en la búsqueda de alineamientos posibles = **10**
  
2. Resultados que el AlgCl mostró en cada corrida por base:
  - Mostrar la iteración número (0 equivale a mostrar la última iteración del algoritmo) = **1**
  - Alineamientos = **activado**
  - Bindings = **activado**
  - Clusters = **activado**
  - Filtrado → *Sólo los que generan bindings* = **activado**

ROL	papel	que	<i>actúa</i>	un	<i>ser</i>	vivo	en	la	<i>vida</i>
ROL	papel	que	<i>desempeña</i>	un	<i>sujeto</i>		en	la	<i>sociedad</i>
equal	equal	equal	matched	equal	matched	null	equal	equal	matched

Tabla 10: Ejemplo de la generación de pares semánticos

3. Por cada base el programa dio los siguientes resultados:

a) **lingüística**

<b>Iteración : 1</b>
Estadísticas:
Términos disponibles: 791
Términos procesados: 791
Bindings generados hasta esta corrida: 288
Clusters identificados hasta esta corrida 79

b) **física**

<b>Iteración : 1</b>
Estadísticas:
Términos disponibles: 299
Términos procesados: 299
Bindings generados hasta esta corrida: 190
Clusters identificados hasta esta corrida 74

c) **sexualidad**

<b>Iteración : 1</b>
Estadísticas:
Términos disponibles: 73
Términos procesados: 73
Bindings generados hasta esta corrida: 22
Clusters identificados hasta esta corrida 16

d) **desastres**

<b>Iteración : 1</b>
Estadísticas:
Términos disponibles: 351
Términos procesados: 351
Bindings generados hasta esta corrida: 130
Clusters identificados hasta esta corrida 86

APÉNDICE C: 51 PARES DE VERBOS

registrar - representar	actuar - desempeñar	alterar - modificar
utilizar - hablar	afectar - impactar	causar - provocar
desempeñar - expresar	continuar - preservar	afirmar - decir
marcar - distinguir	variar - preservar	tener - presentar
afectar - caracterizar	variar - mantener	matar - eliminar
explicar - relacionar	considerar - importar	emplear - invertir
obtener - liberar	impresionar - afectar	iniciar - comenzar
emplear - seguir	describir - barrer	unir - enlazar
decir - predicar	atravesar - experimentar	usar - emplear
denotar - implicar	representar - designar	poseer - tener
existir - haber	entender - referir	contener - poseer
llamar - denominar	propagar - viajar	significar - expresar
suplir - sustituir	suponer - considerar	señalar - expresar
demostrar - mostrar	determinar - modificar	especificar - expresar
añadir - agregar	recibir - realizar	reproducir - producir
someter - sujetar	generalizar - usar	existir - ocurrir
transmitir - emitir	incrementar - aumentar	

## APÉNDICE D: PARES RELACIONADOS LÉXICAMENTE

---

Pares con relación sinonímica:

<b>existir -haber</b>	<b>llamar - denominar</b>	<b>usar - emplear</b>
<b>iniciar - comenzar</b>	<b>unir - enlazar</b>	<b>suplir - sustituir</b>
<b>incrementar - aumentar</b>	<b>poseer - tener</b>	<b>alterar - modificar</b>
<b>someter -sujetar</b>	<b>transmitir - emitir</b>	<b>actuar - desempeñar</b>
<b>afirmar – decir</b>	<b>tener - presentar</b>	<b>matar - eliminar</b>
<b>demostrar - mostrar</b>	<b>añadir - agregar</b>	<b>contener - poseer</b>
<b>causar - provocar</b>	<b>afectar - impactar</b>	

Pares con relación antonímica:

<b>variar - preservar</b>	<b>variar - mantener</b>
---------------------------	--------------------------

## APÉNDICE E: DEFINICIONES ALINEADAS

---

- 1) Fuente Def 1 [D008 Cardona. Diccionario] Fuente Def 2 [D009 Mounin. Diccionario]  
**atlas lingüístico**
  - e) Conjunto de mapas que **registran** la distribución de las varias lenguas en el ámbito de una determinada área, una región, una nación, etc.
  - f) Conjunto de mapas que **representan** la distribución de las lenguas y dialectos para un área geográfica dada
  
