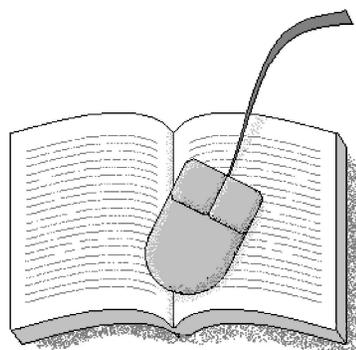
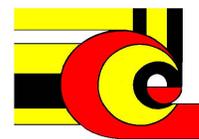




**Instituto Politécnico Nacional**  
**Centro de Investigación en Computación**  
**Maestría en Ciencias de la Computación**  
**Laboratorio de Lenguaje Natural y Procesamiento de Texto**



**Selección automática  
de primitivas semánticas  
para un diccionario explicativo  
del idioma español**

TESIS QUE PRESENTA  
**Lic. GABRIELA RIVERA LOZA**  
PARA OBTENER EL GRADO DE  
**MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**  
DIRECTOR DE TESIS  
**DR. ALEXANDER GELBUKH**  
CO-DIRECTOR  
**DR. GRIGORI SIDOROV**

México, D.F., 2003

## ÍNDICE GENERAL

<b>Resumen .....</b>	<b>7</b>
<b>Abstract .....</b>	<b>8</b>
<b>1    <b>Introducción.....</b></b>	<b>9</b>
<b>2    <b>Métodos de procesamiento automático de información semántica.....</b></b>	<b>13</b>
<b>3    <b>Algoritmo para la selección automática de primitivas semánticas .....</b></b>	<b>53</b>
<b>4    <b>Metodología experimental y resultados obtenidos.....</b></b>	<b>97</b>
<b>5    <b>Conclusiones .....</b></b>	<b>109</b>
<b>6    <b>Bibliografía .....</b></b>	<b>113</b>
<b>Anexo 1. Formato del archivo de entrada.....</b>	<b>119</b>
<b>Anexo 2. Lista de las primitivas encontradas .....</b>	<b>126</b>
<b>Anexo 3. Lista de los ciclos encontrados .....</b>	<b>131</b>
<b>Anexo 4. Ejemplos de ciclos en el diccionario. ....</b>	<b>147</b>

# ÍNDICE DETALLADO

<b>Resumen .....</b>	<b>7</b>
<b>Abstract .....</b>	<b>8</b>
<b>1    <b>Introducción.....</b></b>	<b>9</b>
1.1 <i>Objetivo principal.....</i>	12
1.2 <i>Metas particulares.....</i>	12
<b>2    <b>Métodos de procesamiento automático de información semántica.....</b></b>	<b>13</b>
2.1 <i>Teorías semánticas modernas .....</i>	13
2.1.1   Teoría clásica del significado .....	14
2.1.2   Aproximación prototípica.....	22
2.1.3   Aproximación relacional .....	23
2.2 <i>Semántica léxica.....</i>	25
2.2.1   Aproximaciones basadas en diccionarios.....	37
2.3 <i>Polisemia desde el punto de vista computacional.....</i>	44
2.4 <i>Diccionarios computacionales .....</i>	48
<b>3    <b>Algoritmo para la selección automática de primitivas semánticas .....</b></b>	<b>53</b>
3.1 <i>Conceptos básicos .....</i>	53
3.2 <i>Modelo matemático y algoritmo principal.....</i>	56
3.3 <i>Implementación del algoritmo.....</i>	67
3.3.1   Módulo graph .....	68
3.3.2   Módulo circle .....	71
3.3.3   Módulo visualizer.....	72
3.3.4   Librería circle .....	73
3.3.5   Explicación del código. ....	73
3.3.5.1   Módulo graph .....	74
3.3.5.2   Módulo circle.....	90
3.3.5.3   Módulo visualizer .....	93
3.3.5.4   Librería circle.....	95

<b>4</b>	<b>Metodología experimental y resultados obtenidos.....</b>	<b>97</b>
4.1	<i>Metodología experimental.....</i>	<i>97</i>
4.2	<i>Resultados obtenidos.....</i>	<i>99</i>
<b>5</b>	<b>Conclusiones .....</b>	<b>109</b>
5.1	<i>Resultados obtenidos.....</i>	<i>110</i>
5.2	<i>Trabajo futuro .....</i>	<i>111</i>
<b>6</b>	<b>Bibliografía .....</b>	<b>113</b>
	<b>Anexo 1. Formato del archivo de entrada.....</b>	<b>119</b>
	<b>Anexo 2. Lista de las primitivas encontradas .....</b>	<b>126</b>
	<b>Anexo 3. Lista de los ciclos encontrados .....</b>	<b>131</b>
	<b>Anexo 4. Ejemplos de ciclos en el diccionario. ....</b>	<b>147</b>

## LISTA DE FIGURAS

Figura 1. El uso de notaciones lexicográficas para distinguir diferentes sentidos de una palabra.....	15
Figura 2. Representación del significado de la palabra "planta".....	26
Figura 3. Modelo de Miller y Johnson Laird.....	27
Figura 4. Relación tipo-subtipo.....	35
Figura 5. Parte de una red semántica.....	40
Figura 6. Definición de la palabra "corrido".....	54
Figura 7. Uno de los sentidos de la palabra "prolepsis".....	55
Figura 8. Algoritmo A (principal).....	59
Figura 9. Algoritmo B (depuración).....	61
Figura 10. Contraejemplo.....	62
Figura 11. Ejemplo: la palabra seleccionada como primitiva es "suministrar".....	65
Figura 12. Ejemplo: la palabra seleccionada como primitiva es "abastecer".....	66
Figura 13. Archivos de entrada y archivos de salidas del sistema.....	67
Figura 14. Módulo Graph (primera parte).....	69
Figura 15. Módulo Graph (continuación).....	69
Figura 16. Módulo Circle.....	72
Figura 17. Módulo Visualizer.....	73

## LISTA DE GRÁFICAS

Gráfica 1. Entradas totales contra entradas significativas.....	101
Gráfica 2. Resultados obtenidos por grafo y por método.....	101

## LISTA DE TABLAS

Tabla 1. Primitivas semánticas de Wierzbicka. ....	21
Tabla 2. Número de vértices en los grafos. ....	100
Tabla 3. Número de definidores, con diferentes algoritmos.....	100
Tabla 4. Longitudes de los ciclos más cortos (ejemplo).....	104
Tabla 5. Número de palabras que aparecen en los distintos conjuntos. ....	104
Tabla 6. Tipos de palabras del grafo depurado.....	105
Tabla 7. Tipos de palabras de cada conjunto de primitivas semánticas obtenidas.....	105
Tabla 8. Tipos de las 478 palabras que están en los 4 conjuntos obtenidos. ....	105
Tabla 9. Número de palabras del conjunto de primitivas comunes a los 4 conjuntos ordenados la longitud de uno de sus ciclos (LC). ....	106
Tabla 10. Algunos ejemplos de datos de palabras relacionadas y sus frecuencias. ....	107
Tabla 11. Algunas primitivas halladas por el método 4. ....	108

## Resumen

---

Esta tesis se sitúa en el campo de la lingüística computacional, también conocida como procesamiento del lenguaje natural. En esta tecnología, los diccionarios de un lenguaje específico, como el español, tienen una importancia crucial. Una de las formas para construir diccionarios semánticos es la conversión de diccionarios explicativos tradicionales. Sin embargo, un diccionario semántico para el razonamiento lógico no debe tener círculos viciosos en el sistema de sus definiciones, mientras que los diccionarios explicativos tradicionales presentan este problema con mucha frecuencia. En esta tesis, presentamos un algoritmo capaz de detectar dichos círculos y seleccionar un conjunto minimal de palabras a través de las cuales la definición del resto de las palabras puede realizarse, sin causar círculos viciosos. Para su aplicación práctica el número de elementos en un conjunto de este tipo deberá ser tan pequeño como sea posible, por lo que presentamos también algunas técnicas para crear un conjunto tan pequeño como dichas técnicas lo permitan (no se garantiza que éste es el conjunto más pequeño posible). Partiendo de este algoritmo, hemos desarrollado una herramienta para ayudar a los lexicógrafos a elegir las palabras a través de las cuales deseen definir todas las demás y corregir los defectos en el diccionario relacionados a los círculos en el sistema de definiciones. Esta tesis da una introducción a la semántica general y computacional en relación con la tarea mencionada, y usa los métodos de la teoría de grafos y de la ciencia de la computación.

## **Abstract**

---

This thesis is situated in the field of computational linguistic or automatic natural language processing. In this technology, dictionaries of a specific language, e.g., Spanish, have crucial importance. One of the possibilities to construct semantic dictionaries is to convert existing large traditional explanatory dictionaries. However, a semantic dictionary for logical reasoning should not have vicious circles in the system of its definitions, while the traditional explanatory dictionaries do present this problem very frequently. In this thesis, we present an algorithm for detecting such circles and selecting a minimal (though not the smallest possible) set of words through which the definition of the rest of the words can be done without causing the vicious circles. Since for practical applications the number of elements in such a set should be as small as possible, we present also the some techniques to find such a set as small as we can (though still not guaranteed to be the smallest one). Basing on this algorithm we have developed a tool to help lexicographers to choose such words and correct the defects in the dictionary related to the circles in the system of definitions. The thesis gives an introduction to general and computational semantics to the extent as the task described above is concerned, and uses the methods of graph theory and computer science.

# **1 Introducción**

---

El lenguaje natural es una de las características más importantes en el ser humano. A través de él logra la comunicación y la transmisión de conocimientos. La lingüística es la ciencia que se ocupa de los métodos de investigación y de las cuestiones comunes a las diversas lenguas; aborda su objeto de estudio desde distintos puntos de vista, como el temporal, el metodológico, el social, el cultural, el de aprendizaje, etc. Por otro lado, una buena parte del conocimiento humano se encuentra en forma de textos, por lo que hay una demanda de procesamiento a gran escala. La computadora es actualmente una herramienta indispensable para realizar el procesamiento de grandes volúmenes de datos. La lingüística computacional es la ciencia que trata de los modelos computacionales de la lingüística y de cómo hacer entender a la computadora el lenguaje natural.

La finalidad última de lingüística computacional es que las personas puedan acceder a todas las facilidades ofrecidas por las computadoras, mediante órdenes expresadas lo más libremente posible con el vocabulario y la sintaxis de su propia lengua y, al mismo tiempo, que las computadoras presenten los resultados de sus aplicaciones en ese mismo idioma de manera natural y comprensible para las personas.

Las líneas de trabajo dentro de la lingüística computacional son: la identificación automática del idioma, el reconocimiento y síntesis de habla, la comprensión y generación del lenguaje natural y la integración de información

visual en el procesamiento del lenguaje y del habla. Para facilitar la interacción persona-computadora, la lingüística computacional se ha centrado en el desarrollo de técnicas o sistemas de comprensión y de generación de lenguaje natural. En el procesamiento de texto las aplicaciones más importantes son: la generación automática de resúmenes, la clasificación automática de los documentos, la búsqueda inteligente en bases de datos y la traducción automática. Su relevancia se debe a la gran incidencia social que tienen los desarrollos en estos campos.

El lenguaje natural está constituido por cinco niveles: fonético/fonológico, morfológico, sintáctico, semántico y pragmático. Cada uno de estos entraña sus propios problemas teórico-metodológicos y los problemas de su manejo computacional. La semántica, por ejemplo, se encarga de entender la frase, es decir, entender el sentido de todas las palabras y dar una interpretación a las relaciones sintácticas. Dentro del análisis semántico existe el gran problema de la polisemia, es decir, diferentes sentidos para una misma palabra. Definir los sentidos de las palabras es una tarea muy ardua que se ha hecho conforme a ciertos principios, a través de una rama de la semántica llamada lexicografía. Los resultados de esas investigaciones son los diccionarios.

En los últimos años la semántica ha tenido un gran desarrollo. Una de las razones es la necesidad de crear modelos léxico-semánticos claros y modelos de análisis semánticos de la oración que pudieran mejorar los resultados obtenidos en el análisis sintáctico. Las aplicaciones de la lingüística computacional, en su intento por mejorar el procesamiento del lenguaje natural y manejar así las grandes masas de información escrita y oral que existen en formato electrónico, se han encontrado con el obstáculo de tratar los contenidos de los textos y no sólo su forma, por ejemplo, los sistemas de recuperación de información. La semántica léxica, junto con otras disciplinas, ha contribuido con herramientas teóricas y metodológicas para el desarrollo de la lingüística computacional, abriendo nuevas vías para el tratamiento de las lenguas naturales, y de hecho ha ocupado un lugar privilegiado en esta evolución. La tradición semántico-estructural, entre otras,

como la gramática generativa y sus estructuras de papeles semánticos y semántica de rasgos, está siendo aprovechada por la lingüística computacional para la elaboración de resúmenes automáticos, la clasificación de documentos y la traducción automática.

Dentro de las tareas en lingüística computacional ocupa un lugar preponderante la integración de diccionarios computacionales. Construir vocabularios definidores compuestos por primitivas semánticas, gracias a ciertos principios, es de gran importancia para contar con diccionarios computacionales que puedan aplicarse a aquellas tareas donde la semántica tiene un papel fundamental. La construcción de un diccionario computacional se puede hacer a partir de un diccionario tradicional, sin embargo ese tipo de diccionario presenta el problema de los ciclos, que se dan cuando una palabra tiene referencias a otras palabras y éstas a su vez a más palabras, dentro de las cuales puede llegarse de nuevo a la palabra de inicio. Por esta razón, es importante implementar algoritmos que, partiendo de un diccionario explicativo, puedan procesarlo para obtener un diccionario más adecuado para su uso computacional.

Este trabajo tiene como objetivo presentar los algoritmos correspondientes para la detección de ciclos y selección de un conjunto definidor minimal de palabras con la función de primitivas semánticas a través de varios métodos. Como marco teórico presento los métodos de procesamiento semántico, algunas de las corrientes más importantes en relación con las teorías semánticas (como la teoría de las primitivas semánticas); posteriormente presento el modelo matemático de un diccionario (representado como grafo), una descripción del algoritmo principal para la selección automática de las primitivas semánticas, un algoritmo de depuración de la red semántica, la implementación de dichos algoritmo, la descripción de los experimentos realizados y la descripción e interpretación de los resultados.

## **1.1 Objetivo principal**

Procesar un diccionario tradicional en español para detectar ciclos en las definiciones y, a través de distintos métodos, seleccionar conjuntos minimales de palabras que puedan integrar un vocabulario definidor.

## **1.2 Metas particulares**

- La construcción de una red semántica a partir de un diccionario tradicional.
- La depuración inicial de dicha red semántica.
- El diseño del algoritmo que encuentre un conjunto P minimal de palabras.
- La aplicación de diversos métodos para obtener varios conjuntos minimales, justificar su aplicación y encontrar el conjunto más pequeño posible, dados los métodos usados.

## 2 Métodos de procesamiento automático de información semántica

---

### 2.1 Teorías semánticas modernas

Existen distintas teorías semánticas modernas. Algunas de ellas están relacionadas con la aproximación clásica, otras más se inscriben dentro de la aproximación prototípica o bien de la aproximación relacional. Dado que la existencia de significados distintos para una misma palabra es un fenómeno que complica la descripción semántica y el uso de las palabras dentro de un contexto, la polisemia es un factor que es considerado y manejado por la mayoría de las teorías semánticas. En este inciso veremos algunas teorías semánticas modernas y su relación con la polisemia.

"La semántica relaciona al mundo extralingüístico con expresiones lingüísticas que lo describen"<sup>1</sup>. Una expresión léxica se aplica a un tipo extralingüístico a través de un proceso de abstracción; dicha abstracción es capturada en las teorías semánticas por medio de dos principios: la generalización, la cual reduce la polisemia para incrementar el poder explicativo de la teoría o bien

---

<sup>1</sup> Ravin, Y. and Leacock, C. (2000) *Polysemy: An Overview*, p. 11.

el incrementar la polisemia, haciendo distinciones, para reunir el mayor número de detalles semánticos posibles.

### **2.1.1 Teoría clásica del significado**

Para los clásicos o "Aristotelianos" una entidad individual es un miembro de una categoría conceptual. Cada categoría se define por 1) condiciones necesarias o características definitorias y 2) condiciones suficientes o propiedades nucleares.

Las categorías conceptuales pueden ser vistas como un conjunto de listas de características conectadas por operadores lógicos. Estas categorías se organizan jerárquicamente, donde conceptos del mismo nivel de la jerarquía heredan y comparten las propiedades esenciales o nucleares de los conceptos que están por arriba de ellos, pero poseen un conjunto de características que son mutuamente exclusivas. Un objeto pertenece o no pertenece a una categoría conceptual; el problema que se presenta con los objetos que son ambiguos se soluciona eligiendo "arbitrariamente" alguna de las categorías posibles.

Esta forma de entender la semántica es retomada hoy en día por numerosos investigadores, gracias a su enfoque jerárquico y clasificatorio, y por el hecho de que se fundamenta en la lógica. Una corriente moderna que puede inscribirse dentro de la aproximación clásica es la de Anna Wierzbicka (*Natural Semantic Metalanguage - NSM*) (Wierzbicka 1996), quién indica que el problema del análisis semántico no se encuentra en hacerlo en función de características necesarias y suficientes, sino en que implícitamente se crea que existe una correspondencia observable y mensurable de éstas con aspectos certeros de la realidad externa. Más adelante en esta misma sección se mencionará más sobre el trabajo recién citado.

Tradicionalmente se ha visto a la polisemia como diversidad de significados para una misma palabra, a la sinonimia como un mismo significado para varias palabras. Por otro lado, la homonimia se refiere a varios significados

para una misma palabra, la cual sin embargo proviene etimológicamente de orígenes distintos y por lo tanto los sentidos y los dominios en el que se emplean la palabra también son muy distintos. Como ejemplo de ello tenemos en español la palabra "estela", con dos sentidos diversos, uno de los cuales hace referencia al rastro que deja un objeto al atravesar un fluido, y el otro, a un monumento monolítico de carácter conmemorativo. El primer sentido deriva el latín *aestuaría*, mientras que el segundo proviene del griego *stèle*. Un ejemplo de polisemia es la palabra "brillante" (del verbo brillar, éste último del italiano *brillare*), que puede tener varios sentidos, como adjetivo significa que brilla, o que es admirable o sobresaliente. Tiene otro sentido más, que es, por supuesto, el de un diamante tallado. Todos estos sentidos tienen cierto parecido semántico. Se podría decir, según la aproximación de Katz (1972), que comparten marcadores, esto se debe a que provienen etimológicamente de la misma palabra (el concepto de marcadores se explicará mas adelante). Cada teoría del significado caracteriza y maneja a la polisemia en sus propios términos.