- 2) Fuente Def 1 [D010 Cerdá Massó. Diccionario] Fuente Def 2 [D188 Diccionario Lingüística.] **bilíngüismo**
  - d) Aptitud de un hablante para **utilizar** indistintamente dos lenguas
  - e) Capacidad de un individuo para **hablar** indistintamente dos lenguas con igual dominio
  
- 3) Fuente Def 1 [D006 Dubois. Diccionario] Fuente Def 2 [D188 Diccionario Lingüística.] **adjetivo atributivo**
  - d) Adjetivo que **determina** a un sustantivo sin palabra de unión
  - e) Adjetivo que **modifica** a un sustantivo sin ninguna palabra de unión
  
- 4) Fuente Def 1 [D010 Cerdá Massó. Diccionario] Fuente Def 2 [D008 Cardona. Diccionario] **sujeto**
  - c) Elemento o sintagma nominal que **recibe** la acción expresada por el verbo
  - d) Posición sintáctica del actante que **realiza** la acción expresada por el verbo
  
- 5) Fuente Def 1 [T076 Alarcos. Gramática de] Fuente Def 2 [T084 Pérez-Rioja. Gramática de] **laísmo**

- d) Tendencia a suprimir la diferencia de funciones entre objeto directo e indirecto en beneficio de la distinción de género que consiste en **generalizar** la y las para las referencias a sustantivos femeninos en la función de objeto indirecto
  - e) Construcción vulgar e incorrecta que consiste en **usar** la y las del pronombre ella tanto para el acusativo como para el dativo
- 6) Fuente Def 1 [D086 Lázaro. Diccionario] Fuente Def 2 [T072 Gili y Gaya. Curso superi] **predicado verbal**
- f) Predicado que se **desempeña** por un verbo
  - g) Predicado que se **expresa** por un verbo
- 7) Fuente Def 1 [D009 Mounin. Diccionario] Fuente Def 2 [D188 Diccionario Lingüística.] **caso ergativo**
- e) Caso que **marca** al agente de los verbos transitivos
  - f) Caso que **distingue** al sujeto activo de un verbo transitivo del sujeto de un verbo intransitivo en las lenguas ergativas
- 8) Fuente Def 1 [T050 Quilis. Tratado de f] Fuente Def 2 [D188 Diccionario Lingüística.] **base de articulación**
- c) Conjunto de hábitos articulatorios y de características acústicas que **afectan** a todos los elementos del plano de expresión de una lengua
  - d) Conjunto de hábitos articulatorios que **caracteriza** a todas los sonidos de una lengua
- 9) Fuente Def 1 [Desconocida] Fuente Def 2 [Física de emergencia] **cinemática**
- d) Parte de la mecánica que estudia el movimiento **sin considerar** las causas que lo producen
  - e) Es la parte de la mecánica que estudia el movimiento de los cuerpos **sin importar** las causas que lo producen
- 10) Fuente Def 1 [Física General Aplicada] Fuente Def 2 [Física de emergencia] **luz**

- e) Energía radiante en forma de ondas de longitud tal que **impresionan** al ojo humano y producen la sensación llamada visión
  - f) Podemos definir la luz como aquel fenómeno físico que **afecta** al ojo humano produciendo lo que llamamos la visión
- 11) Fuente Def 1 [Physics] Fuente Def 2 [Física de emergencia] **segunda ley de Kepler**
- e) La línea entre un planeta y el sol **describe** áreas iguales en el mismo tiempo
  - f) La línea imaginaria que une a un planeta con el sol **barre** áreas iguales en tiempos iguales
- 12) Fuente Def 1 [Larousse] Fuente Def 2 [RAE92] **teoría**
- d) Conjunto organizado de ideas o leyes que sirven para **explicar** determinado orden de fenómenos
  - e) Serie de las leyes que sirven para **relacionar** determinado orden de fenómenos
- 13) Fuente Def 1 [RAE92] Fuente Def 2 [Vox] **energía nuclear**
- e) La **obtenida** por la fusión o fisión de núcleos atómicos
  - f) La **liberada** por la desintegración de los núcleos de los átomos
- 14) Fuente Def 1 [PersonalKyr#2] Fuente Def 2 [PersonalKyr#4] **anticonceptivo**
- d) Métodos con productos o con procesos que se **emplean** para evitar un embarazo
  - e) Métodos y procesos que se **siguen** para evitar la concepción de un ser
- 15) Fuente Def 1 [PersonalKyr#4] Fuente Def 2 [PersonalKyr#5] **adolescente**
- e) Persona que **atraviesa** por la edad del cambio entre ser niño y adulto
  - f) Persona que **experimenta** los cambios correspondientes a la adolescencia
- 16) Fuente Def 1 [13] Fuente Def 2 [21] **sistema afectable (sa)**
- e) sa **representa** cualquier sistema integrado por el hombre y sus elementos necesarios para subsistir

- f) sa **designa** cualquier sistema integrado por el hombre y los elementos que necesita para subsistir: ciudad, regiones productoras de elementos, etc.

17) Fuente Def 1 [51] Fuente Def 2 [30] **desastre (disaster catastrophe)**

- d) Por desastre se **entiende** a los estados mismos de daño
- e) No sólo se **refiere** a los estados mismos del daño sino a todas las consecuencias adversas que se caracterizan por múltiples alteraciones -y hasta rupturas- del orden normal de las relaciones productivas comerciales sociales y políticas de la sociedad

18) Fuente Def 1 [D086 Lázaro. Diccionario] Fuente Def 3 [T056 Alonso. Gramática de] **adjetivo numeral múltiplo**

- o) Adjetivo numeral que **significa** multiplicación
- p) Adjetivo numeral que **expresa** multiplicación por la serie natural numérica

Fuente Def 1 [T092 Seco. Manual de gr] Fuente Def 2 [T146 Bello. Gramática de] **adjetivo numeral partitivo**

- q) Adjetivo numeral que **expresa** división
- r) Adjetivo numeral que **significa** división

Fuente Def 1 [T093 Bello. Gramática de] Fuente Def 2 [D188 Diccionario Lingüística.] **presente**

- s) Tiempo verbal que **significa** la coexistencia del atributo con el momento en que proferimos el verbo
- t) Tiempo verbal que **expresa** la coexistencia de la acción con el acto de la palabra

19) Fuente Def 1 [T148 Beristáin. Español. Edu] Fuente Def 2 [T154 Popova. La gramática] **grado positivo**

- c) Grado que **señala** la simple cualidad sin indicar intensidad
- d) Grado que **expresa** la cualidad de un objeto tal como es y sirve para la formación de otros grados de comparación



- 20) Fuente Def 1 [D086 Lázaro. Diccionario] Fuente Def 2 [D010 Cerdá Massó. Diccionario] **artículo**
- d) Es un morfema exclusivo del sustantivo que **expresa** el género y el número del mismo
  - e) Clase de determinante actualizador del nombre virtual del que **especifica** el género y el número
- 21) Fuente Def 1 [D011 Lara. Diccionario] Fuente Def 2 [T148 Beristáin. Español. Edu] **predicado**
- d) Conjunto de signos lingüísticos que manifiesta aquello que se **dice** del sujeto de la oración
  - e) Lo que se dice -lo que se **predica-** del sujeto
- 22) Fuente Def 1 [T154 Popova. La gramática] Fuente Def 2 [T056 Alonso. Gramática de] **presente actual**
- e) Presente que **denota** una acción que coincide con el momento de hablar
  - f) Presente que **implica** una acción continua ya que dentro de su duración se incluye el momento en que hablamos
- 23) Fuente Def 1 [Desconocida] Fuente Def 2 [María Moliner] **velocidad**
- c) Relación entre el espacio recorrido por un móvil y el tiempo **empleado** en recorrerlo
  - d) Relación entre la cantidad de movimiento de trabajo etc. realizados y el tiempo **invertido** en realizarlos
- 24) Fuente Def 1 [Física Universitaria Sears] Fuente Def 2 [Física Moderna] **primera ley de newton**
- d) Todo cuerpo **continúa** en su estado de reposo o movimiento rectilíneo uniforme a menos que sea impelido a cambiar dicho estado por fuerzas que actúen sobre él
  - e) Todos los cuerpos **preservan** su estado de reposo o movimiento uniforme a menos que se le aplique una fuerza exterior