Homonimia			Polisemia		
estela	I. Del lat. <i>aestuaría</i> .	1. Rastro que deja un objeto al atravesar un fluido.	brillante	I. Del ital. <i>brillare</i> .	1. Que brilla
	II. Del gr. <i>stèle</i> .	1. Monumento monolítico de carácter conmemorativo.		2. Admirable o sobresaliente.	
					3. Diamante tallado por sus dos caras

**Figura 1. El uso de notaciones lexicográficas para distinguir diferentes sentidos de una palabra.**

Como se mencionó en el párrafo anterior, una corriente moderna de la teoría clásica está en la teoría intencionalista y racionalista de la semántica, desarrollada por Katz y Fodor (1963) y más adelante por Katz (1972). Ellos parten de la teoría estándar de Chomsky y su objetivo es articular los principios de una teoría semántica universal que explicaría el sistema de reglas internalizadas que constituyen la habilidad o capacidad lingüística del hablante nativo. Semánticamente, esto se refiere a las reglas que relacionan las estructuras lingüísticas del mismo.

Katz (1972) comienza el debate haciendo a un lado la discusión ontológica sobre el significado. Para él, la meta de una teoría semántica es hallar un sistema general para representar el significado. Este esquema provee un mecanismo para definir cómo la forma lógica o significado de proposiciones u oraciones está definido por los sentidos de sus componentes, y de las relaciones sintácticas entre esos componentes.

Una teoría semántica provee una representación semántica para cada sentido de cada elemento léxico o palabra. Las palabras ambiguas se representan en un diccionario con más de un sentido o representación de sentidos.

Las representaciones semánticas de sentidos tienen unidades más pequeñas llamadas marcadores semánticos, que es algo más o menos equivalente a las propiedades nucleares y definitorias, y reflejan los conceptos componentes que constituyen un sentido particular.

Los marcadores semánticos no son necesariamente primitivos, esto es, pueden subdividirse en otros marcadores semánticos.

Los marcadores semánticos permiten inferencia, pues al heredarse la serie de características que constituyen una clase particular, se puede inferir que una instancia de una clase en especial también es un miembro de la clase situada en un nivel superior, pues todos sus marcadores semánticos se encuentran también en

dicha clase. Ejemplo: En la oración "Una silla está en la habitación" se deriva que "Un objeto está en la habitación" como una inferencia válida

Sobre este concepto también es conveniente señalar lo siguiente:

- Los sentidos relacionados, se refieran o no de la misma palabra, comparten marcadores semánticos (uno o más).
- Existe una medida llamada similitud semántica que está en función del grado en que se comparten marcadores semánticos.
- La sinonimia es aquella en la que todos los marcadores semánticos se comparten.
- La ambigüedad se da cuando una palabra tiene más de una representación semántica.

Katz no distingue polisemia de homonimia, su objetivo no es estudiar el origen de los diferentes sentidos, tampoco explica lo que es la polisemia regular, fenómeno a través de cual un elemento léxico con una representación de sentido adquiere otra representación que difiere de la primera de maneras predecibles. Sin embargo otras teorías clásicas sí dan cuenta de la misma.

Apresjan define la polisemia como la similitud en las representaciones de los sentidos de una palabra, basta que cada significado esté ligado por lo menos a otro, no es necesario que en todos exista una parte común. Apresjan estudió el caso en que las palabras son sistemáticamente ambiguas, como el caso en inglés de la frase "*the construction is complete*", pues esta frase puede significar o bien que el objeto físico construido está terminado o bien que la jornada de trabajo en el espacio físico donde se realiza esta acción ha concluido. En español podríamos dar el ejemplo de la frase "en la reunión tocaron música", pues también tiene dos sentidos: que unas personas realizaron la acción de tocar instrumentos o bien que a través de un aparato reproductor, se escuchó música en dicha reunión.

La polisemia regular está gobernada por procesos que son productivos, muy similares a los de formación de palabras, que se desarrollan conforme a reglas, y que son procesos predecibles. Ejemplos de estos procesos son: la metonimia (responsable de que surjan nuevos sentidos como sucedió con la palabra "pie" en "al pie de la montaña") o la relación entre el nombre de un contenedor y la cantidad que se le adjudica a dicho contenedor, como en "Una taza de azúcar".

Aunque el fenómeno de polisemia regular no es un problema mayor dentro de la aproximación clásica, la distinción entre sentidos sí lo es. La teoría clásica postula que con cada diferencia conceptual se tiene un nuevo sentido, pero esto implica el peligro de la proliferación infinita de sentidos. Katz acota este peligro indicando que el significado debe quitar las características variables de las cosas y enfocarse sólo en las características invariables por virtud de las cuales una cosa es *cierta* cosa y no otra. De otra manera ninguna palabra se podría volver a usar con el mismo significado.

Esta afirmación es importante porque permite la abstracción o indeterminación, aún dentro de la aproximación clásica, en la construcción de representaciones semánticas. Las otras teorías no clásicas llevan esta idea de indeterminación mucho más allá, haciéndose preguntas de este tipo: ¿porqué el implemento para comer no forma parte del significado de "comer", ¿porqué no es esencial en muchos casos y en algunos sí?. Ravin investiga en este sentido, aplicándolo a los verbos; para él no hay un criterio claro para decidir que aspectos del mundo real son relevantes en la semántica particular de un verbo. Sin embargo, Ravin propone una metodología, la cuál se basa en la idea de que, en las representaciones semánticas, deben reunirse todas las propiedades semánticas y las relaciones de expresiones lingüísticas, por ejemplo "*break into pieces*" y "*shatter*", con significados idénticos, deben tener representaciones idénticas, estableciendo restricciones en la representación de las palabras constituyentes "*break*", "*into*",

"*pieces*". Representaciones idénticas deben tener también, es español, la palabra "guaquero" y la expresión "buscador de tesoros".

El metalenguaje semántico natural (NSM - *Natural Semantic Metalanguage*) es una teoría desarrollada por Anna Wierzbicka y más tarde también por Cliff Goddard (1999). NSM se basa en la evidencia de que existe un conjunto pequeño de significados universales y básicos, llamados primitivas semánticas, que pueden encontrarse como palabras u otras expresiones lingüísticas en todos los lenguajes. NSM incluye ciertos principios y un método para construir un vocabulario de primitivas semánticas.

La metodología que propone Wierzbicka para distinguir una palabra polisémica de la que no lo es e identificar su significado, consiste en lo siguiente: se asume que hay un solo significado, construyendo su definición, y después esa definición se compara con el uso de esa palabra en el mundo real. Si es necesario (ya que esta palabra no funcione para algunos casos de uso común) se postula otro significado, y así sucesivamente.

Ella usa la idea intuitiva de que la representación semántica de una palabra puede ser sustituida por la palabra en todos sus usos (intercambiabilidad definición-palabra). El problema de esto es que muchas veces es difícil obtener las intuiciones semánticas de los hablantes de una lengua y que no hay pruebas objetivas confiables que ayuden en esta labor.

Uno de los principios de NSM es la paráfrasis reductora. La paráfrasis reductora (*reductive paraphrase*) consiste en describir un significado complejo en términos de otros más simples. Para consignar el significado de una palabra compleja semánticamente, se dará una paráfrasis formada por otras palabras más "simples" y más fáciles de entender que la original, por ejemplo:

**X ama a Y =**

X frecuentemente piensa en Y

X piensa cosas buenas sobre Y

X quiere hacer cosas buenas por Y

X quiere que cosas buenas le sucedan a Y

Cuando X piensa en Y, X frecuentemente quiere estar con Y

Cuando X piensa en Y, X frecuentemente siente algo bueno

Según esta teoría, cada lenguaje tiene un conjunto de primitivas que es irreductible y del cual hacen uso los hablantes para entender todo pensamiento complejo. Este conjunto tiene la característica de ser el mismo en todos los idiomas, pues es un reflejo del pensamiento humano. Cada concepto específico a una cultura puede traducirse a una configuración de primitivas semánticas

Las primitivas semánticas y sus principios de combinación serían como un mini-lenguaje con la misma expresividad que el lenguaje natural. El modelo actual de primitivas semánticas de AnnaWierzbicka(1996) <sup>2</sup> es el siguiente:

---

<b>Sustantivos</b>	YO, TÚ, ALGUIEN, GENTE, ALGO, CUERPO
<b>Determinadores</b>	ESTO, LO MISMO, OTRO
<b>Cuantificadores</b>	UNO, DOS, ALGO, TODO, MUCHO/MUY
<b>Evalúadores</b>	BUENO, MALO
<b>Descriptorios</b>	GRANDE, PEQUEÑO, (LARGO)
<b>Predicados mentales</b>	PENSAR, SABER, QUERER, SENTIR, VER, ESCUCHAR
<b>Habla (discurso)</b>	DECIR, PALABRA, CIERTO
<b>Acciones, eventos y movimientos</b>	HACER, PASAR, MOVER, (TOCAR)
<b>Existencia y posesión</b>	HAY, TENER
<b>Vida y muerte</b>	VIVIR, MORIR
<b>Tiempo</b>	CUANDO/TIEMPO, AHORA, INICIO, DESPUÉS, UN TIEMPO GRANDE, UN TIEMPO CORTO, POR ALGÚN TIEMPO, (MOMENTO)
<b>Espacio</b>	DONDE/LUGAR, AQUÍ, SOBRE, BAJO, LEJOS, CERCA, LADO, DENTRO

---

<sup>2</sup> Esta tabla tiene las traducciones de las palabras que esta investigadora determinó a partir del idioma inglés. Se realizó una traducción al español con las imprecisiones que esto conyeva.

Lógica	NO, MAY BE, PORQUE, SI
Intensificador, aumentador	TODO, MÁS
Taxonomía, "partonomía" <sup>3</sup>	TIPO DE, PARTE DE
Similaridad	COMO

---

**Tabla 1. Primitivas semánticas de Wierzbicka.**

Como se puede ver en esta tabla, Wierzbicka ha encontrado aproximadamente 60 primitivas semánticas.

Algunas observaciones que se desprenden respecto de esta lista de primitivas (mismas que la autora ha reconocido) son:

- Una lista de primitivas no es suficiente pues hay términos que también tienen múltiples sentidos.
- Una caracterización completa incluye los "contextos canónicos" (conjunto de oraciones o partes de oraciones ejemplificando contextos gramaticales para cada primitiva).
- Las primitivas pueden ser *frasemas*, por ejemplo: "un tiempo grande".
- Las primitivas no siempre son morfológicamente simples (en inglés: *someone, inside*)
- Pueden ser alomorfos (por ejemplo, en inglés *thing-something*)

Como ya se mencionó en la sección 2.1.1 los *scripts* son convenciones locales de discurso que se expresan a través del metalenguaje de primitivas semánticas inventado por Wierzbicka, por ejemplo, en japonés:

*Las personas piensan que:*

*Si algo malo pasó porque yo hice algo.*

*Yo tengo que decir algo a alguien porque yo hice algo.*

*"Yo siento algo malo".*

*Yo tengo que hacer algo debido a esto.*

Este *script* representa la necesidad de disculparse que está presente en la cultura japonesa y la cual es distinta en el mundo occidental.

Existen diccionarios que ya usan un vocabulario definidor, aunque hasta el momento no hay ninguno cuyo conjunto definidor esté formado por "primitivas semánticas", con el sentido que Wierzbicka les da y la metodología que ha aplicado. Un ejemplo es el vocabulario definidor del *Longman Dictionary*, el cual posee más de 2000 palabras.

En núcleo de primitivas semánticas puede ser usado como una herramienta para la lingüística y para el análisis cultural: explica palabras complejas, palabras específicas a una cultura y construcciones gramaticales, y articula valores específicos de una cultura y actitudes o *scripts* culturales<sup>4</sup> en términos que son claros y traducibles. La teoría también provee las bases semánticas para una gramática universal y para la tipología lingüística. Tiene aplicación en la comunicación, lexicografía, enseñanza de un idioma y el estudio de adquisición de lenguaje en la infancia, semántica legal, etc.

### **2.1.2 Aproximación prototípica**

En un principio, Wittgenstein (1958), refiriéndose a los sentidos de las palabras, señaló que cualquier categoría tiene fronteras difusas y los significados exhiben reminiscencias familiares con elementos comunes que se superponen y cruzan de la misma forma que los rasgos familiares aparecen en los miembros de cada familia de personas. La aproximación prototípica está más cercana a la psicología

---

<sup>3</sup> En inglés *partonomy*, aludiendo a la acción de dividir en partes algo.

que a la lógica y a la filosofía (como la aproximación clásica). Rosch, en la década de 1960, demostró que las personas no categorizan objetos en la base de condiciones necesarias y suficientes sino en la base de la reminiscencia de los objetos a un miembro prototípico de la categoría.

Esta corriente respeta la jerarquía entre conceptos pero ve la membresía a una categoría como un asunto de grados, algo relativo, que se mide respecto a un "prototipo" o tipo ideal.

Según Fillmore (1982), un sentido está formado por la disyunción de otros sentidos; cuando aparecen todos los sentidos en una oración, el sentido del referente es más prototípico, si no se cumple ninguna condición, el tratamiento es anómalo. Ejemplo:

La palabra "escalar" está compuesta por la disyunción: "subir" ó "trepar". En la oración "Juan escaló la montaña" el uso de "escalar" es prototípico (se cumplen ambas condiciones, es decir, tanto "subir" como "trepar" son aplicables). En "Juan escaló hacia abajo la montaña" el sentido "subir" no es aplicable, por lo que el uso no es totalmente prototípico. En la frase "La serpiente escaló el árbol" el sentido "trepar" no es aplicable, por lo tanto, el uso no es prototípico.

### **2.1.3 Aproximación relacional**

Según esta aproximación las palabras están organizadas acorde a sus significados usando relaciones semánticas o enlaces para formar redes semánticas. Hace explícita la organización estructural que está implícita en otros modelos y describe cómo elementos de un dominio están relacionados entre sí. La aproximación relacional trata de ubicar cada palabra del lexicón o diccionario, en un lugar

---

<sup>4</sup> Un *script* cultural es una forma de llamar a distintas convenciones de discurso "locales" usando el metalenguaje de primitivas semánticas inventado por Wierzbicka.

particular de esta red, dependiendo del dominio y conforme a cierto tipo de relaciones, sobre todo la relación de sinonimia, antonimia, hyponimia (relación jerárquica - 'is\_a') y troponimia (se da entre verbos, con variedades de un mismo verbo, como la relación entre "trotar" y "correr"). Un autor importante de esta corriente es Fellbaum (1990, 1998).

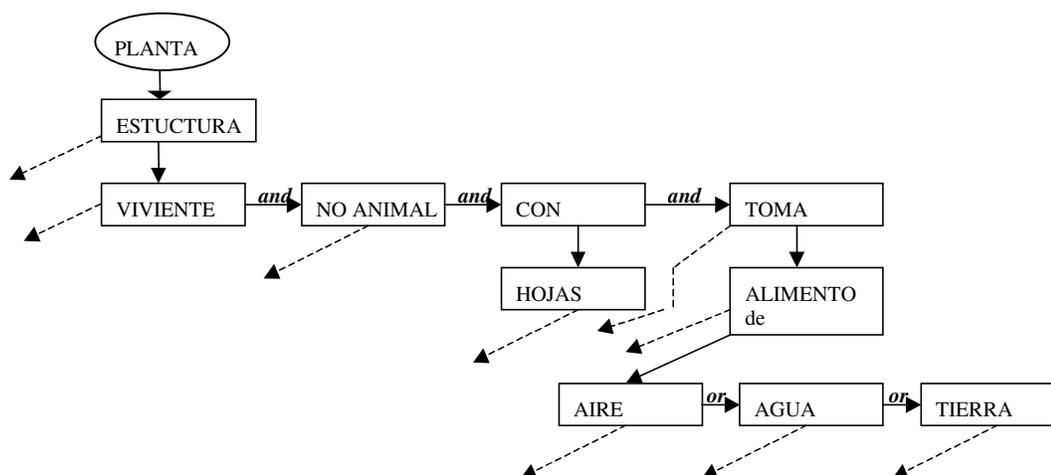
La representación de la polisemia en esta corriente es problemática, pues sentidos con polisemia regular (ya mencionada en el apartado 3.2), pueden quedar alejados en la red semántica conceptual. Mel'cuk y Zholkovsky (1988) usan índices para establecer relaciones de polisemia regular o no-regular usando una relación especial de similitud.

## **2.2 Semántica léxica**

Actualmente la semántica léxica juega un papel muy importante en la lingüística computacional, junto con los formalismos gramaticales para el análisis y la generación del lenguaje, así como la producción de una representación semántica del discurso y del enunciado. En este apartado veremos algunos modelos desarrollados desde la perspectiva psicolingüística y lingüística.

El papel central de la semántica léxica en la lingüística computacional puede explicarse por el hecho de que cada entrada léxica contiene una parte considerable de información relacionada al sentido de la palabra que representa. El campo del procesamiento del lenguaje natural se ha apoyado en algunos conceptos de la psicolingüística, sobre todo en aquellos aspectos relacionados a la forma en que se organizan los conceptos en la memoria, y también a la forma en que esta información es recuperada del "lexicón mental". La investigación en psicolingüística tiene mucha influencia en relación con los modelos de representación del conocimiento para los sistemas de inteligencia artificial. Los seres humanos procesamos lexemas para nombrar objetos, elaborar pensamientos y reconocer palabras escritas, escuchadas o pronunciadas. Según Collins y Quillian (1972), a través de su modelo de organización de la memoria, los conceptos forman parte de una red de nodos "tipo" y nodos de "ocurrencia", y las relaciones que los asocian (*and, or*), por ejemplo, según un diccionario cualquiera, la definición de la palabra "planta" podría ser la siguiente:

*Planta: Estructura viviente que no es un animal, que con frecuencia tiene hojas, toma su alimento del aire, del agua y de la tierra.*



**Figura 2. Representación del significado de la palabra "planta".<sup>5</sup>**

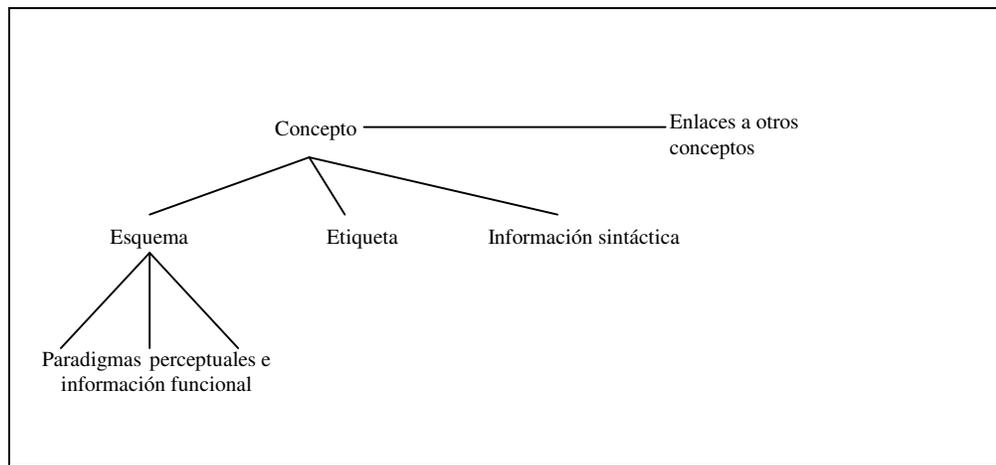
Básicamente una red semántica es un grafo con nodos que representan conceptos, y aristas que son relaciones binarias entre los nodos. En este modelo los conceptos no son primitivos. El modelo de Collins y Quillian (1972) está centrado en un almacenamiento de memoria a largo plazo, en otras palabras, en "conocimiento enciclopédico". Otros modelos distinguen entre memoria a largo plazo y memoria de trabajo o a corto plazo. Existen modelos que no reconocen que exista una diferencia entre ambas. Para Collins y Quillian, el modelo debe basarse en la economía de espacio, mientras que otros modelos (Conrad, 1972) prefieren tener como principio la eficiencia, a través de la flexibilidad de acceso, aunque haya redundancia en la información.

El modelo de Miller y Johnson Laird (1976) destaca el papel de la percepción. En su modelo la percepción es una actividad estructurada que destaca características distintivas de un objeto; a la representación de la información

<sup>5</sup> Los nodos tipo se representan con elipses, los nodos de ocurrencia con rectángulos, las relaciones asociativas con flechas o arcos entre los nodos. Cada nodo de ocurrencia se relaciona con un nodo tipo, en este diagrama estas relaciones aparecen con flechas punteadas.

relevante asociada a un concepto la llaman esquema. El nivel conceptual es donde convergen los esquemas perceptuales y los elementos léxicos. Un concepto léxico consiste en una etiqueta (cualquier palabra de un diccionario), un esquema, y las reglas que gobiernan el comportamiento sintáctico de la etiqueta.

El esquema integra información funcional y perceptual y podría incluir información almacenada que no tiene ninguna consecuencia perceptual directa, como se ve en la Figura 3:



**Figura 3. Modelo de Miller y Johnson Laird.**

El significado lingüístico de los elementos léxicos se deriva de las relaciones que pueden darse entre ellos, comúnmente: IS-A (es un(a)), HAS-A(tiene un(a)), IS-IN(esta en).

Según estos autores, un lexicón mental tiene dos componentes:

1. Un componente léxico como su inclusión en una relación de clase todo-parte
2. Un componente conceptual, que se refiere a las características perceptuales, información sintáctica mínima, información "mnésica" (emoción, funcionalidad, etc.).

Según Allport (1985), las investigaciones recientes se inclinan más a no establecer una separación entre la memoria a corto plazo y la memoria a largo plazo, y tampoco a considerar que la percepción es la entrada a una "caja negra" de la que se obtendrá la memoria/conocimiento. Contrariamente a esto, se inclinan más a considerar a la memoria como una capacidad que interviene en cada nivel de actividad de la "caja negra", y que la percepción y la memoria/conocimiento están interrelacionadas. La memoria intervendría así en aspectos sensoriales, motores, y en aspectos cognitivos más elevados.

Actualmente se cree que el acceso a la memoria es distinto dependiendo si se trata de reconocimiento o de producción.

El acceso léxico se refiere a cómo reconocemos que un elemento léxico es simple o compuesto y de qué manera obtenemos esta información del sitio donde está almacenada. Uno de los aspectos que se toman en cuenta es que no solo tenemos acceso a las palabras individuales, sino también a las relaciones entre ellas, en términos de sintaxis y del significado del enunciado.

Todos estos modelos tienen en común la búsqueda de un sistema para detectar palabras. En este sentido se parte de dos hipótesis, una en la que las palabras se recuperan siguiendo un procesamiento secuencial (como se hace a través de un diccionario), modelo de Forster (1976), y otra en la que el procesamiento es paralelo y se realiza gracias a la activación de palabras dentro de un tratamiento interactivo del lenguaje (Morton, 1982; Marslen-Wilson, 1984).

Forster también da algunas claves de cómo estructurar lexicones a través de su "*semantic priming*", esto es, que el tiempo de reacción para el reconocimiento de estímulos es menor, cuando un elemento léxico ha sido presentado después de otro al que está relacionado, semántica y contextualmente, es decir, las personas reconocen más rápido la palabra "enfermera" después de "doctor", que la palabra "abogado".

Morton (1982) ha desarrollado el modelo "logonen", en el que define a los *logonens* como unidades relacionadas a los elementos léxicos y que permanecen activados durante el proceso de recuperación. Morton sostiene que un sistema de reconocimiento de palabras, para ser operacional, necesita de toda la información del contexto y que la recuperación de los elementos debe hacerse en paralelo.

Marslen-Wilson (1984) desarrollaron el modelo "cohort", un sistema en que la identificación de palabras no es resultado de la activación selectiva de *logonens*, sino de la eliminación progresiva de ellos. Para Marslen -Wilson la elaboración de significado inicia tan pronto como una oración comienza.

En términos de la producción, el acceso léxico depende de cómo asociamos un objeto dado o un evento con la etiqueta léxica apropiada, y cómo funciona el proceso de recuperación léxica dentro del contexto de la producción de expresiones verbales.

El estudio de la producción se ha hecho basándose en observaciones, no en experimentos, dado que esto es inaccesible, por lo que hoy se ha hecho más fácil abordarlo observando los errores que cometen los hablantes. Estos estudios han llevado a la teorizar alrededor de dos componentes de la producción: el "qué decir" y el "cómo decir". Respecto al primero -"qué decir"-, este componente involucra la planeación a un nivel pre-lingüístico, donde el hablante decide, acorde a la representación que tiene de su co-hablante y del mensaje que quiere comunicar, cómo elegir y ordenar su "qué decir".

Garrett(1982) analizó errores en hablantes normales y concluyó que hay dos niveles de planeación: el funcional, en el que el elemento se funda en una forma abstracta, alcanzando solo la información gramatical y semántica, y el nivel posicional, donde los morfemas gramaticales y fonológicos son insertados y donde se provee el orden de la producción.

Butterworth(1980) analizó pausas en hablantes espontáneos y afirma que la frecuencia de las pausas depende de la complejidad de la planeación del contenido semántico del discurso, la programación semántica no es inconsciente, está en parte controlada; también asegura que la frecuencia de las pausas no se afecta por la complejidad sintáctica, lo cual indica que la programación sintáctica involucra procesos automáticos, por último, dice que la frecuencia de las pausas es alta durante la selección léxica de palabras las cuales tienen un alto grado de incertidumbre.

El resultado de estas investigaciones puede aplicarse a la perspectiva computacional en términos de heurísticas.

El segundo componente ("cómo decir") se refiere a la traducción del lexicon mental en una expresión verbal para la que se requiere haber tomado varias decisiones: la elección de la palabra, el orden de la enunciación y el acto de habla.

Para un contenido semántico dado hay muchas realizaciones lingüísticas. Hay varios parámetros que tomar en consideración:

- El efecto automático al usar una palabra que tenemos en la mente, esto se relaciona a la frecuencia de uso.
- El efecto preparatorio, que selecciona una palabra pre-activada por el contexto, por ejemplo, las personas seleccionamos con mayor facilidad la palabra "cliente" en el contexto de un consorcio de abogados, que la palabra "paciente".
- La elección del orden de la enunciación y del acto de habla determinado, esto es, si es una afirmación o una pregunta, si usar o no modalidades, como "Creo que..." y el orden de la argumentación.

La psicolingüística no ha llegado a una conclusión sobre la recuperación de información, como un proceso serial o paralelo, pues no se sabe si el reconocimiento de las palabras es un proceso controlado o bien tenemos un acceso directo y automático a los elementos léxicos. Los experimentos con estímulos visuales indican que el contexto no rige la existencia del nivel automático de reconocimiento de palabras, que es indiferente a la información contextual. Marslen-Wilson no están de acuerdo con esto en lo que toca a los actos de habla. Además, el acceso a distintos componentes de la expresión (por ejemplo a los niveles fonológico, léxico y sintáctico) no se hace en forma serial y dirigida, en la que cada componente es autónomo, recibiendo información del nivel superior y dando información al siguiente nivel más abajo. En lugar de esto, los componentes son accedidos en paralelo, con una parte desarrollada en un modo dirigido por el concepto, con guía de análisis semántico, por ejemplo, análisis fonológico y sintáctico.

Para Rumelhart, et al. (1986) tanto el tratamiento automático como el tratamiento controlado son necesarios. Distinguen como procesos automáticos a los que tardan menos de 0.25 a 0.5 segundos. Éstos, dicen, son paralelos y su modelo tendrá que ser paralelo. Los más lentos, tienen un componente serial y su modelo será de modelo de procesamiento secuencial.

Como vemos, el lexicón mental es esencialmente subjetivo (producto de la experiencia individual) y se maneja por el principio de la eficiencia, más que por el de economía (con trayectorias de acceso múltiple que permiten gran flexibilidad y la coherencia de su buen funcionamiento), por lo que aún ningún sistema propuesto por los psicolingüistas representa una solución aplicable y completamente satisfactoria computacionalmente.

La forma en que los lexicones están estructurados y organizados tiene un alto impacto en las formas en que deben ser accedidos. Una de las teorías que abordan este aspecto es la semántica componencial de Katz y Fodor. Esta teoría

fue inaugurada por Hjelmslev, y más adelante desarrollada por Katz y Fodor. Según ésta, una palabra se identifica por:

- marcadores sintácticos
- marcadores semánticos
- diferenciadores semánticos
- delimitación sintáctica o semántica contextual.

El significado de un ítem léxico se analiza en términos de átomos de sentidos (elementos terminales de la estructura de red).

Otros trabajos han añadido la noción de prototipo que determina a las palabras más típicas en una familia de palabras, por ejemplo "manzana" y "naranja" en la familia "frutas", también se refiere a encontrar los atributos prototípicos de la palabra, como por ejemplo, de "casa" serían: extensión del terreno, número de pisos, número de habitaciones, funcionalidad, ubicación, precio, etc. Asimismo, estos atributos pueden tener valores prototípicos, como número de pisos, en "casa" es prototípicamente entre 1 y 4 (si fueran más pisos estaríamos hablando de otro tipo de construcción).

Otra teoría que se refiere a la estructuración y organización de los lexicones es la desarrollada por Miller y Johnson-Laird (1976). Ésta consiste en definir el significado de una palabra por los procedimientos involucrados en su uso. La palabra no se considera una entidad estática (un conjunto de características o proposiciones), sino como una subrutina la cual es llamada por el programa en turno que esté "ejecutándose", ya se encuentre en su fase de producción o reconocimiento. El sentido de "oración" o frase se relaciona al de "programa de computadora". El modelo toma en cuenta la producción y el reconocimiento usando el mismo conjunto de reglas, lo cual es ciertamente una ventaja para los sistemas de inteligencia artificial. Sin embargo esto no los hace más apegados a la

investigación psicolingüística, al contrario. Un ejemplo es el siguiente: la expresión: "¿trajo Lucy el postre?" según Miller y Johnson-Laird, generaría un programa cuyos pasos son los siguientes:

- encontrar en memoria el episodio donde  $F(x, y)$  existe.
- asignar a  $F$  el valor "traer", asignar a  $x$  "Lucy" y asignar a  $y$  "postre".
- Si la descripción se encuentra, la respuesta es "Sí"; si nada se encuentra, la respuesta es "No lo sé"; si alguna información contradice la descripción, la respuesta es "No".

Desde el punto de vista léxico, "traer" puede llamar a una subrutina la cual consiste en:

- encontrar en el dominio a consideración un evento  $e$  donde  $venir(x)$  exista.
- verificar si, durante  $e$ ,  $traer(x, y)$  exista.

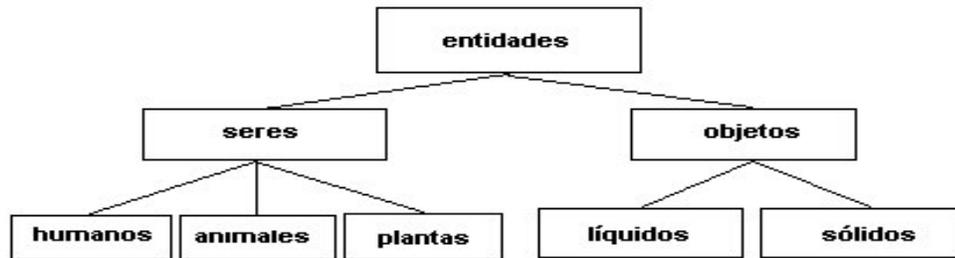
En esta perspectiva una palabra no está nunca aislada y no es independiente del contexto, y su significado se ve como una subrutina dentro de un programa. Sin embargo se ha dicho muy poco sobre la semántica y la sintaxis de las palabras.

La aplicación de la semántica léxica en la lingüística actualmente está relacionada principalmente con: la estructura de argumento, los papeles temáticos, las restricciones selectivas, la semántica léxica, las relaciones esenciales, las ontologías, la estructura conceptual léxica, la estructura de eventos y la estructura *qualia*.

La **estructura de argumento** es una estructura que representa conceptos a través de predicados y un número de argumentos. Los argumentos representan los



técnicas de representación del conocimiento de la Inteligencia Artificial. El principio de esta técnica es asociar cada argumento de un predicado a una lista de características  $F_i$  que expresan restricciones en la naturaleza semántica de los argumentos.  $F_i$  podría ser una simple característica, como "humano", "animal", etc., mientras que  $(F_1 \wedge F_2 \dots \wedge F_n)$  es una conjunción de restricciones, y  $(F_1 \vee F_2 \dots \vee F_3)$  es cuando alguna restricción debe cumplirse. También se hacen combinaciones de estas formas. Las restricciones selectivas se basan frecuentemente en bases empíricas, en los últimos años se han tratado de definir principios para formarlas, una de estas orientaciones organiza los tipos semánticos usados como restricciones selectivas a través de un árbol de tipos. Estos tipos están estructurados a través de una relación tipo-subtipo, por ejemplo, un humano es un subtipo de una entidad animada.



**Figura 4. Relación tipo-subtipo.**

Las **relaciones semánticas léxicas** permiten hacer un análisis de las palabras en función a la relación entre las palabras. Las relaciones de congruencia son: identidad (cuando las entidades  $E_1$  y  $E_2$  son iguales:  $E_1 = E_2$ ), de inclusión ( $E_2$  está incluida en  $E_1$ ), la superposición (cuando  $E_1$  y  $E_2$  tienen una intersección no vacía, pero no son subconjuntos alguno del otro) y la disyunción (no tienen elementos en común  $E_1$  y  $E_2$ ). También pueden estar relacionadas

jerárquicamente, las relaciones de este tipo son: la taxonomía (*'is\_a\_kind\_of'*, *'is\_a\_way\_of'*), las meronimias (*'has\_a'*, *'is\_part\_of'*) y las series proposicionales.

También hay relaciones no-jerárquicas como la sinonimia, la antonimia y los opuestos.

Una **ontología** es un sistema formal que permite la representación de un dominio dado a través de sus elementos básicos, los distintos conceptos y sus realizaciones lingüísticas relacionadas. Una mirada extensa hacia el conocimiento ontológico puede incluir varias formas de conocimiento enciclopédico sobre el dominio, a la vez que conocimiento de sentido común, y conocimiento retórico y metafórico, así como expresiones. Las descripciones ontológicas son usadas en múltiples situaciones y clases de aplicaciones. En términos de la aplicación, algunos sistemas incorporan conocimiento ontológico para mejorar su calidad y generalidad. Este tipo de conocimiento puede ser crucial para interfaces de usuario inteligentes de lenguaje natural y para la recuperación automática de documentos, aplicándose mucho a las redes semánticas, a los grafos conceptuales, a la estructura conceptual léxica, a los léxicos generativos y a los sistemas de administración de escenarios, *scripts* y discursos.

La **estructura conceptual léxica** permite establecer una correspondencia basada en la noción del concepto primitivo entre el lenguaje externo y el lenguaje interno. Los conceptos básicos son léxicos y el significado de una oración se construye a través de los significados léxicos. Es así como cada constituyente léxico tiene una representación conceptual, cada categoría conceptual codifica una o más unidades lingüísticas, además de que establecen la diferencia entre "tipo" y "objeto"; las categorías son también cuantificables. La estructura conceptual de un elemento léxico es una entidad con cero o más argumentos.

Una **estructura de evento** permite identificar un tipo particular de evento asociado a la palabra o frase, por ejemplo, el logro, el estado. Básicamente una palabra o frase denota un estado, un proceso o una transición.

Las **estructuras de herencia Qualia** son estructuras que proveen los atributos esenciales para un objeto, distribuido entre 4 aspectos del significado de la palabra, y capturada en alguno de los siguientes papeles: el papel formal (el que lo distingue en otros dominios más amplios), el papel constitutivo (que da la relación entre el elemento léxico y sus constituyentes), el papel "téllico" (el que define propósito y meta) y el papel de agente (que trae información diversa). Estas estructuras forman parte de una teoría que busca sustentar la generación de un lexicón altamente organizado y estructurado. Pustejovsky(1991) es el autor de dicha teoría y de dicho lexicón, se basa en la idea de que esta teoría del significado léxico puede determinar el significado esencial de los elementos léxicos.

### **2.2.1 Aproximaciones basadas en diccionarios.**

La medición de la relación semántica tiene como fin revelar que tan parecidas son dos palabras entre sí y en qué sentido son parecidas. Lo anterior tiene una aplicabilidad muy amplia en el procesamiento del lenguaje natural, por ejemplo en la desambiguación de términos polisémicos. Se puede encontrar el sentido correcto de una palabra polisémica al calcular la relación semántica de la palabra en cada uno de sus sentidos, a una tabla de palabras no ambiguas que la rodean en el texto, y escogiendo el sentido que presenta el valor acumulativo de relación más alto. Otra aplicación es determinar la estructura de un texto, pues el conocimiento de la relación entre palabras permite identificar secuencias o cadenas de palabras altamente relacionadas que a su vez pueden hacerse "corresponder" (mapearse) en el texto y así encontrar los límites entre segmentos de texto que forman dominios o unidades "tópicas". Si un algoritmo es capaz de formar estas cadenas y diferenciarlas en función a su "fuerza", podrían hacerse resúmenes de un texto dado extrayendo segmentos de texto que correspondan a las cadenas más fuertes que cierto "umbral" o valor mínimo. Un uso evidente dentro del área de recuperación de información es reemplazar la equivalencia léxica convencionalmente usada, pues en lugar de recuperar documentos gracias

únicamente a la ocurrencia de los términos de la consulta en el documento, podemos incluir documentos que contienen términos que están semánticamente relacionados a los términos de la consulta. En el reconocimiento de texto y de los mensajes del habla, cuando aparece un lexema irreconocible, la interpretación del mismo puede realizarse escogiendo el candidato más cercanamente relacionado a un subconjunto de lexemas reconocidos con anterioridad.

Entre los pioneros de la ola contemporánea de investigación, están Osgood (1952), Quillian (1968), y Collin y Loftus (1975). Osgood definió un "diferencial semántico" como un intento de representar palabras como entidades en un espacio n-dimensional, donde la medición de la distancia entre ellas proviene de nuestro conocimiento de la geometría euclídeana. Osgood encontró que este sistema se sostenía en las "emociones connotativas" que se asocian a una palabra más que a su "significado denotativo" y lo discontinuó. Quillian, Collins y Loftus iniciaron entonces una aproximación basada en un procedimiento llamada "activación extensiva" o "*spreading activation*", el cual continúa en boga por algunos investigadores.