- 25) Fuente Def 1 [Física Conceptual] Fuente Def 2 [Personal] **fuerza de fricción cinética**
- d) Existe una fuerza de fricción llamada fuerza de fricción cinética cuando **existe** un desplazamiento de un cuerpo sobre otro
  - e) Es la fuerza que existe cuando **ocurre** un desplazamiento de un cuerpo sobre otro
- 26) Fuente Def 1 [Diccionario Enciclop Grijalbo] Fuente Def 2 [Personal] **teoría electromagnética de la luz**
- d) Maxwell desarrollo la teoría del campo electromagnético; demostró que la acción electromagnética se **propaga** en formas de ondas transversales y que las ondas luminosas son de naturaleza electromagnética.
  - e) Lo que la teoría electromagnética de la luz quiere demostrar es que, la luz **viaja** en forma de ondas electromagnéticas (en sí, son ondas electromagnéticas) y que no modifica en lo absoluto sus características o propiedades el medio.
- 27) Fuente Def 1 [Vox] Fuente Def 2 [Clave] **eje**
- e) Recta alrededor de la cual se **supone** que gira una línea para engendrar una superficie o una superficie para engendrar un sólido
  - f) Recta fija alrededor de la cual se **considera** que gira una línea para engendrar una superficie o una superficie para engendrar un cuerpo geométrico
- 28) Fuente Def 1 [35] Fuente Def 2 [8] **deforestación y desertificación**
- d) Cambio ecológico que despoja a la tierra de su capacidad para sostener y **reproducir** vegetación, actividades agropecuarias y condiciones de habitación humana
  - e) La desertificación se define como un cambio ecológico que despoja a la tierra de su capacidad para sostener y **producir** vegetación, actividad agropecuaria y condiciones de habitación humana

## APÉNDICE F: BWANANET

---

Liga electrónica:

<http://www.iula.upf.es/corpus/corpus.htm>

Cada texto que conforma el corpus es seleccionado por especialistas de cada área representada, asimismo son agrupados sobre la base de una clasificación temática y de uso propuesta por los propios especialistas.

Las áreas que constituyen al BwanaNet son:

- i. derecho,
- ii. economía,
- iii. medio ambiente,
- iv. medicina,
- v. informática.

Para su explotación en la red, los textos son etiquetados de acuerdo con el estándar [SGML](#) y siguiendo las directrices marcadas por el "*Corpus Encoding Standard*" ([CES](#)) de la iniciativa [EAGLES](#).<sup>1</sup>

La estructura de los bancos terminológicos, tal y como aparece en la interfaz de búsqueda del BwanaNet es la siguiente:

1. Informática
  - Hardware
  - Organización de las computadoras
  - Software
  - Estructura de datos
  - Teoría de la computación
  - Sistemas de información
  - Metodología de la computación
  - Aplicaciones
  - Comunicación hombre – máquina
  - Entorno

---

<sup>1</sup> Para profundizar más en el diseño y características del BwanaNet, consultar su página: <http://bwananet.iula.upf.edu/indexes.htm>

2. Medio ambiente
  - Medio natural
  - Asentamientos humanos
  - Impacto ambiental
  - Política y derecho ambiental
  - Ciencia y tecnología ambiental y energética
  - Medio social y organizaciones
  
3. Derecho
  - Civil
  - Mercantil
  - Laboral
  - Penal
  - Canónico
  - Constitucional
  - Administrativo
  - Financiero y tributario
  - Internacional público
  - Teoría del derecho
  
4. Medicina
  - Anatomía
  - Organismos
  - Enfermedades
  - Productos químicos y fármacos
  - Técnicas y equipamientos analíticos, diagnósticos y terapias
  - Psiquiatría y psicología
  - Ciencias biológicas
  - Antropología, educación, sociología y fenómenos sociales
  - Planificación y gestión sanitaria
  - Ciencias físicas
  - Tecnología, industria y agricultura
  - Humanidades
  - Información científica
  - Grupos nominales
  
5. Genoma
  - Estructura interna
  - Ingeniería genética
  - Biotecnología
  - Enfermedades
  - Farmacogenómica
  - Inmunología
  - Neurociencia
  - Filogenia
  - Eugenesia
  - Diferenciación