Desde hace dos décadas, cuando McGill y colegas determinaron 67 medidas de similitud para la recuperación de información en general, la medición de la relación semántica se convirtió en un motor muy importante de dicha investigación. Hoy en día se le llama por lo menos de tres formas: relación, similitud y distancia. Resnik (1995) ve la similitud como un caso especial de relación. La relación semántica o relacionismo semántico agrupa varios tipos de relaciones: meronimia (ventana-casa), antonimia (caliente-frío), y asociación funcional (océano - barco). El término distancia semántica puede ser más confuso, pues puede aludir tanto a qué tan similares son, como a la relación, es mejor verlo como lo inverso al relacionismo y no solamente lo inverso a la similitud.

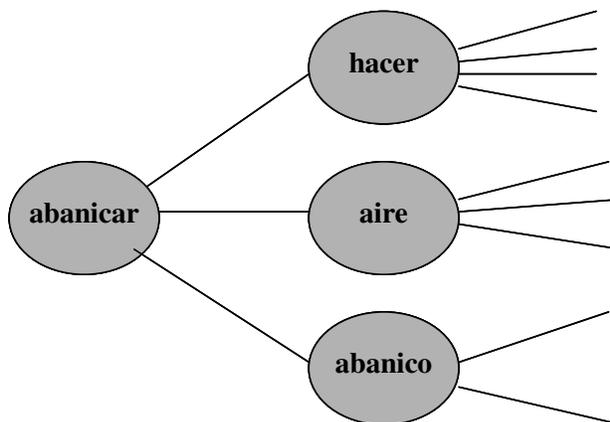
Las aproximaciones recientes para medir la relación semántica entre las palabras usan diccionarios, tesauros o redes semánticas, y algunas aproximaciones

combinan varios de estos enfoques. Aquí trataremos un ejemplo del primer caso: una aproximación basada en un diccionario; después, al final de este apartado, se comentarán muy brevemente otras aproximaciones.

Los diccionarios representan la fuente de conocimiento lingüístico más común. Por esta razón es natural que se haya pensado en aplicar algún diccionario al problema de determinar la distancia semántica entre las palabras. El primer diccionario que estuvo disponible en un formato leíble por computadora y el que probablemente siga siendo el más usado actualmente (Guthrie et. al. 1996) es el *Longman Dictionary of Contemporary English* (LDOCE). Una característica especial es que posee un vocabulario definidor, el *Longman Defining Vocabulary* (LDV) de 2851 palabras, que se emplean para formar las 56,000 definiciones de palabras restantes. Este vocabulario definidor se basó en la investigación de West (1953).

Kozima y Furugori (1993) estructuraron una red semántica a partir de LDOCE en la que a cada nodo corresponde una palabra del diccionario. Su trabajo aparece referido en Budanitsky (1999). Los enlaces van de cada nodo *n* a otros nodos, que son las palabras contenidas en la definición de *n*. Por ejemplo, la palabra abanicar aparece en el diccionario de la siguiente manera:

*Abanicar. I. Hacer aire con un abanico.*



**Figura 5. Parte de una red semántica.**

El nodo que representa a la palabra “abanicar” tiene enlaces con los 3 nodos que representan a las palabras que forman parte de su definición. A su vez, estas palabras deberán tener enlaces a los nodos de las palabras que componen sus definiciones respectivas. Si la palabra “abanicar” aparece en la definición de otra u otras palabras, existirán enlaces desde los nodos de éstas palabras hacia el nodo de “abanicar”. Se han omitido preposiciones, artículos o conjunciones, como sucede en las redes semánticas en donde se considera únicamente la relación de la entrada del diccionario en cuestión, con las palabras significativas de su definición.

Un buen diccionario debe ser un sistema cerrado de lenguaje natural, donde todas las palabras que se usen en la definición esten a su vez definidos por el mismo diccionario.

Si existe un vocabulario definidor, le corresponde la parte más densa de la red, mientras que las palabras que no son definidoras no están ligadas entre sí, por lo tanto se encuentran en la periferia.

A partir de estas observaciones, Kozima y Furugori tomaron al LDOCE para estructurar una red semántica. En un principio, separaron las palabras que pertenecen al LDV. Al subdiccionario resultante lo llamaron "Glosema", y estuvo compuesto de 2,851 entradas y 101,861 palabras en total. Cada entrada del Glosema está compuesta del término principal o "*headword*", la clase de palabra, o "*word class*", y una o más unidades correspondientes a las definiciones de los sentidos numerados para esa palabra, en el LDOCE. Cada unidad consiste en una "*head-part*" que corresponde al "*genus*" y una o más "*det-parts*" que corresponden a la diferencia.

Posteriormente el Glosema se tradujo subsecuentemente en una red semántica llamada "Paradigma". El Paradigma cuenta con 2,851 nodos que corresponden a las entradas del Glosema, interconectados con 295,914 ligas no etiquetadas. Cada nodo en el paradigma incluye un "*headword*", un "*word-class*" y un "*activity-value*" y se encuentra ligado a las palabras que aparecen en la definición de la entrada del Glosema (aproximadamente 104 ligas por nodo). Las ligas pueden ser de dos tipos, la "*référants*" es la que refleja la intención del "*headword*" del nodo. Contiene varios subconjuntos llamados "*subréférénts*", cada uno de los cuales corresponde a una unidad del Glosema (sus enlaces conectan cada nodo a los nodos que representan las palabras de la unidad). Estas ligas pueden llamarse "ligas salientes" o "*outgoing links*". La "*référé*" de un nodo  $n$  nos da información sobre su extensión, al ligar  $n$  a los nodos que se refieren a él en la definición respectiva del Glosema; a estas ligas se les puede llamar "entrantes" o "*incoming*".

Hecho esto, la similitud entre palabras en el LDV se calcula a través de la función de activación "*spreading activation*". Cada nodo tiene cierta "actividad", expresada a través del campo "*activity-value*" o  $v_n$ , la cual es recibida y transmitida a través de las ligas de los nodos. La actividad del nodo  $n$  en el tiempo  $T+1$  se calcula como sigue:

$$v_n(T+1) = \phi \left( \frac{R_n(T) + R'_n(T)}{2} + e_n(T) \right),$$

donde:

$R_n(T)$  es la actividad compuesta de los nodos referidos ("référant") por  $n$  en el tiempo  $T$

$R'_n(T)$  es la actividad compuesta de los nodos referidos ("référé") hacia  $n$  en el tiempo  $T$

$e_n(T)$  es la actividad impartida en el tiempo  $T$ , en "n" desde fuera

$\phi$  es una función que limita la salida al intervalo  $[0,1]$

El algoritmo que calcula la similitud  $\text{sim}_{\text{KF}}(w_k, w_l)$  entre las palabras  $w_k$  y  $w_l$  es:

1. Reiniciar valores de actividad de todos los nodos de la red.
2. Activar el nodo  $k$  correspondiente a la palabra  $w_k$ , con la fuerza  $e_k = s(w_k)$  por 10 pasos para obtener el patrón de activación  $P(w_k)$ . La *significancia* de  $w_k$ ,  $s(w_k)$  se define como la información normalizada de la palabra  $w_k$ , en el corpus de 5,487,056 palabras.
3. Observar  $a(P(w_k), w_l)$ , el valor de actividad del nodo  $l$  en el patrón  $P(w_k)$  (calculado según la fórmula (1)). La similitud es:

$$\text{sim}_{\text{KF}}(w_k, w_l) = s(w_l) \cdot a(P(w_k), w_l).$$

$$s(w_l) = \frac{-\log(n(w_l)/N)}{-\log(1/N)}$$

Donde:

$n(w_l)$  es el número de veces que aparece  $w_l$  en el corpus.

$N$  es el número de palabras del corpus.

Este algoritmo es para  $LDV \times LDV \diamond [0,1]$ , que representa solo el 5% de LDOCE.

Para que abarque todo LDOCE se aplica la siguiente fórmula:

$$\text{sim}_{\text{KF}}(W, W') = \psi \left( \sum_{w' \in W'} s(w') \cdot a(P(W), w') \right)$$

Donde:

$P(W)$  es el patrón resultante de la activación de cada  $w_i \in W$  con la fuerza  $s(w_i)^2 / \sum s(w_k)$  para 10 pasos o etapas y  $\psi$  es una función que restringe la salida al intervalo  $[0,1]$ .

El método de Kozima y Furugori (1993), así como los de Osgood (1952), Morris y Hirst (1991) son métodos libres de contexto o estáticos porque miden la distancia de las palabras sin tomar en consideración la influencia del contexto. Posteriormente Kozima e Ito (1997) derivaron, del método explicado arriba, otro que es más dinámico y sensible al contexto; la medida que utilizan toma en consideración la "dirección asociativa" de un par dado de palabras.

Las aproximaciones basadas en tesauros, como su nombre lo indica, buscan la relación entre las palabras a través de un tesoro. Éste se caracteriza por agrupar las palabras en una estructura de clases, categorías y subcategorías. Es una forma de agrupar ideas sin preguntarse cómo o porqué las palabras están relacionadas. Como consecuencia, la distancia semántica no se mide numéricamente, solo a través de valores como "palabra cercana" o "palabra no cercana". Analizando un tesoro, Morris y Hirst (1991) encontraron cinco tipos de relaciones semánticas entre las palabras: cuando tienen una categoría en común en sus entradas de índice, cuando una tiene una categoría en su índice que contiene un apuntador a una

categoría de otro, cuando uno es una etiqueta en la entrada indexada del otro o se encuentra dentro de una de sus categorías, cuando están ambas contenidas en la misma subcategoría, o cuando ambas tienen categorías en sus entradas de índice que apuntan a una categoría común. Por último, las redes semánticas (como se ha dado en llamar a cualquier representación que enlace nodos con arcos, donde los nodos son conceptos y los enlaces son tipos de relaciones entre conceptos), han medido la similitud entre las palabras calculando la longitud de una trayectoria entre un par de nodos (mientras más corta sea, mayor similitud), como en Resnik (1995), o distinguiendo distintos "niveles" de similitud, como lo hizo Hirst y St-Onge' s (1998) basándose en WordNet (Fellbaum 1998): el más alto (extra fuerte) cuando las palabras son una repetición literal de sí mismas, en siguiente nivel (fuerte) cuando tienen al mismo conjunto de sinónimos, o hay un enlace entre dichos conjuntos, y el nivel medio o regular, cuando existe una trayectoria "permisible" entre los conjuntos de sinónimos de las palabras de las que se quiere saber su grado de relación semántica.

### **2.3 Polisemia desde el punto de vista computacional**

La polisemia es uno de los principales problemas del procesamiento de lenguaje natural; su estudio ha estado muy ligado al de la traducción automática. El tratamiento computacional de la polisemia busca relacionar expresiones con sus significados correspondientes automáticamente. Una de las formas en que esto se resuelve es contar con diccionarios leñbles por computadora, de conocimiento lexicográfico de los diferentes sentidos de las palabras polisémicas y otra forma es simular el entendimiento humano usando procedimientos estadísticos para capturar patrones de co-ocurrencias de las palabras en el contexto. Bar-Hillel (1964), quien en un principio estuvo a favor de la traducción automática concluyó que no era posible identificar el sentido correcto automáticamente pues la desambiguación

requiere de una caracterización completa del conocimiento del mundo, lo cual es inabarcable.

Katz y Fodor (1963) publicaron su teoría de interpretación semántica, la cual establece unas estructuras que están contenidas en el lexicón mental, independientes del conocimiento del mundo. Para esta teoría crearon algoritmos y métodos para la desambiguación de sentido automático, logrando desambiguar 671 palabras en un período de 7 años, una cantidad de trabajo demasiado costosa. No obstante lo anterior hay un interés renovado, gracias a los diccionarios computacionales y al análisis de enormes corpus de texto con la capacidad computacional actual.

El grado óptimo de abstracción sobre múltiples significados está determinado frecuentemente por la meta computacional. En la traducción automática, consideramos que hay distintos sentidos si existe una elección léxica en el lenguaje destino, por ejemplo, la palabra "*know*" en inglés en la oración "*know a person*", se traduce en francés como "*connaitre*", mientras que "*know a fact*" se traduce en el mismo idioma como "*savoir*". El grado de abstracción está determinado en ocasiones por los límites de la propia tecnología. En la recuperación de información a gran escala, por ejemplo, la polisemia es sólo un aspecto de un problema mayor de ambigüedad, el cual incluye ambigüedad semántica (el verbo "*table*" contra el sustantivo "*table*") y la homonimia ("*river bank*" contra "*bank*" como institución financiera).

En los 80' s el uso de diccionarios leíbles computacionalmente (MRD' s *Machine Readable Dictionaries*) se propuso para la desambiguación de sentidos. Lesk (1986), enlaza definiciones en diccionarios si estas comparten palabras, los algoritmos de Lesk escogen el sentido contenido buscando la intersección entre palabras, por ejemplo, si "*pine*" tiene 2 sentidos y "*cone*" tiene otros tres, tenemos 6 sentidos posibles resultantes para las oraciones que tengan esas dos palabras, se elige el sentido en el que aparezcan mayor número de palabras en común.

Otra forma de realizar la desambiguación de sentidos es el uso de códigos temáticos. El código temático de un texto es aquél que está asociado con el mayor número de palabras del mismo texto. El sentido elegido es aquel asociado con el código temático. Otra forma es la desarrollada por Guthrie et al. (1991) en la que se obtienen "vecindarios" para las palabras polisémicas, a través de la búsqueda de co-ocurrencias en las definiciones y oraciones de ejemplo de todas las palabras que comparten el mismo código temático.

Recientemente los métodos de desambiguación basados en el corpus están en boga. Actualmente se puede usar *corpora* de millones de palabras, anteriormente el análisis a gran escala de palabras que co-ocurren con las palabras polisémicas no estaba al alcance. La identificación de sentidos basándose en el corpus, esto es, el análisis estadístico basado en el corpus, de co-ocurrencia de patrones, se convirtió en la corriente principal de la lingüística computacional. El entrenamiento de los sistemas de identificación se hace a través de oraciones de ejemplo para cada sentido de una palabra (desde 50 hasta cientos de ellas) y con ellas el sistema crea un modelo para cada sentido. Una vez que el modelo ha sido creado, el sistema escoge el modelo más apropiado para un nuevo uso de la palabra. Esto lo realiza a través de una medida de similitud entre las características del modelo y aquellas del contexto de la nueva ocurrencia. Miller y Leacock caracterizaron los tipos de información contextual que una computadora puede extraer del corpus. Así, llamaron contexto local a las que están cerca o adyacentes a la palabra polisémica mientras que el contexto tópico, son sustantivos que co-ocurren con un sentido en la discusión de un tópico dado. Los sistemas de identificación automática de sentido que hace explícito el uso de contextos tópicos o locales los desarrollaron Towell y Voorhees(1998) y Leacock et al. (1998).

En la reunión de Senseval<sup>7</sup> en 1998 se llevó a cabo la desambiguación de 335 palabras polisémicas, los resultados condujeron a lo que Bar-Hillel llamaron la

---

<sup>7</sup> Senseval es una conferencia para evaluar el estado del arte en relación con los sistemas de desambiguación de sentidos, basados en el corpus (*International Workshop on Evaluating Word*

"falacia del 80%", a partir del cual el esfuerzo para desambiguar el restante 20% es mucho mayor que el realizado con antelación.

Otro aspecto problemático de esta aproximación es el llamado "cuello de botella de adquisición del conocimiento" (Gale et al. 1992), en el que el material para entrenar un programa requiere, para cada palabra polisémica, de un corpus del contexto que tiene que ser dividido en diferentes sentidos, y el esfuerzo manual requerido debe hacerse en cada caso sin posibilidad de aprovechar el esfuerzo realizado en otras palabras.

Schütze toma una dirección distinta para hacer una identificación de sentido basado en el corpus, donde su preocupación central es diseñar un algoritmo que sea psicológicamente plausible y generalmente aplicable. En lugar de asignar una palabra polisémica a un sentido discreto, Schütze reúne sentidos de palabras que son similares, pues comparten contextos similares, y después define estos *clusters* como sentidos de palabras.

Otros autores como Dolan, Vanderwende y Richardson (1999) combinan las dos aproximaciones principales: el uso de MRD' s y las basadas en el corpus. Como Shutze, utilizan una medida de similitud en el contexto de las palabras para determinar si están cercanas en significado. Sin embargo, los contextos que ellos usan son muy especializados y, además, son enlaces en una red semántica llamada MidNet, la que se obtiene del análisis de las definiciones de un MRD. MidNet no trata de clasificar sentidos individuales de palabras, sino reconocer conceptos que son similares a aquellos a los que se les da un uso nuevo. La aproximación de Schütze y Dolan puede ser útil para la recuperación de información donde la tarea es vincular el contexto de la consulta a contextos similares en la base de datos de los documentos, pero no es obvio cómo esas aproximaciones pueden ser usadas en

---

*Sense Disambiguation Systems*). El primer SENSEVAL se realizó el verano de 1998, culminando en un taller realizado en Herstmonceux Castle, en Sussex, Inglaterra, del 2 al 4 de septiembre de dicho año.

aplicaciones como la traducción automática, sin una codificación extensa y manual.

En la lingüística computacional es más fácil manejar los homónimos que las palabras polisémicas, pues el contexto de los homónimos tiene un vocabulario muy distinto, mientras que cuando hay polisemia, puede ser muy similar. Por otro lado, la recuperación de información, donde a un usuario se le dan a escoger varios documentos en función a una consulta, es un ambiente mucho menos restrictivo que la traducción automática.

El campo de la lingüística computacional crece en dos direcciones. La colección de textos en línea está creciendo en tamaño y así la demanda de procesamiento a gran escala. La investigación a gran escala se especifica como una meta importante de muchos proyectos de investigación. La otra dirección es la aplicación de computadoras a nuevos dominios, los editores y compañías de computación experimentan con diccionarios en línea, bibliotecas digitales, métodos para navegar, buscar y acceder a datos relevantes en ambientes multilingües, etc.

## **2.4 Diccionarios computacionales**

Una de las tareas de la lingüística computacional es la de construir diccionarios para su uso por computadora. La utilidad más importante de los diccionarios computacionales se haya en su aplicación en la recuperación de información y en la traducción automática. En el primer caso, por ejemplo, un diccionario computacional puede utilizarse en la mejor resolución de consultas, analizando el contenido de los textos a través del diccionario. En la traducción automática, se aplicaría un diccionario de este tipo con el fin de lograr traducciones más correctas

identificando el sentido preciso que tiene una palabra polisémica<sup>8</sup> en un idioma y su traducción a otro. Un método natural para la construcción de un diccionario semántico utilizable por los sistemas computacionales es la definición de unas palabras a través de otras, como se hace en los diccionarios explicativos tradicionales. Esta sustitución se repite iterativamente hasta llegar a conceptos cada vez más "simples"; en algún momento es necesario detener este proceso y así se llega a un conjunto de palabras "primitivas".

El concepto de primitivas semánticas ha despertado mucho interés entre los lingüistas, pues se ha discutido si existe realmente un conjunto de palabras a través de las cuales se puedan definir las demás palabras, y si estas primitivas semánticas forman un conocimiento humano innato independiente del idioma del que se trate.

Más allá de esta discusión, una forma de obtener diccionarios computacionales es a través de la construcción de vocabularios definidores compuestos por primitivas semánticas. Como se mencionó anteriormente, la construcción de un diccionario computacional a partir de un diccionario tradicional, es una aproximación que presenta el problema de los ciclos en las definiciones al sustituir iterativamente una palabra por su definición.

La construcción de un diccionario computacional, en el que las definiciones de las palabras se realicen a través de otras palabras, despierta algunas dudas: ¿Se puede transformar un diccionario tradicional en un diccionario "para la máquina"? ¿Qué problemas se enfrentarán?. En geometría, por ejemplo, se puede desarrollar cualquier término (digamos, "bisectriz") sustituyendo en su definición cada palabra (digamos, "ángulo") con la definición de esta última. El sistema de definiciones se puede construir de tal manera que al repetir este proceso iterativamente, se llegará a una definición larga, compuesta sólo por los términos más simples como "punto" y "línea".

---

<sup>8</sup> Polisemia: Fenómeno muy común en el lenguaje en el que una misma palabra tiene múltiples

Ahora bien, ¿qué va a pasar si en cada definición de un diccionario explicativo, sustituimos iterativamente cada palabra con su definición? ¿Cuáles palabras se deben considerar primitivas?.

El problema de las primitivas semánticas es uno de los problemas más interesantes en semántica. ¿Existe un pequeño número de palabras con las cuales se pueden definir todas las demás palabras? ¿Y si tales palabras existen, qué realidad se refleja en este fenómeno? ¿Existe la *lingua mentalis* (Wierzbicka, 1980) en realidad o es un artificio creado en el marco de una teoría?.

Existen diferentes puntos de vista sobre este problema. La teoría de Wierzbicka (1980, 1996) proclama un número muy reducido (alrededor de 60) de las llamadas primitivas, es decir, las palabras que se pueden usar para definir, en su totalidad, los cientos de miles de palabras que existen en el lenguaje.

Por otro lado, por ejemplo, Apresjan (1974, 1995) sugiere usar un conjunto restringido de las palabras del lenguaje como la base para construir las definiciones de otras palabras, pero en su opinión este conjunto no sería muy pequeño.

Esta idea también se usa en la lexicografía práctica. En el diccionario *Longman dictionary of contemporary English* se usan en las definiciones sólo las palabras del vocabulario definidor (*Defining Vocabulary o DF* en inglés), el cual es un vocabulario restringido. El tamaño de este vocabulario, conocido como *Longman defining vocabulary*, es alrededor de dos mil.

Sin involucrarse en la discusión sobre la naturaleza de las primitivas y el vocabulario definidor, notemos que un problema importante que impide considerar un diccionario tradicional como un diccionario “para la máquina” es la presencia

---

significados. Para la teoría semántica y las aplicaciones semánticas es un problema muy importante, especialmente en la traducción y en la lexicografía.

de ciclos en las definiciones, esto es, cuando una palabra se define a través de otra y ésta a través de la primera. Por ejemplo:

*abeja: insecto que segrega miel.*

*miel: sustancia que producen las abejas.*

Pueden existir ciclos más largos, por ejemplo:

*convenio: pacto, acuerdo.*

*acuerdo: pacto, tratado.*

*tratado: convenio.*

Estas son definiciones tomadas (y un tanto simplificadas) del Diccionario Anaya de la lengua española (1997)<sup>9</sup>. Por supuesto, este sistema de definiciones no puede ser utilizado como un diccionario computacional.

Los ciclos en un diccionario explicativo representan un gran problema para cualquier trabajo semántico que involucra algún tipo de razonamiento. Un ejemplo es la resolución de la ambigüedad<sup>10</sup> en el sentido de las palabras. Este caso se presenta cuando una palabra tiene varios sentidos (polisemia y homonimia, como se explicó en la sección 2.1.1). Un método para encontrar el sentido preciso de una palabra podría hacer uso de un diccionario con el fin de convertir una expresión compleja (como lo es cualquier expresión lingüística) en una expresión formada por términos "primitivos" o simples sobre las cuales se posea un dominio semántico que permita al sistema "entender" la pregunta y buscar una solución aceptable. Una forma muy simple que se nos podría ocurrir sería buscar las

---

<sup>9</sup> En el anexo 4 se pueden consultar ejemplos reales de definiciones del diccionario y ciclos involucrados.

<sup>10</sup> La desambiguación de las palabras es indispensable en la traducción automática y también en la recuperación de información, sobre todo en la resolución de consultas donde es necesario traducir las consultas a ciertos términos que un sistema pueda manipular para encontrar la respuesta más apropiada

definiciones de los sentidos que esa palabra tenga en el diccionario, y, para elegir el sentido más indicado en función al contexto que nos da el resto de la pregunta formulada, se podría realizar una búsqueda "a profundidad" en cada sentido, es decir, buscar ocurrencias de las palabras que conforman la pregunta en la definición en cada uno de los sentidos, aquel sentido en donde se presenten más ocurrencias, se considerará el sentido correcto. Por supuesto este método en muchos casos podría arrojar resultados insatisfactorios o no dar ninguna solución, pero hay otros métodos más complejos que se basan en la co-ocurrencia de términos recientemente explicada, aplicando tesauros, palabras clave, etcétera. Mientras más ciclos se encuentran en las definiciones menos oportunidad habrá de diferenciar unos sentidos de otros, además de que, computacionalmente, se vuelve muy costoso el manejo de los ciclos.

Concluyendo, un "buen" diccionario explicativo no debe tener ciclos. Para lograrlo, o bien hay que cambiar las definiciones, o bien, excluir las definiciones de algunas palabras del diccionario convirtiéndolas en las palabras primitivas que posteriormente se puedan definir en un sistema computacional con otros medios, por ejemplo a través de un lenguaje de programación y no a través de otras palabras. En esta tesis se explora esta última posibilidad.

## 3 Algoritmo para la selección automática de primitivas semánticas

---

### 3.1 Conceptos básicos

Nuestra solución utiliza los métodos de la lexicografía computacional (Saint-Dizier and Viegas, 1995).

Específicamente, representamos el diccionario como un grafo dirigido. La idea de representar el diccionario como un grafo (una red semántica), no es nueva, es una idea desarrollada por autores como Evens (1988) y Fellbaum (1990). Kozima y Furugori (1993) también analizan una red semántica, en este caso, para saber qué palabras se "activan" empezando de alguna palabra determinada.

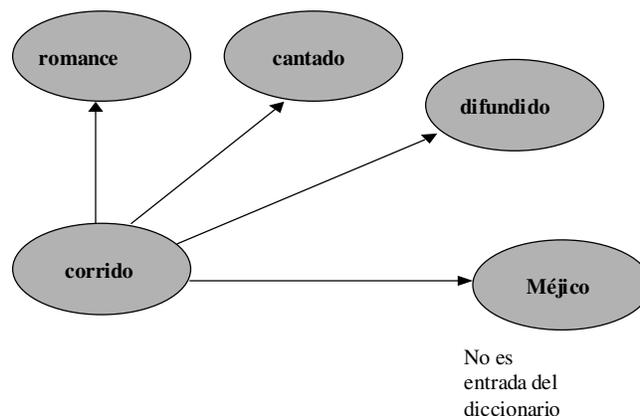
En nuestro grafo, los vértices o nodos son las palabras que se mencionan en el diccionario, tanto las palabras que se definen como entradas, como las que se usan en las definiciones. La misma palabra puede ocurrir tanto como una entrada como varias veces en las definiciones, pero se cuenta como el mismo vértice.

Hay diferentes maneras de considerar que dos ocurrencias (*tokens* en inglés) son la misma palabra (*type* en inglés): como cadenas de letras, por la forma base, por la raíz, por el significado específico, etc. En la sección 4.1 se dan más

detalles sobre los dos grafos a los que aplicamos el algoritmo de depuración y el algoritmo principal (Grafo 1, por forma base, y Grafo 2, por significado).

Definidos los vértices, las flechas o arcos del grafo se definen como sigue: la flecha desde la palabra  $v_1$  hasta la palabra  $v_2$  significa que en la definición de la palabra  $v_1$  se usa la palabra  $v_2$ . Las palabras que no son entradas no tienen flechas salientes, mientras que las que no se usan en las definiciones de otras palabras, no tienen flechas entrantes. Un ejemplo del primer caso (palabras que no son entradas del diccionario) es la palabra "Méjico". Esta palabra aparece en la definición de "corrido", como se ve a continuación:

*Corrido.- I. De correr. Romance cantado, difundido principalmente en Méjico.*



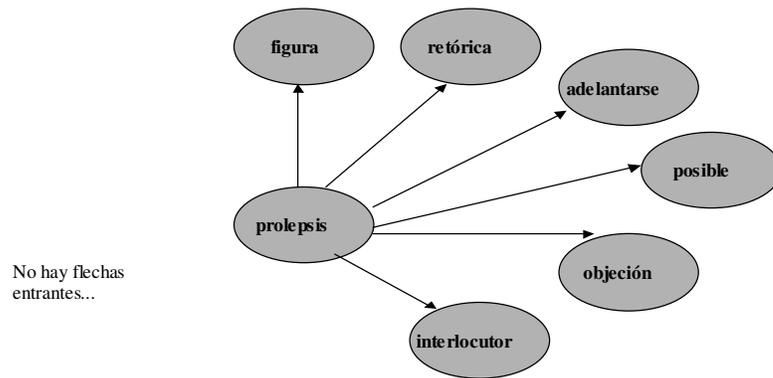
**Figura 6. Definición de la palabra "corrido".**

Aunque "Méjico" también aparece en las definiciones de muchas otras palabras como "ajolote", "cayo", "peso", etcétera, no es considerada como una palabra sustancial en el vocabulario, al menos en este diccionario, ya que no es un diccionario enciclopédico. Como vemos éste será el caso de nombres propios de

lugares o personajes. Otras palabras de este tipo son: "Grecia", que aparece en la definición de la palabra "griego", "María" que aparece en la definición de "magnificat", "Cataluña" que aparece en la definición de "cobla", "Platón", en la definición de "platónico", entre otras.

Otro caso es el de entradas como "prolepsis"; uno de los sentidos que esta palabra tiene es:

*Prolepsis.- I. De gr. pro-lepsis: percepción. Figura retórica consistente en adelantarse a las posibles objeciones del interlocutor.*



**Figura 7. Uno de los sentidos de la palabra "prolepsis".**

Esta palabra es una entrada del diccionario, por lo tanto tiene flechas salientes hacia las palabras de su definición, como se muestra en la Figura 7, pero no hay flechas entrantes hacia ella, pues el resto de las palabras del diccionario no la usan en sus definiciones. Otras palabras que tienen esta característica son: "ababol", "frambuesa", "guachinango", "metempsicosis", "zaino", etcétera. Si una palabra no tiene flechas entrantes, es una palabra que se consideró innecesaria para las otras definiciones. Todas las palabras que no tienen flechas entrantes o salientes podrán ser vértices a eliminar en el proceso de depuración del grafo, pues como se verá en la siguiente sección, éstas características evitará que formen

ciclos y la detección de ciclos es nuestra estrategia para incluir una palabra dentro de conjunto de primitivas. También tenemos una razón lingüística para no considerarlas candidatas a primitivas semánticas, pues o no formaron partes de las entradas del diccionario, o bien son innecesarias para las definiciones de otras entradas del mismo, por lo que todas estas palabras no son tan "esenciales" como las otras para el diccionario, y no es conveniente que formen parte de nuestras primitivas semánticas.

### **3.2 Modelo matemático y algoritmo principal**

Desde el punto de vista matemático, el problema y su solución son los siguientes:

Sea  $\mathbf{G} = \{\mathbf{V}, \mathbf{F}\}$  un grafo dirigido definido por los conjuntos  $\mathbf{V}$  de  $N$  vértices y  $\mathbf{F} \subseteq \mathbf{V} \times \mathbf{V}$  de flechas. Un ciclo en este grafo, lo entenderemos como un ciclo dirigido.

Digamos que un subconjunto  $\mathbf{P} \subseteq \mathbf{V}$  es un subconjunto *definidor* si cualquier ciclo en el grafo  $\mathbf{G}$  contiene un vértice de  $\mathbf{P}$ . O sea, si el grafo  $\mathbf{G}' = \{\mathbf{V}', \mathbf{F}'\}$  que se obtiene del  $\mathbf{G}$  eliminando los vértices que pertenecen al  $\mathbf{P}$  –es decir,  $\mathbf{V}' = \mathbf{V} \setminus \mathbf{P}$  y  $\mathbf{F}' = \mathbf{F} \cap (\mathbf{V}' \times \mathbf{V}')$ – no tiene ciclos. Llamaremos definidores a los vértices  $p \in \mathbf{P}$ .

Digamos que un subconjunto definidor  $\mathbf{P} \subseteq \mathbf{V}$  es minimal si ningún subconjunto  $\mathbf{P}' \subset \mathbf{P}$  es definidor. O sea, si la eliminación de cualquier elemento de  $\mathbf{P}$  produce un subconjunto no definidor. Es decir, para cada vértice  $p \in \mathbf{P}$ , existe un ciclo en  $\mathbf{G}$  que contiene  $p$  y no contiene ningún otro vértice de  $\mathbf{P}$ .

El conjunto *definidor minimal* no tiene que ser el más pequeño (el que está contenido en todos los subconjuntos definidores; tal subconjunto usualmente no existe) ni siquiera del tamaño más pequeño posible, en el mismo grafo pueden

existir muchos conjuntos definidores minimales de tamaños diferentes, como es muy fácil mostrar con ejemplos y como también será claro en nuestro algoritmo.

El algoritmo que aquí presentamos resuelve el siguiente problema: dado un grafo dirigido  $\mathbf{G}$ , encontrar un subconjunto definidor  $\mathbf{P}$  minimal, aunque no sea del tamaño más pequeño.

El algoritmo encuentra una de múltiples variantes posibles de un subconjunto definidor minimal. La selección de la variante se puede controlar con un ordenamiento  $\sigma: \{1, \dots, N\} \leftrightarrow \{1, \dots, N\}$  que ordena los números de 1 a  $N$  en una secuencia  $\sigma(1), \dots, \sigma(N)$ . Por ejemplo: 6, 3, 5, 2, 4, 1 es un ordenamiento para  $N = 6$ .

Los conjuntos  $\mathbf{P}$  generados basándose en diferentes ordenamientos  $\sigma$  son usualmente diferentes.

Como se observará en el algoritmo, el sentido del ordenamiento es el siguiente: los primeros vértices tienden a ser no definidores y los últimos tienden a entrar en  $\mathbf{P}$ . Específicamente, el primer vértice en el ordenamiento, si no tiene un lazo, siempre es no definidor.

Ahora bien, dado  $\mathbf{G}$  y  $\sigma$ , el algoritmo funciona como sigue. Se construye, paso a paso, un subgrafo  $\mathbf{G}' \subseteq \mathbf{G}$  sin ciclos. Inicialmente está vacío. Se construye, insertándole uno por uno (en el orden dado  $\sigma$ ), los vértices de  $\mathbf{G}$  con sus relaciones hacia los vértices ya insertados. En cada paso,  $\mathbf{G}'$  se mantiene sin ciclos: si el vértice a insertar le causa ciclos, se considera un definidor y no se inserta en  $\mathbf{G}'$ . Al terminar el proceso, se tiene el conjunto definidor  $\mathbf{P}$ , que por su construcción es minimal (porque cada  $p \in \mathbf{P}$  produce ciclos incluso en un subconjunto del  $\mathbf{G}'$ ).

El paso computacionalmente más costoso del algoritmo es la verificación de que el nuevo vértice no le causaría ciclos al subconjunto  $\mathbf{G}'$ . Por esto, para cada

vértice  $v_i \in \mathbf{V}$  el algoritmo mantiene un conjunto  $\mathbf{A}_i$  de vértices accesibles en  $\mathbf{G}'$  desde  $v_i$ , diciéndose del vértice  $u$  que es accesible desde  $v$  si existe un camino dirigido en el grafo desde  $v$  hasta  $u$ . El algoritmo se basa en el hecho de que es una relación transitiva: si  $a$  es accesible desde  $b$  y  $b$  desde  $c$ , entonces  $a$  es accesible desde  $c$ .

Con el fin de disminuir el conjunto de datos a procesar se puede llevar a cabo la depuración del grafo, véase más adelante en esta misma sección el algoritmo B. Éste consiste en agregar elementos al conjunto  $\mathbf{P}$  y eliminar los vértices (y sus relaciones) del conjunto  $\mathbf{G}$ .

En la página siguiente se detalla el algoritmo principal (algoritmo A), en forma de pseudocódigo, incluyendo algunas etiquetas para identificar aquellos pasos de los que convendrá hacer comentarios.

---

**Selección\_de\_primitivas\_semánticas**

---

---

- Formar el conjunto de los definidores  $\mathbf{P} = \emptyset$  y de los no definidores  $\mathbf{V}' = \emptyset$ .
  - Depuración del grafo  $\mathbf{G}$ , según algoritmo B(opcional).
  - Verifica\_cada\_nodo
- 
- 

**Verifica\_cada\_nodo**

---

---

Para  $i = 1, \dots, N$  repetir:

- Seleccionar  $\sigma(i)$ -ésimo vértice  $v \in \mathbf{V}$
- Verificar si  $v$  provoca ciclos en  $\mathbf{G}'$ . Para esto:
- Para cada flecha  $(v \rightarrow u) \in \mathbf{F}$  que va desde  $v$  tal que  $u \in \mathbf{V}'$ , repetir:
  - Para cada flecha  $(v \leftarrow w) \in \mathbf{F}$  que va hasta  $v$  tal que  $w \in \mathbf{V}'$ , repetir:
    - Verificar que  $w \notin \mathbf{A}_u$ . (Paso crítico 1)
    - Si falla, agregar  $v$  a  $\mathbf{P}$ .
    - Si no falla, agregar  $v$  a  $\mathbf{G}'$ .

---

---

**Agregar\_nodo\_a\_G'**

---

---

- Agregar  $v$  a  $\mathbf{V}'$ .
  - Formar  $\mathbf{A}_v = \cup \mathbf{A}_u$ , donde la unión se forma por todas las flechas  $(v \rightarrow u) \in \mathbf{F}$  que van desde  $v$  tales que  $u \in \mathbf{V}'$ .
  - Para cada flecha  $(v \leftarrow w) \in \mathbf{F}$  que va hasta  $v$  tal que  $w \in \mathbf{V}'$  repetir:
    - Para cada vértice  $q \in \mathbf{V}'$  tal que  $q = w$  ó  $w \in \mathbf{A}_q$  repetir:
    - Agregar  $\{v\} \cup \mathbf{A}_v$  a  $\mathbf{A}_q$ . (Paso crítico 2)
- 
- 

**Figura 8. Algoritmo A (principal).**

Después del paso de depuración, se supone que  $\mathbf{V}$ ,  $\mathbf{F}$  y  $N$  están definidos por el grafo  $\mathbf{G}$  ya depurado, es decir, posiblemente sean menores que el original. También el algoritmo mantiene el conjunto  $\mathbf{V}'$  de los vértices ya incluidos en el subconjunto  $\mathbf{G}'$  y el conjunto  $\mathbf{P}$  (se pueden considerar como unas coloraciones de

los vértices del conjunto  $\mathbf{G}$  pues su intersección es vacía). Al terminar el algoritmo,  $\mathbf{P}$  será un conjunto definidor minimal. Véase el algoritmo detallado en la Figura 8.