- Investigación genética

## 6. Economía

- General
- Metodología e historia del pensamiento económico
- Métodos matemáticos y cuantitativos
- Microeconomía
- Macroeconomía y economía monetaria
- Internacional
- Financiera
- Pública
- Salud, educación y bienestar
- Trabajo y economía demográfica
- Derecho y economía
- Organización industrial
- Administración y economía de empresa, marketing, contabilidad
- Historia económica
- Desarrollo económico y cambios tecnológicos
- Sistemas económicos
- Economía de la agricultura y de los recursos naturales
- Economía urbana, rural y regional
- Temas especiales

## 7. General

- Prensa
- Lingüística
- Minerales
- Física
- Corpus de Chile

## APÉNDICE G: CRITERIOS DE BÚSQUEDA EN EL BWANANET

### Criterios

#### Ámbitos temáticos

1. Informática: 4 primeros pares
2. Medio ambiente
3. Derecho
4. Medicina
5. Genoma
6. Economía
7. General

Cada dominio consta de una serie de subdominios referentes al área (Figura 12).

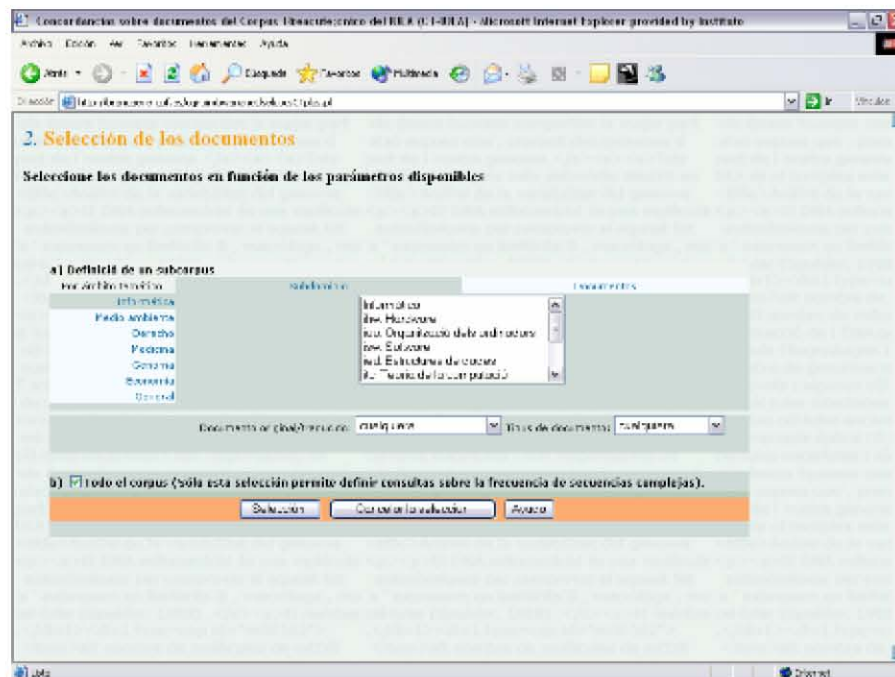


Figura 12: Interfaz de elección de dominios y subdominios

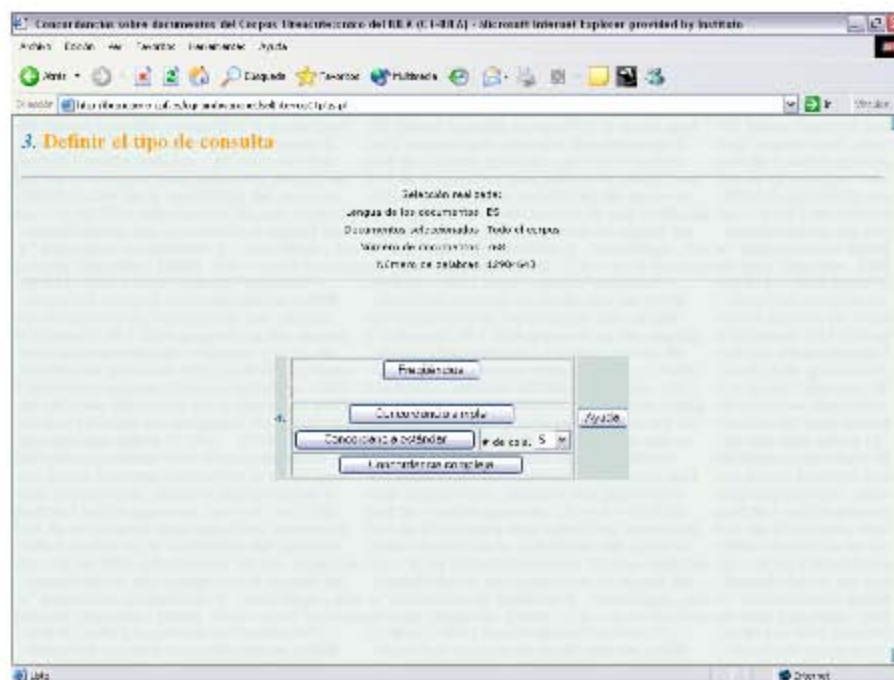
Idioma ESPAÑOL

Un ámbito temático por cada 4 pares.

Seleccionados todos los subdominios de cada ámbito temático.

Documentos originales.

Tipo de documento: CUALQUIERA.



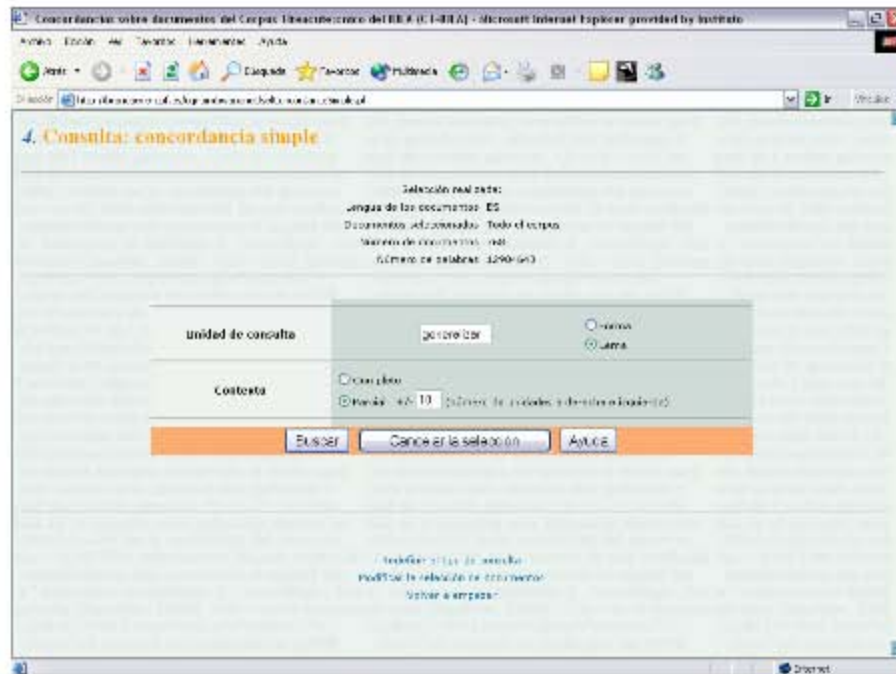
**Figura 13: Interfaz de tipo de consulta**

El tipo de consulta fue por CONCORDANCIA SIMPLE

Selección de todo el corpus.

La unidad de consulta fue el LEMA.

El contexto fue parcial con un rango de 10 unidades a la izquierda y 10 a la derecha de la forma.



**Figura 14: Interfaz de elección de consulta**

Los resultados proporcionados por BwanaNet se limitan sólo a presentar 50 concordancias aunque el total de ellas sea mucho mayor.



## APÉNDICE H: EXTRACTO DE CONSTRUCCIONES VÁLIDAS

---

1. A pesar de que en la matriz solamente se **representan** dos países, el resultado del juego es perfectamente
2. A pesar de que en la matriz solamente se **registran** dos países, el resultado del juego es perfectamente
  
3. ponía a cargo de la filial española a un directivo **hablando** algo de castellano.
4. ponía a cargo de la filial española a un directivo **utilizando** algo de castellano.
  