El algoritmo tiene complejidad exponencial, siendo las operaciones que se repiten el mayor número de veces, los pasos críticos 1 y 2. Dado el gran tamaño del diccionario (30 mil vértices en nuestro caso), se deben tomar dos precauciones para mantener su tiempo de espera en los límites razonables:

- Implementar eficientemente la operación más crítica y
- Disminuir el tamaño de los datos a procesar.

En cuanto a la implementación eficiente, en una computadora PC nosotros representamos los conjuntos  $\mathbf{A}_i$  como una matriz de  $N \times N$  bits empacados en  $N \times \left( \frac{N}{32} + 1 \right)$  números enteros de 32 bits. La operación de unión de dos conjuntos (dos renglones de esta matriz) se implementa a través de la disyunción lógica (OR) de los números enteros, lo que es 32 veces más rápido que operar sobre cada elemento independientemente.

También en la misma PC Pentium II, probamos la implementación a través de las unidades de 64 bits usando la tecnología MMX, pero no observamos incremento considerable en la rapidez en comparación con las operaciones en 32 bits. La depuración de cualquiera de los dos grafos, la herramienta puede realizarla en menos de 5 minutos y es el mismo caso para la ejecución del algoritmo principal.

En cuanto a disminuir el tamaño de los datos a procesar, se pueden observar dos hechos:

1) Los vértices  $v$  con un lazo  $(v \rightarrow v) \in \mathbf{F}$  tienen que estar en  $\mathbf{P}$  pues no pueden estar en  $\mathbf{G}'$ . Nótese que la prueba en el paso crítico 1 falla para tales

---

```
Depurar_grafo
Para cada vértice con lazo, repetir:
    • Agregar el vértice al P y eliminarlo del G
      junto con sus flechas.
Para cada vértice que no tiene salientes o entrantes,
repetir:
    • Eliminarlo de G junto con sus flechas,
      considerándolo no definidor.
```

---

---

**Figura 9. Algoritmo B (depuración).**

vértices. Entonces, estos vértices –aunque son pocos– se pueden de antemano agregar a **P** y quitar (con todas sus relaciones) de **G** para fines del procesamiento posterior.

2) Los vértices que no tienen ninguna flecha entrante o ninguna flecha saliente no pueden estar en **P** (recuerde que **P** es un subconjunto definidor minimal) pues para tal vértice  $v$ , si **Q** es un conjunto definidor entonces  $\mathbf{Q} \setminus \{v\}$  también lo es. Nótese que la prueba en el paso crítico 1 no se ejecuta para tales vértices y de este modo no puede fallar. Entonces, estos vértices –y en nuestro caso son muchísimos– se pueden de antemano quitar (con todas sus relaciones) de **G** considerándolos no definidores.

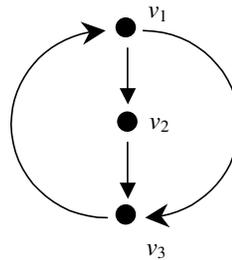
La eliminación de un vértice puede hacer que otros vértices pierdan todas sus flechas entrantes o salientes. Entonces, estas dos observaciones sugieren el algoritmo de depuración mencionado (algoritmo B), véase la Figura 9 .

La eliminación de un vértice se puede implementar como una coloración. El grafo **G** que se obtiene después de la depuración, satisface las siguientes condiciones:

1. No tiene lazos.
2. Cada vértice es transitivo, es decir, tiene flechas tanto entrantes como salientes. En tal grafo, para cada vértice  $v$  existe un conjunto definidor minimal **P**

tal que  $v \notin \mathbf{P}$ . Más generalmente, para cada conjunto compatible  $\mathbf{Q} \subseteq \mathbf{V}$  (diciéndose que el conjunto  $\mathbf{Q}$  es compatible si no existen ciclos en  $\mathbf{G}$  que contengan únicamente los vértices de  $\mathbf{Q}$ ) existe un conjunto definidor minimal  $\mathbf{P}$  tal que  $\mathbf{Q} \cap \mathbf{P} = \emptyset$ . Esto se comprueba por la aplicación del algoritmo A con un ordenamiento  $\sigma$  que empieza con los elementos de  $\mathbf{Q}$ .

Entonces, en tal grafo la propiedad de ser un definidor no es propia del vértice sino sólo es relativa a la selección de un conjunto  $\mathbf{P}$ . Es decir, ningún vértice, salvo los vértices con lazos, es un definidor “por sí mismo”, o bien, que entra en cualquier conjunto definidor.



**Figura 10. Contraejemplo.**

Lo contrario no es cierto: en el grafo depurado con el algoritmo B todavía pueden existir los vértices que no entran en ningún conjunto definidor del tamaño minimal, es decir, se pueden eliminar de cualquier conjunto definidor. Un ejemplo de tal situación –el vértice  $v_2$ – se muestra en la Figura 10. Nosotros no tenemos ningún algoritmo rápido para detectar tales vértices. Sin embargo, nuestros experimentos con el diccionario real mostraron que en este diccionario los vértices de este tipo, si existen, no constituyen más del 20% del grafo depurado, pues encontramos los conjuntos definidores minimales cuya unión cubre un 80% del grafo. Por lo tanto, la detección previa de tales vértices no contribuiría significativamente en la rapidez del algoritmo.

Empleando cualquiera de los cuatro métodos u ordenamientos establecidos, el sistema toma cada palabra y verifica si la inclusión de esta palabra en el grafo sin

ciclos, provoca un ciclo entre los nodos que ya se encuentran en dicho grafo. Por esta razón las primeras palabras tienden a ser no-definidoras, pues entran con mayor facilidad en un grafo sin ciclos en el cual hay pocas palabras. Conforme el algoritmo va incrementando el número de nodos dentro del grafo sin ciclos, encontrará ciclos con mucho mayor frecuencia y las palabras que se analizan al final son las que más fácilmente quedarán en el conjunto minimal definidor.

Por ejemplo, la definición de la palabra "abastecer" que tenemos en el diccionario Anaya es la siguiente (origen etimológico I, sentido 1):

abastecer I. De a- + ant. bastecer < bastir = proveer < germ. bastjan = construir 1. (verbo transitivo, pronominal). Proveer de cosas necesarias. FAM. Abastecedor, -a, abastecimiento; desabastecer. SIN. Surtir, suministrar, aprovisionar, avituallar, abastar.

Como esta palabra no tiene un lazo y es una entrada del diccionario (no solamente una palabra empleada en la definición de alguna otra que sí es entrada), está codificada en el archivo **net1.txt**, teniendo como flechas entrantes y salientes las siguientes (debajo de la palabra están las palabras de las flechas salientes de "abastecer", más abajo están las palabras de las flechas entrantes a la misma palabra):

<b>Palabra</b>  9:abastecer	<b>Flechas salientes a:</b> 2637:cosa, 6846:necesario, 8116:proveer,
	<b>Flechas entrantes de:</b> 8115:proveedor, 8618:repostar, 9401:suministrar,

Si la palabra "abastecer" es la primera palabra que el algoritmo evalúa, esta palabra necesariamente entra en el grafo depurado (el algoritmo la marcaría), pues sus flechas entrantes y salientes apuntan a y desde nodos que no están aún incluidos.

Más tarde le tocará su turno a la palabra "proveer", origen etimológico I, sentido 2:

proveer I. Del lat. providere < pro = hacia adelante + videre = ver 2. (verbo transitivo, pronominal). Suministrar lo necesario FAM. Proveedor, -a, providencia, providente, provisión, provisor, -a; desproveer. SIN. 2. Abastecer, equipar, surtir, proporcionar.

<b>Palabra</b>  8116:proveer	<b>Flechas salientes a:</b> 6846:necesario, 8672:resolución, 9401:suministrar, 9604:tener, 9890:tribunal,
	<b>Flechas entrantes de:</b> 9:abastecer, 846:armario, 1716:cargar, 1850:celular, 1878:cepillo, 2796:cumplir, 3508:dotar, 3936:equipar, 4391:fábrica, 5584:injertar, 6020:lepidóptero, 6451:medusa, 7485:percebe, 8121:provisión, 8172:pulpo,

Supongamos que esta palabra (representada por un nodo) no tiene aún flechas entrantes y salientes en el grafo sin ciclos o que aún cuando las tenga, estas flechas entrantes y salientes no formarían un ciclo con la inclusión de "proveer". De esta forma, "proveer" se marcaría como parte del grafo sin ciclos y se actualizaría el conjunto de nodos accesibles a "proveer" y los conjuntos de nodos accesibles de las palabras que usen "proveer" en su definición.

Posteriormente, la palabra a analizar es "suministrar":

suministrar I. Del lat. subministrare 1. (verbo transitivo). Proporcionar lo necesario, abastecer s. víveres; s. datos. FAM. Suministrable, suministrador, -a, suministro. SIN. Proveer, surtir, equipar, aprovisionar. ANT. Desproveer, desguarnecer.

<b>Palabra</b>  9401:suministrar	<b>Flechas salientes a:</b> 9:abastecer, 6846:necesario, 8080:proporcionar,
--	--

	<b>Flechas entrantes de:</b> 226:administrar, 443:alimentar, 2964:deparar, 3325:digital, 8116:proveer, 9121:servir, 9402:suministro, 9437:surtidor, 9438:surtir,
--	--

En este caso el algoritmo encontraría por lo menos un ciclo al estar evaluando si existe un camino en el grafo depurado, que nos conduzca de algún nodo accesible desde esta palabra hasta algún nodo que tenga entre sus nodos accesibles a "suministrar" (lo cual nos indicaría que hay un ciclo) y no la marcaría, es decir, la consideraría primitiva.

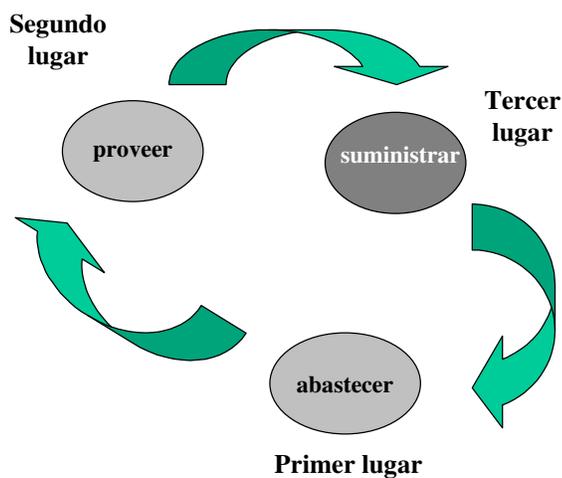
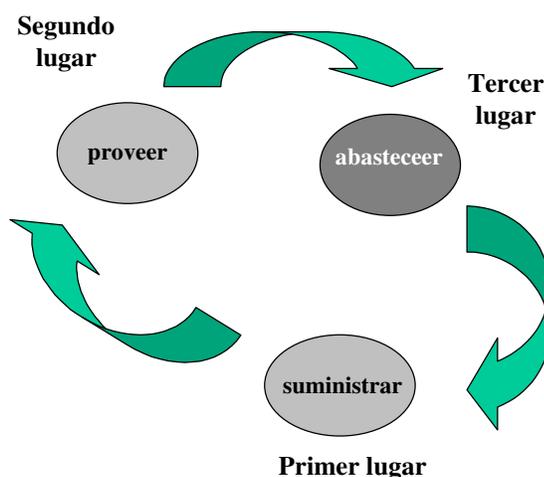


Figura 11. Ejemplo: la palabra seleccionada como primitiva es "suministrar".

En cambio, si la primera palabra que se marca como nodo del grafo depurado, es "suministrar", seguida de "proveer", al llegar el proceso de análisis a "abastecer", el algoritmo encontrará que es "abastecer" la que provoca un ciclo e incluirá ésta como primitiva.



**Figura 12. Ejemplo: la palabra seleccionada como primitiva es "abastecer".**

De esta manera se ve que, tratándose de las mismas palabras, y las mismas definiciones, el proceso de análisis produce resultados distintos dependiendo de la existencia de salidas de una palabra que conduzcan a entradas a esa misma palabra en el grafo sin ciclos, y a su vez, este suceso depende del ordenamiento de los nodos a analizar y, por supuesto, de la presencia de los ciclos.

### 3.3 Implementación del algoritmo

El sistema por medio del cual procesamos el diccionario está formado de 3 módulos o programas y una librería, los cuales son:

- a) el módulo **graph**,
- b) el módulo **circle**,
- c) el módulo **visualizer**, y
- d) la librería **circle**.

En términos generales es módulo **graph** controla la lectura, depuración y visualización del grafo. También se encarga de ejecutar alguno de los 4 métodos u ordenamientos explicados en la sección 4.

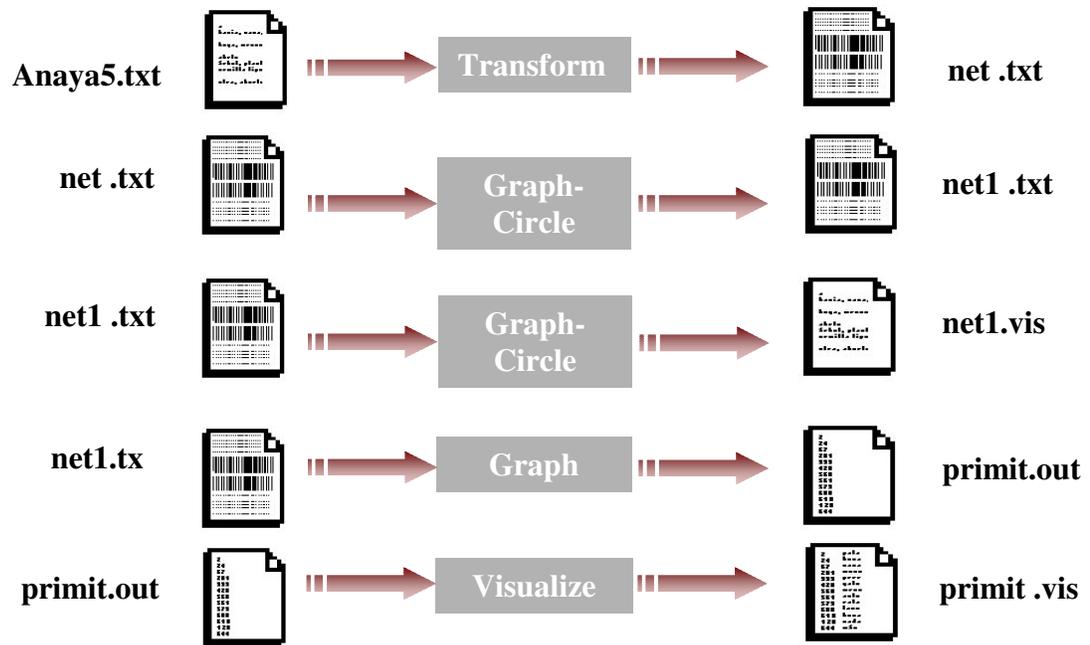


Figura 13. Archivos de entrada y archivos de salidas del sistema.

En lo referente a la lectura, depuración y visualización, **graph** llama a métodos que se encuentran desarrollados en el módulo **circle**, por lo que éste módulo no se ejecuta directamente por el usuario. El módulo **visualizer**, en cambio, es un módulo que ejecuta el usuario para realizar una verificación de las primitivas

seleccionadas, o bien, para crear un archivo en el que se visualice el conjunto de primitivas elegido de cualquiera de las siguientes formas: con los datos básicos de cada primitiva o con los datos básicos más uno de los ciclos más cortos que rompe cada primitiva en la red depurada.

La librería `circle` declara algunas de las clases y métodos usados por los módulos **graph**, **circle** y **visualizer**. Los archivos de entrada y los archivos de salida del sistema están en la Figura 13

### 3.3.1 Módulo **graph**

Este módulo es el que controla la depuración del grafo de entrada y la elección de un subconjunto de palabras que cortan los ciclos en el diccionario y nos permite obtener lo que llamaremos un "vocabulario definidor" minimal. Es así como **graph**, lee el grafo y aplica el algoritmo B llamando a un método del módulo **circle**. Asimismo, **graph** es el módulo que contiene en sus métodos el algoritmo A. Ambos algoritmos se mostraron en la sección 3.2. En la siguiente figura aparecen las partes que forman al módulo **graph**:

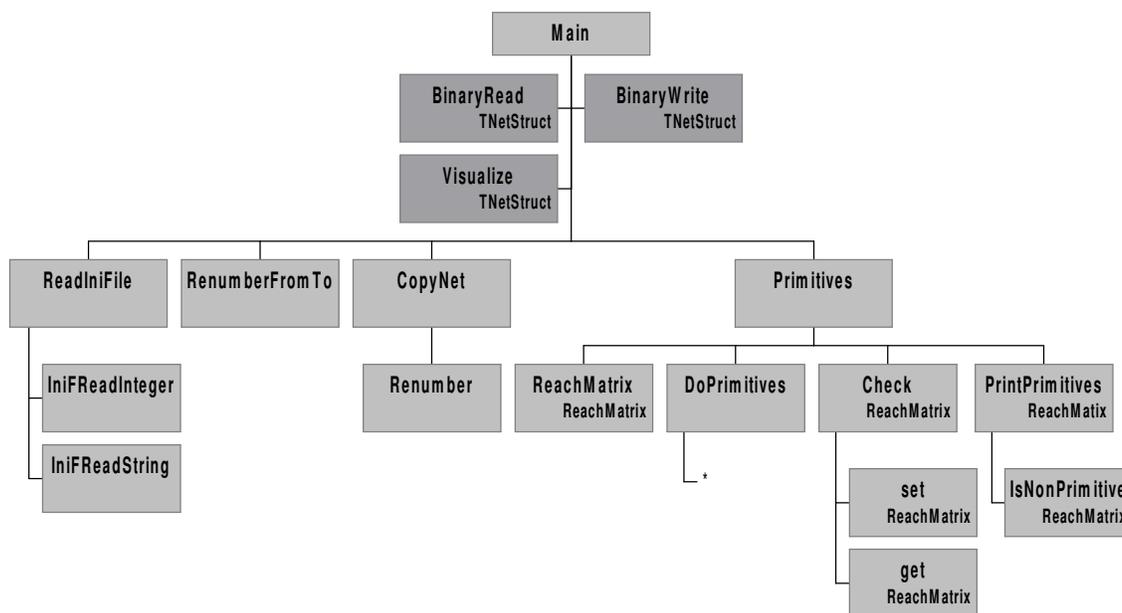


Figura 14. Módulo Graph (primera parte).<sup>11</sup>

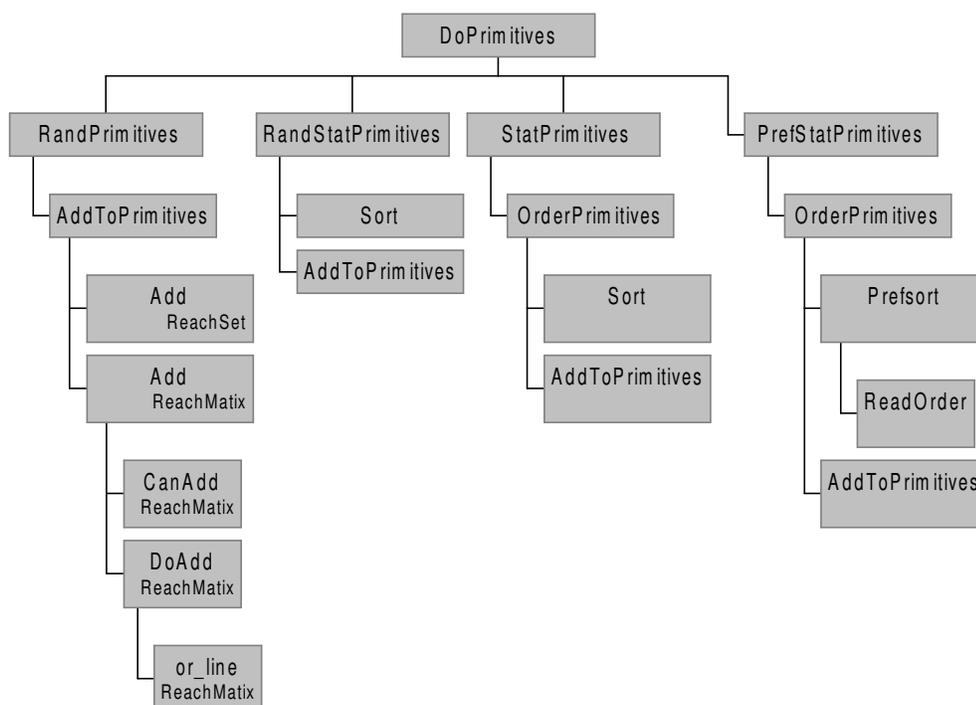


Figura 15. Módulo Graph (continuación).