5. la incidencia impositiva se preocupa de cómo diversas estructuras impositivas **modifican** los precios de equilibrio y por lo tanto la asignación de recursos
6. la incidencia impositiva se preocupa de cómo diversas estructuras impositivas **determinan** los precios de equilibrio y por lo tanto la asignación de recursos
7. El cociente  $(-)(1/6)/(1/3) = (-)1/2$  **determina** la tasa marginal de sustitución técnica de X2 por X1
8. El cociente  $(-)(1/6)/(1/3) = (-)1/2$  **modifica** la tasa marginal de sustitución técnica de X2 por X1
  
9. El Miembro Liquidador Custodio será responsable de las Garantías cuyo Depósito haya **recibido**, habiéndose comprometido frente a MEFF RENTA FIJA a la comunicación
10. El Miembro Liquidador Custodio será responsable de las Garantías cuyo Depósito haya **realizado**, habiéndose comprometido frente a MEFF RENTA FIJA a la comunicación
  
11. Esta regla puede **generalizarse** para cualquiera sea el número de inputs.
12. Esta regla puede **usarse** para cualquiera sea el número de inputs.
  
13. A pesar de ser un factor que **desempeña** un papel en la determinación de los precios relativos,
14. A pesar de ser un factor que **expresa** un papel en la determinación de los precios relativos,
  
15. utilizar los recursos es el objetivo económico inaplazable que **distingue** a los países subdesarrollados de los que no lo son
16. utilizar los recursos es el objetivo económico inaplazable que **marca** a los países subdesarrollados de los que no lo son

17. Dado que nuestro énfasis está centrado en los aspectos que **caracterizan** este mercado por el lado de la demanda,
18. Dado que nuestro énfasis está centrado en los aspectos que **afectan** este mercado por el lado de la demanda,
  
19. Por otro de cantidades fijas (donde únicamente **importan** los resultados cuantitativos).
20. Por otro de cantidades fijas (donde únicamente **consideran** los resultados cuantitativos).
  
21. otra parte, justificaban plenamente la ola de indignación que **impresionó** hasta a la gran Isabel.
22. otra parte, justificaban plenamente la ola de indignación que **afectó** hasta a la gran Isabel.
23. un amigo suyo ha escrito un libro que le ha **impresionado**
24. un amigo suyo ha escrito un libro que le ha **afectado**
  
25. el sol verdadero gira un ángulo xxx y su área **barrida** es igual a: xxx siendo xxx: la ecuación
26. el sol verdadero gira un ángulo xxx y su área **descrita** es igual a: xxx siendo xxx: la ecuación
27. visto en la fórmula de Betz, de la superficie **barrida** por la hélice cuando gira.
28. visto en la fórmula de Betz, de la superficie **descrita** por la hélice cuando gira.
  
29. al cambio de monedas (o de tickets) se **relacionan** tres fenómenos a diferentes niveles y de diferentes mecanismos
30. al cambio de monedas (o de tickets) se **explican** tres fenómenos a diferentes niveles y de diferentes mecanismos
  
31. Esta es precisamente la que se **libera** cuando se quema carbón, madera o petróleo
32. Esta es precisamente la que se **obtiene** cuando se quema carbón, madera o petróleo
33. El capital **liberado** de esta forma se convierte en capital de producción.
34. El capital **obtenido** de esta forma se convierte en capital de producción.
  
35. Los economistas actuales, por otra parte, **emplean** los modelos de optimización intertemporal
36. Los economistas actuales, por otra parte, **siguen** los modelos de optimización intertemporal
37. en conjunto quiere decir que estamos comparando los costes de producción **empleando** métodos distintos
38. en conjunto quiere decir que estamos comparando los costes de producción **siguiendo** métodos distintos

39. era el momento en que Cataluña **atravesaba** dificultades económicas
40. era el momento en que Cataluña **experimentaba** dificultades económicas
41. los trabajadores deben adaptarse a un mundo que está **experimentando** profundos cambios.
42. los trabajadores deben adaptarse a un mundo que está **atravesando** profundos cambios.
43. se remonta a 1810, año en el que Suecia **atravesaba** un difícil momento político.
44. se remonta a 1810, año en el que Suecia **experimentaba** un difícil momento político.
45. Utilizaremos la letra xxx para **designar** la renta per cápita media de todo el país.
46. Utilizaremos la letra xxx para **representar** la renta per cápita media de todo el país.
47. el término gana cierta precisión si hacemos constar que para nosotros **designa** aquella clase de planificación
48. el término gana cierta precisión si hacemos constar que para nosotros **representa** aquella clase de planificación
49. la siguiente función: xxx en la que por xxx **entendemos** la producción agregada.
50. la siguiente función: xxx en la que por xxx **referimos** la producción agregada.
51. La noción de progresiva liberalización **significa** el compromiso por parte de los Miembros
52. La noción de progresiva liberalización **expresa** el compromiso por parte de los Miembros
53. Como se ha **señalado** repetidamente, existe una interrelación entre el derecho
54. Como se ha **expresado** repetidamente, existe una interrelación entre el derecho
55. El profesor Díez de Velasco **señala** con acierto que “no debe confundirse el consensus
56. El profesor Díez de Velasco **expresa** con acierto que “no debe confundirse el consensus
57. las Garantías Diarias es fijado por MEFF RENTA FIJA de acuerdo con lo **especificado** en el presente Reglamento
58. las Garantías Diarias es fijado por MEFF RENTA FIJA de acuerdo con lo **expresado** en el presente Reglamento
59. A través de esta discusión, excepto cuando se **especifique** lo contrario
60. A través de esta discusión, excepto cuando se **exprese** lo contrario
61. Tampoco se puede **decir** que el crecimiento económico y la mayor integración hayan resuelto el problema
62. Tampoco se puede **predicar** que el crecimiento económico y la mayor integración hayan resuelto el problema

63. Y ya en mis mismas palabras se **denota** que esto no es así, cosa que debiéramos resolver
64. Y ya en mis mismas palabras se **implica** que esto no es así, cosa que debiéramos resolver
65. se ve, globalmente hablando, que el consultor, **emplea** entre una hora a hora y media en cada uno
66. se ve, globalmente hablando, que el consultor, **invierte** entre una hora a hora y media en cada uno
67. tiene la ventaja para la firma o el consultor de **invertir** poco tiempo y por lo tanto de rentabilizar al máximo
68. tiene la ventaja para la firma o el consultor de **emplear** poco tiempo y por lo tanto de rentabilizar al máximo
69. Las horas **invertidas** en buscar a un ejecutivo se detraerán del tiempo
70. Las horas **empleadas** en buscar a un ejecutivo se detraerán del tiempo
71. Observando se en ciertos casos la tendencia de **preservar** la vida rural
72. Observando se en ciertos casos la tendencia de **continuar** la vida rural
73. Conservar y restaurar el hábitat para **preservar** dicha especie.
74. Conservar y restaurar el hábitat para **continuar** dicha especie.
75. Un aumento de la ocupación sólo puede **ocurrir** acompañada de un descenso en la tasa de salarios reales
76. Un aumento de la ocupación sólo puede **existir** acompañada de un descenso en la tasa de salarios reales
77. El sonido se **propaga** en todas direcciones a través de l aire
78. El sonido **viaja** en todas direcciones a través de l aire
79. **Supongamos** que existen siete centros (A G)
80. **Consideremos** que existen siete centros (A G)
81. Además, muchas empresas que normalmente se **consideren** parte del sector moderno suelen no cumplir
82. Además, muchas empresas que normalmente se **suponen** parte del sector moderno suelen no cumplir
83. está, por lo tanto, utilizando más de ambos inputs para **producir** el mismo nivel de actividad que el hospital más eficiente
84. está, por lo tanto, utilizando más de ambos inputs para **reproducir** el mismo nivel de actividad que el hospital más eficiente

## TABLA DE ABREVIATURAS

---

AlgCl: Algoritmo de Clustering  
BCL: Base de Conocimiento Léxico  
CA: Corpus de Análisis  
CP: Corpus Prueba  
CxG: Construction GRammar  
DEBO: Diccionario Electrónico de Búsquedas Onomasiológicas  
GIL: Grupo de Ingeniería Lingüística  
SRI: Sistema de Recuperación de Información  
SW: Sperber y Wilson  
TR: Teoría de la Relevancia