<sup>11</sup> Los bloques de color más oscuro de estos diagramas representan métodos desarrollados en otros módulos.

El módulo **graph** puede recibir dos parámetros. El parámetro *c* indica que el programa tomará como grafo de entrada el contenido en el archivo **net.txt** el cual está en formato binario, para obtener como salida el grafo depurado **net1.txt**, el cual también está en formato binario, conforme al algoritmo B. En **net1.txt** se han eliminado todas aquellas palabras que no presentan ningún ciclo en su definición, dado que estas palabras, según nuestra aproximación, no serán consideradas como definidoras. Si no usamos la opción *c* entonces el archivo de entrada será **net1.txt** por defecto. El parámetro *v* sirve para crear el archivo **net1.vis**, el cual contiene las 10,358 entradas del diccionario que forman el grafo depurado, con los nodos a los que apuntan sus flechas de entrada y de salida, en un formato adecuado para su lectura, estas palabras crean conjuntos de nodos altamente interconectados. El programa **graph.exe** sin parámetros, creará otro archivo, llamado **primit.out**, el cual contendrá la lista de primitivas encontradas según alguno de los cuatro métodos o estrategias de ordenamiento.

La forma como trabaja el algoritmo A se explica a continuación. Como primer paso **graph** lee el archivo de inicio **graph.ini**, el cual contiene una cadena que le indica cuál método de selección de primitivas usar, por ejemplo: Algorithm = 4, lo cual indicaría que se trata del método 4, por votación aleatoria. Enseguida, si la opción elegida es *c*, es decir, *clean*, se hace una lectura binaria de la red **net** a través del método **BinaryRead** de la clase **TNetStruct**. Posteriormente ejecuta la función **Cleanup** que elimina a las entradas que no conducen a ciclos, ya que hemos considerado a éstas, por dicha característica, como no-primitivas. El siguiente paso del algoritmo es la inicialización de la red, es decir, la propiedad "*marcado*" de los nodos se establece como **false** para todos ellos, y reenumera para después copiar la parte de la red con la que va a trabajar al grafo **net1**, convirtiéndolo también a formato binario.

Este proceso de copiar una parte de la red **net** a **net1** no se realiza si se omite el parámetro *c* y por lo tanto se toma simplemente, la red **net1** como la

proveedora de posibles conjuntos de primitivas semánticas. Cuando el parámetro es  $v$ , el método para visualizar la red **net1** es **visualize** (módulo **circle**).

La función que obtiene las primitivas se llama **primitives** y tiene como parámetros a **net1** y al archivo de salida **primit.out**.

Los métodos a elegir son los siguientes<sup>12</sup>:

**Método 1: aleatorio uniforme (RandPrimitives).** Genera un número aleatorio entre los  $N$  nodos de la red, de manera que cada uno de los nodos sea evaluado como candidato a ser primitiva.

**Método 2: por frecuencias (StatPrimitives).** Ordenamos los vértices por el número de las flechas entrantes de menor a mayor y en ese orden se procesan los nodos para provocar que los nodos con mayor frecuencia sean los que queden en el conjunto de primitivas.

**Método 3: aleatorio por frecuencias (RandStatPrimitives).** Calcula la frecuencia máxima y genera un número aleatorio, pero con las probabilidades en función inversa a las frecuencias.

**Método 4: por votación aleatoria (PrefStatPrimitives).** Para este método genera 20 diferentes conjuntos definidores minimales  $P_i$  con el método 1, asociando así con cada vértice un número entre 20 y 0. Ordena primero los vértices que nunca entraron en los  $P_i$  usando sus frecuencias, como en el método 2, y después los que entraron, en el orden inverso al número de los  $P_i$  en que entró, es decir, desde 1 hasta 20.

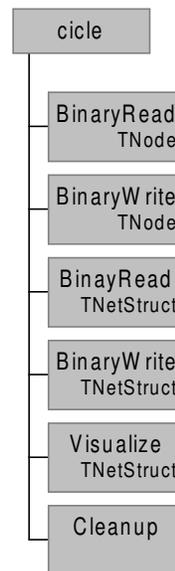
### 3.3.2 Módulo circle

En este módulo se desarrollan los métodos para hacer una escritura y lectura binaria sobre cada nodo y sobre la estructura de la red. La escritura binaria de cada nodo se realiza a través del método **BinaryWrite** de la clase **TNode**, por medio

---

<sup>12</sup> Estos métodos se detallarán mejor en la sección 4.1

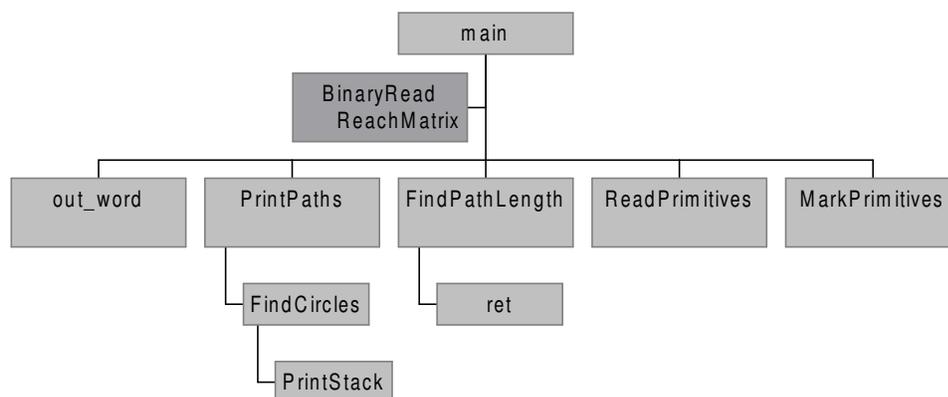
del cual se guarda la información esencial de cada nodo para la posterior depuración del grafo y para su procesamiento. El método **BinaryRead** de la clase **TNode** sirve para hacer una lectura desde el archivo **net1.txt** de las características de cada nodo. El método **BinaryWrite** de la clase **TNode** controla la escritura de los nodos en el archivo que contiene al grafo (depurado o no). El método **BinaryRead** de la clase **TNetStruct** abre el archivo que contiene el grafo y un leerá las características de cada nodo a través del método **BinaryRead** de la clase **TNode**. El método **cleanup** consiste en marcar los nodos que no tienen entradas o salidas, removiéndolos del grafo **net**; después escribe en un archivo cuántos nodos fueron removidos y cuántos se conservaron. Posteriormente el módulo **graph** copia el resultado en el archivo **net1.txt**.



**Figura 16. Módulo Circle.**

### 3.3.3 Módulo visualizer

El módulo visualizer puede recibir los siguientes parámetros: *l* indica modo de lista, por lo en este modo, el módulo toma el archivo de primitivas dado por el



**Figura 17. Módulo Visualizer.**

usuario como segundo parámetro (en estos archivos sólo está una lista con los identificadores de las primitivas halladas) y el nombre del archivo en el que escribirá las primitivas conforme al modo de lista. En este último escribe la frecuencia, el número de salidas, el número de entradas, la palabra y el identificador de cada primitiva (véase Anexo 2. Lista de las primitivas encontradas). Si no se da argumento alguno, el módulo tomará el archivo de entrada de primitivas dado por el usuario y lo procesará para obtener otro archivo, el cual mostrará, con respecto a cada primitiva, la longitud de uno de los ciclos más cortos, el número de salidas, el número de entradas, la palabra y la serie de palabras de dicho ciclo. Véase Anexo 3. Lista de los ciclos encontrados

### **3.3.4 Librería circle**

Esta librería contiene los atributos de la clase **TNode**, el constructor de dicha clase, la clase **TNetStruct** y el constructor de esa clase.

### **3.3.5 Explicación del código.**

A continuación describiremos las principales funciones que desempeñan los tres módulos que forman al sistema.

3.3.5.1 Módulo *graph*

El módulo **graph** define dos clases de datos, la clase **ReachSet** y la clase **ReachMatrix**.

<b>ReachSet</b> <sup>13</sup>
<b>Métodos públicos</b> ReachSet () ~ReachSet () GetItemsInContainer () Flush () Add (int n)
<b>Métodos privados</b> TArray<unsigned> s

**Clase ReachSet.-** Clase que representa a un contenedor de datos que se pueden agregar y quitar. Con este contenedor se realizaran las siguientes operaciones:

ReachSet () .- Crea un objeto de esta clase que es un conjunto *s* (0, 0, 1000) al que llamaremos contenedor.

~ReachSet () .- Destructor del objeto

GetItemsInContainer () .- Obtiene los datos del contenedor *s*.

Flush () .- Elimina los datos del contenedor *s*.

Add (int n).- Agrega el entero *n* al contenedor *s*, *s* representa un nodo que tiene una flecha entrante o una flecha saliente.

**Clase ReachMatrix .-** Clase que representa una matriz cuyos elementos se pueden agregar, quitar, revisar, etc. para finalmente identificar si un nodo es primitivo o

---

<sup>13</sup> Para describir las clases se utiliza en este trabajo la notación denominada *Unified Notation*

no. En el caso de que lo sea, el identificador de este nodo se escribe en un archivo. A continuación se detallan los métodos públicos y privados que se aplican a esta clase, sus parámetros y su funcionamiento:

<b>ReachMatrix</b>
<b>Métodos públicos</b>  ReachMatrix (unsigned size); ~ReachMatrix (); void Flush (); bool Add (unsigned n, const ReachSet & from_which, const ReachSet & which); bool Check () const; void PrintPrimitives (ostream & f) const; bool IsNonPrimitive (int i) const
<b>Métodos privados</b>  Bool CanAdd (unsigned n, const ReachSet & from_which, const ReachSet & which); Void DoAdd (unsigned n, const ReachSet & from_which, const ReachSet & which); Void or_line (unsigned i, unsigned j); // desde-hasta (from to) Void or_line (char *pi, char *pj); // desde.hasta (from to) Char *line (unsigned i) const; Void set (unsigned i, unsigned j); Void set (char *pi, unsigned j); Bool get (unsigned i, unsigned j) const; Bool get (char *pi, unsigned j) const; Unsigned cols; Unsigned lins; Buftype *m; Char *non_primitives;

Métodos públicos y sus parámetros:

### 2.1 ReachMatrix::ReachMatrix (unsigned size)

Método: ReachMatrix.

Clase: Reach Matrix.

Tipo: Constructor de la clase.

Parámetros: Recibe **size** que indica el número de nodos de la red.

Llamadas: no hay.

Constructor de la clase ReachMartix, determina el número de líneas y columnas en función al tamaño de la red (**size**), es decir, al número de nodos de la red.

## 2.2 ReachMatrix:: ~ReachMatrix (unsigned size)

Método: ~ReachMatrix.

Clase: Reach Matrix.

Tipo: Desctructor de la clase.

Parámetros: Recibe **size** que indica el número de nodos de la red.

Llamadas: no hay.

Destructor de la clase ReachMatrix

## 2.3 void ReachMatrix::Flush ()

Método: Flush().

Clase: Reach Matrix

Tipo: void

Parámetros: no hay.

Llamadas: no hay

Quita una cantidad de bytes en un bloque de memoria que era para la matriz **m** y para el arreglo **non\_primitives**.

2.4 bool ReachMatrix::Check () const

Método: Check().

Clase: Reach Matrix

Tipo: booleano

Parámetros: no hay.

Llamadas: no hay

Crea y llena un archivo "matrix.out", numera los renglones de 0 a lins/100 y la llena con ceros y unos. Después revisa la matriz, y si encuentra errores regresa "falso"; necesita devolver "verdadero" para que el programa continúe.

2.5 void ReachMatrix::or\_line (unsigned i, unsigned j)

Método: or\_line

Clase: Reach Matrix

Tipo: void

Parámetros: Recibe dos enteros positivos **i** y **j**.

Llamadas: no hay

Forma alternativa de manejar la memoria

2.6 void ReachMatrix::or\_line (char \*pi, char \*pj)

Método: or\_line

Clase: Reach Matrix

Tipo: void

Parámetros: Recibe dos apuntadores a carácter **i** y **j**.

Llamadas: no hay

Forma alternativa de manejar la memoria cuando se trata de caracteres.

2.7 char \*ReachMatrix::line (unsigned i) const

Método: line

Clase: Reach Matrix

Tipo: char

Parámetros: Recibe el entero positivo **i**

Llamadas: no hay

Recibe el identificador **i** de un nodo y devuelve un apuntador a carácter **line** con el valor **m+cols\*i**

2.8 void ReachMatrix::set (unsigned i, unsigned j)

Método: set

Clase: Reach Matrix

Tipo: void

Parámetros: Recibe dos enteros positivos **i** y **j**.

Llamadas: no hay

Recibe dos variable **i**, **j**, llama a **line** y con su apuntador a caracter y su número llama a la otra función **set**

2.9 void ReachMatrix::set (char \*pi, unsigned j)

Método: set

Clase: Reach Matrix

Tipo: void

Parámetros: Recibe un apuntador a carácter **pi** y un entero positivo **j**.

Llamadas: no hay

Reserva memoria

2.10 bool ReachMatrix::get (unsigned i, unsigned j) const

Método: get

Clase: Reach Matrix

Tipo: booleano

Parámetros: Recibe **i** y **j**.

Llamadas: no hay

Para sacar una dato de la estructura matriz, primero resuelve un apuntador a través de la función **line**.

2.11 bool ReachMatrix::get (char \*pi, unsigned j) const

Método: get

Clase: Reach Matrix

Tipo: booleano

Parámetros: Recibe el apuntador a caracteres **pi** y el entero positivo **j**.

Llamadas: no hay

Recibe un apuntador a caracter **pi**, y un **id** de nodo **j** para reservar memoria

2.12 bool ReachMatrix::Add (unsigned n,const ReachSet & from\_which,  
const ReachSet & which)

Método: Add

Clase: Reach Matrix

Tipo: booleano

Parámetros: un nodo **n**, conjunto de nodos entrantes a **n** llamado **from\_which** y un conjunto de nodos salientes de **n**, llamado **which**.

Llamadas: CanAdd, DoAdd

Recibe el nodo que queremos agregar como primitiva semántica **n**, a los conjuntos de nodos entrantes y salientes, y evalúa si es factible agregar dicho nodo llamando a la función **CanAdd** y si es así ejecuta **DoAdd** y regresa el valor de **CanAdd**.

```
2.13 bool ReachMatrix::CanAdd (unsigned n, const ReachSet &
    from_which, const ReachSet & which)
```

Método: CanAdd

Clase: Reach Matrix

Tipo: booleano

Parámetros: un nodo **n**, conjunto de nodos entrantes a **n** llamado **from\_which** y un conjunto de nodos salientes de **n**, llamado **which**.

Llamadas: no hay

Recibe un nodo **n**, y los conjuntos de nodos entrantes y salientes representados por **from\_which** y **which**. Se buscan los casos en que no es posible considerar primitiva al nodo y en los demás casos regresa el valor de verdad **true**.

Verifica que pueda agregarse un nodo como primitiva semántica. Regresa False en los casos que se listan a continuación:

- Si la palabra se define con la misma palabra, es decir, si entre los nodos de flechas salientes se encuentra el mismo nodo el nodo **n** no puede agregarse como primitiva.
- Si ninguno de los nodos de flechas entrantes o ninguno de los nodos de flechas salientes al nodo en cuestión, es considerado como primitiva, esto es, si todos los nodos apuntados por flechas desde **n** son No-Primitivas, el nodo **n** tampoco lo es.

2.14 void ReachMatrix::DoAdd (unsigned n, const ReachSet & from\_which, const ReachSet & which)

Método: DoAdd

Clase: Reach Matrix

Tipo: void

Parámetros: un nodo **n**, conjunto de nodos entrantes a **n** llamado **from\_which** y un conjunto de nodos salientes de **n**, llamado **which**.

Llamadas: no hay

Realiza la acción de agregar primitiva

2.15 void ReachMatrix::PrintPrimitives (ostream & f) const

Método: PrintPrimitives

Clase: Reach Matrix

Tipo: void

Parámetros: un archivo de salida **f**

Llamadas: no hay

Recibe un archivo en el que escribirá el identificador de nodo y si el nodo no esta considerado como primitivo, es decir, si la función **IsNonPrimitive**

que a su vez verifica que el nodo **i** esté en la cadena **non\_primitives**, es falsa, escribe el identificador del nodo en dicho archivo, que es **"Primit.out"**

Escribe los identificadores de las palabras primitivas en el archivo **"primit.out"**

2.16 bool IsNonPrimitive (int i)

Método: IsNonPrimitive

Clase: Reach Matrix

Tipo: booleano

Parámetros: entero **i**

Llamadas: no hay

Función booleana que regresa **True** si encuentra al nodo **i** en el conjunto denominado **non\_primitives**.

1.17 void RenumberFromTo (const TNetStruct & net, TArray<unsigned> \*from, TArray<unsigned> \*to)

Método: RenumberFromTo

Clase: ninguna

Tipo: void

Parámetros: la estructura de red net, el arreglo from y el arreglo to.

Llamadas: no hay

Si un nodo de **net** no está marcado, éste método le asigna los **ItemsIncontainer** del arreglo **to** al apuntador **from** e incremento contador

**unmarked**, de otra forma le agrego la constante `UINT_MAX` a dicho apuntador e incrementa el contador **marked**.

1.18 void Renumber (int \*dst, const int \*src, const TArray<unsigned> & from, unsigned n, int \*pn)

Método: Renumber.

Clase: ninguna.

Tipo: booleano

Parámetros: dos apuntadores a entero, **dst** (destino) y **src** (origen), un arreglo **from**, un entero positivo **n** y un apuntador a entero **pn**.

Llamadas: no hay

Reenumera restando 1 a cada número del arreglo **src** y copiándolo al arreglo **dst**.

1.19 void CopyNet (const TNetStruct & net, TNetStruct \* net1, const TArray<unsigned> & from, const TArray<unsigned> & to)

Método: CopyNet.

Clase: ninguna.

Tipo: void.

Parámetros: una red **net**, otra red **net1**, un arreglo **from** y un arreglo **to**.

Llamadas: no

Copia cada nodo de **net** especificado en el arreglo **to** a **net1**.

1.20 void sort (const TNetStruct & net, freq\_id \*freq)

Método: sort.

Clase: ninguna.

Tipo: void

Parámetros: Una red **net** y una estructura **freq\_id**

Llamadas: no

Ordena por la frecuencia

1.21 void read\_order (int p [], unsigned N)

Método: read\_order

Clase: ninguna.

Tipo: void.

Parámetros: recibe un arreglo de entros **p** y un número positivo **N**.

Llamadas: no

Lee el orden de preferencia del archivo **primit.prf**

1.22 void prefsort (const TNetStruct & net, freq\_id \*freq)

Método: prefsort

Clase: ninguna.

Tipo: void.

Parámetros: una red **net**, una estructura **freq\_id**.

Llamadas: no

Para el método 4, calcula la máxima frecuencia de flechas entrantes a un nodo y asigna una frecuencia para cada nodo en función de la frecuencia calculada y la frecuencia máxima

1.23 void AddToPrimitives (ReachMatrix \*m, const TNetStruct & net, unsigned i)

Método: AddToPrimitives

Clase: ninguna.

Tipo: void

Parámetros: Una matriz **m**, una red **net**, un entero positivo **i**

Llamadas: no

Recibe la matriz **m**, la estructura net y el número de nodo **i**. Examina el nodo **i** como posible candidato a ser una primitiva semántica. Para ello mantiene dos conjuntos, uno tiene los identificadores de los nodos que tienen flechas entrantes al nodo en cuestión (**from\_which**) y el otro tiene los de los nodos que tienen flechas salientes de dicho nodo (**which**). Después llama a un método **Add** de la matriz **m** con tres parámetros: el nodo, el conjunto de nodos con flechas entrantes y el conjunto de nodos con flechas salientes.

1.24 void (\*sort\_fun) (const TNetStruct &, freq\_id \*);

Método: \*sort\_fun.

Clase: ninguna.

Tipo: void.

Parámetros: una constante TnetStruct y un número de freq\_id.

Llamadas: no

Ejecuta alguna de las funciones: **sort** o **prefsort** dependiendo del valor de la variable **sort\_fun**.

1.25 void OrderPrimitives (ReachMatrix \*m, const TNetStruct & net, unsigned N)

Método: OrderPrimitives

Clase: ninguna.

Tipo: void.

Parámetros: una matriz **m**, una red **net**, un entero positivo **N**.

Llamadas: no

Ordena conforme al identificador de la palabra

1.26 void RandPrimitives (ReachMatrix \*m, const TNetStruct & net, unsigned N)

Método: RandPrimitives.

Clase: ninguna.

Tipo: void.

Parámetros: una matriz **m**, una constante **net**, un número positivo **N**.

Llamadas: no

METODO 1: aleatorio, uniforme

Parámetros que recibe: la matriz **m**, la red **net** y el número total de nodos de la red, **N**.

Genera una serie de números aleatorios dentro del rango que caracteriza al tamaño de la red. Después llama a **AddToPrimitives**

1.27 void StatPrimitives (ReachMatrix \*m, const TNetStruct & net, unsigned N)

Método: StatPrimitives.

Clase: ninguna.

Tipo: void.

Parámetros: una matriz **m**, una constante **net**, un número positivo **N**.

Llamadas: no

METODO 2: Realiza el análisis en el orden definido **sort**.

Parámetros que recibe: la matriz **m**, la red **net** y el número total de nodos de la red, **N**. Obtiene las frecuencias de cada uno de los nodos en la red dada y los ordena de menor a mayor.

```
1.28 void RandStatPrimitives (ReachMatrix *m, const TNetStruct & net,  
    unsigned N)
```

Método: RandStatPrimitives.

Clase: ninguna.

Tipo: void.

Parámetros: una matriz **m**, una constante **net**, un número positivo **N**.

Llamadas: no

METODO 3 : Aleatorio, por frecuencias

Parámetros que recibe: la matriz **m**, la red **net** y el número total de nodos de la red, **N**.

```
1.29 void PrefStatPrimitives (ReachMatrix *m, const TNetStruct & net,  
    unsigned N)
```

Método: PrefStatPrimitives

Clase: ninguna.

Tipo: void.

Parámetros: una matriz **m**, una constante **net**, un número positivo **N**.

Llamadas: no

MÉTODO 4: Realiza el análisis en el orden definido por **sort\_fun = prefsort**.  
Parámetros que recibe: la matriz **m**, la red **net** y el número total de nodos de la red, **N**.

1.30 void ErrorPrimitives (ReachMatrix \*, const TNetStruct &, unsigned)

Método: ErrorPrimitives

Clase: ninguna.

Tipo: void.

Parámetros: una matriz **m**, una constante **net**, un número positivo **N**.

Llamadas: no

Si se da una opción distinta al elegir el método, **ErrorPrimitives** envía el mensaje siguiente: "ERROR: No se especificó el método."

1.31 void Primitives (const TNetStruct & net, ostream & f, void  
(\*DoPrimitives) (ReachMatrix \*, const NetStruct &, unsigned))

Método: Primitives.

Clase: ninguna.

Tipo: void.

Parámetros: una constante **net**, un archivo de salida **f**, una función determinada por **DoPrimitives**.

Llamadas: no

Parámetros : **net1**, **ostream** (el archivo "**primit.out**"), y **method\_name**.

Recibe una estructura de red **net**, un archivo de salida **f** y el tipo de algoritmo o método de procesamiento de la red elegida, al que a su vez le define tres parámetros ( la matriz **m**, la red **net** y el número total de nodos de la red, **N**)

Llama a: **DoPrimitives**. Esta es sólo una variable que contiene el número de método elegido. Puede ser: **RandPrimitives**, **RandStatPrimitives**, **StatPrimitives** o **PrefStatPrimitives**. También llama a los métodos de ReachMatrix : **Check** y **PrintPrimitives**

Este método crea una matriz **m** de tamaño **N**, identifica las primitivas a través del método elegido (ejecuta el método, pasándole 3 parámetros, la matriz **m**, la red **net** y el número de nodos de la red, **N**) y escribe, tomando como base la misma matriz, las primitivas.

1.32 AnsiString IniFReadString (const char \*file, const char \*name)

Método: IniFReadString.

Clase: ninguna.

Tipo: AnsiString

Parámetros: un apuntador a caracter file, un apuntador a caracter name.

Llamadas: no

Regresa un entero que indica cuál será el método a ejecutar.

1.33 int IniFReadInteger (const char \*file, const char \*name)

Método: IniFReadInteger

Clase: ninguna.

Tipo: int

Parámetros: un apuntador a caracter file, un apuntador a caracter name.

Llamadas: IniFReadString

Lee el archivo "**Graph.ini**" que especifica el método a ejecutar.

1.34 void ReadIniFile ()

Método: ReadIniFile

Clase: ninguna.

Tipo: void.

Parámetros: ninguno

Llamadas: IniFReadInteger

Asigna a la variable **algorithm** el número de algoritmo especificado en el archivo "**Graph.ini**".

1.35 int main(int argc, char\* argv[])

Método: main

Clase: ninguna.

Tipo: int

Parámetros: Entre sus argumentos está -c y -v

Llamadas: no

Función main, evalúa los argumentos recibidos.

### **3.3.5.2**      *Módulo **circle***

1. void TNode::BinaryWrite (ofstream &f)

Método: BinaryWrite

Clase: TNode

Tipo: void

Parámetros: archivo de salida **f**

Llamadas: ninguna.

Escribe el identificador de cada nodo, su longitud, si esta marcado o no, el identificador de cada nodo hacia el que el nodo en cuestión, tiene flechas salientes, y también todos los identificadores de los nodos con flechas entrantes hacia dicho nodo.

2. void TNode::BinaryRead (ifstream &f)

Método: BinaryRead.

Clase: TNode.

Tipo: void

Parámetros: archivo de entrada **f**

Llamadas: ninguna

Lectura binaria de las características de cada nodo en **f**.

3. void TNetStruct::BinaryWrite (AnsiString name)

Método: BinaryWrite.

Clase: TNetStruct.

Tipo: void.

Parámetros: El nombre de un archivo de salida.

Llamadas: TNode::BinaryWrite

Realiza la escritura de un archivo binario **name** a través del método **BinaryWrite** de la clase **TNode**.

4. void TNetStruct::BinaryRead (AnsiString name)

Método: BinaryRead

Clase: TNetStruct

Tipo: void

Parámetros: nombre de archivo **name**.

Llamadas: TNode::BinaryRead

Lectura binaria del archivo **name**, usando el método **BinaryRead** de la clase **node**.

5. void TNetStruct::Visualize (AnsiString fname)

Método: Visualize.

Clase: TNetStruct.

Tipo: void

Parámetros: el nombre del archivo de salida a visualizar.

Llamadas:

Recibe como parámetro el archivo "**net1.vis**" que contiene el diccionario procesado con entradas y salidas, en un formato legible por el usuario, con las 10358 entradas.

6. void Cleanup (TNetStruct \*net, ostream \*f)

Método: Cleanup.

Clase: ninguna..

Tipo: void

Parámetros: Una red **net**, un archivo de salida **f**

Llamadas: ninguna.

Depura la estructura de red de los nodos marcados y registra el número de nodos removidos y los que quedan en el archivo **f**.

### **3.3.5.3**      *Módulo **visualizer***

1. void ReadPrimitives (TArray<unsigned> \*primitives, istream &f)

Método: ReadPrimitives.

Clase: ninguna.

Tipo: void

Parámetros: el arreglo **primitives**, la entrada **f**.

Llamadas: ninguna.

Lee el archivo **f**, encuentra las primitivas y las agrega al arreglo **primitives**.

2. void MarkPrimitives (const TArray<unsigned> & primitives,const TNetStruct & net1)

Método: MarkPrimitives.

Clase: ninguna.

Tipo: void

Parámetros: el arreglo **primitives**, la red **net1**.

Llamadas: ninguna.

Desmarca todos los nodos de **net1** y vuelve a marcar solo los nodos que son primitivas.

3. void out\_word (const char \*s1,unsigned n,const TNetStruct & net1,ostream &f,const char \*s2)

Método: out\_word.

Clase: ninguna.

Tipo: void.

Parámetros: el apuntador a un arreglo **s1**, el número positivo **n**, la red **net1**, el archivo de salida **f** y un apuntador a caracter **s2**.

Llamadas: ninguna.

Imprime la palabra a que se refiere el nodo **n** y su identificador.

```
4. int FindPathLength (unsigned n,const TNetStruct & net,TArray<unsigned> *  
    path,ostream *trace)
```

Método: FindPathLength

Clase: ninguna.

Tipo: int

Llamadas: ninguna.

Tomando como base la red **net**, construye la trayectoria del ciclo que se forma a partir del nodo **n**, con los identificadores de los nodos del ciclo.

```
5. ret
```

Método: ret

Clase, tipo, parámetros y llamadas: no.

Agrega un elemento individual al arreglo **path**.

```
6. int main(int argc, char* argv[])
```

Método: main

Clase: ninguna.

Tipo: int

Llamadas: ninguna.

Evalúa los argumentos, crea un archivo de salida con las primitivas y los ciclos más pequeños que se relacionan a ellas.

**3.3.5.4 Librería *circle***

<b>Clase TNode</b>
<b>Atributos</b> id (n.id), word (n.word), n_of_out (n.n_of_out), dumb (n.dumb), marked (n.marked), tmp_marked (n.tmp_marked), n_pass (n.n_pass), net (n.net), in (0, 0, 100), subst (0, 0, 100), used (0, 0, 100)
<b>Métodos públicos</b>  void AddOut (int add_id) void AssignOutToSubst (); void BinaryWrite (ofstream &f); void BinaryRead (ifstream &f); int FindCircles (int id_where, int depth); bool NoCircles (int id_where, int depth); int NumOfReachable (); int NumOfReachableStupid (ofstream &f, bool inverse); int DoNumOfReachable (int depth); void CalcFreqs (bool inverse); bool HasElement (int element); void Substitute (int start_level, int max_level, int level); void IncFreq (int id, int level, int max_level); int n_of_out; int out [max_out]; TArray <TInt> in; TArray <TInt> used; TArray <TInt> subst; bool dumb; bool marked;

```
bool tmp_marked;  
int n_pass;  
TNetStruct * net;
```

#### **Clase TnetStruct<sup>14</sup>**

##### **Métodos públicos**

```
TNetStruct ()  
~TNetStruct ()  
void EmptyNodes ()  
void FindMaxCircles (ofstream & f);  
void DeleteNonTrans (ofstream & f);  
void AssignOutToSubst ();  
void BinaryWrite (AnsiString name);  
void BinaryRead (AnsiString name);  
void Visualize (AnsiString fname);  
void VisualizeSubst (AnsiString fname, bool JustWords);  
void Substitute (ofstream & f);
```

---

<sup>14</sup> Para describir las clases se utiliza en este trabajo la notación denominada *Unified Notation*

## 4 Metodología experimental y resultados obtenidos

---

### 4.1 Metodología experimental

Una vez que el diccionario tradicional ha sido convertido dos tipos distintos de grafos (por lexemas y por significados), hicimos los siguientes experimentos, cuyos resultados se presentan en la siguiente sección y, específicamente, se muestran en la Tabla 2.

Como se explica en la sección 3.2 el resultado del algoritmo depende del ordenamiento de los vértices del grafo. Se deben probar diferentes ordenamientos para obtener el menor conjunto definidor posible, pues no sabemos de ningún algoritmo que encuentre el conjunto de menor tamaño. Probamos los siguientes ordenamientos.

*Método 1: aleatorio, uniforme.* Usamos el ordenamiento uniformemente aleatorio: en cada iteración del algoritmo A, elegimos aleatoriamente el siguiente vértice  $\sigma(i)$  de los vértices todavía no procesados.

*Método 2: por frecuencias.* Ordenamos los vértices por la frecuencia de su uso en las definiciones del mismo diccionario –es decir, por el número de las

flechas entrantes– de menor a mayor. Entonces, los vértices con menor frecuencia tendieron a entrar en **G'** y los de mayor frecuencia tendieron a ser definidores y entrar en **P**. Esperábamos que con esta heurística, **P** sería menor porque los vértices que lo forman rompen más ciclos en **G**.

*Método 3: aleatorio, por frecuencias.* Este método es una combinación de los métodos 1 y 2. Usamos el ordenamiento aleatorio, pero con las probabilidades en función inversa a las frecuencias: en cada iteración del algoritmo A, elegimos el siguiente vértice  $\sigma(i)$  aleatoriamente de los vértices todavía no procesados, siendo la probabilidad del vértice  $i$  de ser elegido  $p_i \sim f_{max} - f_i + 1$ , donde  $f_i$  es la frecuencia del vértice como en el método 2,  $f_{max}$  es el valor máximo de  $f_j$  y  $\sim$  significa proporción con normalización,  $\sum p_i = 1$ . Esperábamos que alguna alternación del ordenamiento rígido del método 2 produciría un conjunto menor.

*Método 4: por votación aleatoria.* En este método, generamos 20 diferentes conjuntos definidores minimales  $\mathbf{P}_i$  con el método 1, y para cada vértice, contamos el número de los conjuntos  $\mathbf{P}_i$  en los cuáles éste entra, asociando así con cada vértice un número entre 20 (entró en todos) y 0 (nunca entró). Ordenamos primero los vértices que nunca entraron en los  $\mathbf{P}_i$  usando sus frecuencias, como en el método 2, y después los que entraron, en el orden inverso al número de los  $\mathbf{P}_i$  en que entró, desde 1 a 20. Esperábamos que los que entraron en más conjuntos  $\mathbf{P}_i$  serían los “mejores” definidores y debieran entrar en el conjunto que buscamos.

Cabe mencionar, que un 80% de los vértices entraron en por lo menos un conjunto  $\mathbf{P}_i$ .

Hicimos dos experimentos con dos grafos diferentes.

*Grafo 1: por lexemas.* En este experimento, consideramos un lexema como un vértice del grafo, es decir, una palabra normalizada morfológicamente. Por ejemplo, las cadenas *piensa*, *pensó*, *pensaríamos* se consideraron equivalentes y se representaron con el vértice *pensar*. No aplicamos ninguna resolución de

ambigüedad: así, la cadena *fue* se consideraba como si en lugar de ésta, en el texto estuviera *ir ser*, es decir, se formaron las flechas desde la palabra encabezado de la definición que contiene *fue* hasta ambos vértices, *ir* y *ser*. Por supuesto, como las entradas ya están en la forma correcta, no se cambiaron. En este método, las definiciones de todos los significados de la palabra –o, en su caso, de todas las entradas homónimas– se consideraron como una larga definición, pues no hay método para distinguir a qué significado u homónimo refiere una palabra en una definición.

*Grafo 2: por significados.* En este experimento, consideramos un vértice del grafo como un significado específico de la palabra, por ejemplo:  $gato_{1a}$  (animales),  $gato_{1b}$  (tipo de animales),  $gato_{2a}$  (herramienta) etc. Para desambiguar los significados de las palabras que forman las definiciones, empleamos un etiquetador (*tagger*) para la normalización morfológica y para la desambiguación de la parte de oración, y después utilizamos un algoritmo parecido al de Lesk para desambiguar el significado de la palabra en el contexto. Con más detalle este proceso se describe en (Sidorov y Gelbukh, 2002).

En ambos casos, sólo se consideran las palabras significativas, es decir, no se consideran los artículos, preposiciones, conjunciones y verbos auxiliares. Nótese que en el Grafo 2, el número de las flechas es exactamente el mismo que el número de las palabras significativas en las definiciones del diccionario, mientras que en el Grafo 1 este número es ligeramente mayor por la ambigüedad léxica.

## **4.2 Resultados obtenidos**

Para nuestros experimentos usamos el Diccionario de la lengua española del grupo Anaya (1997). Este diccionario contiene 30971 entradas, de los cuales sólo 30725

son palabras significativas. Estas palabras significativas se dividen en 60818 significados específicos<sup>15</sup>. Estas cifras se muestran en la Tabla 2.

Los resultados de estos experimentos y su discusión se presentan a continuación.

<b>Grafo</b>	<b>En total</b>	<b>No definidores</b>	<b>Depurado</b>
Lexemas	30725	20366	<b>10359</b>
Significados	60818	47802	<b>13016</b>

**Tabla 2. Número de vértices en los grafos.**

<b>Grafo</b>	<b>Método 1 Aleatoriamente, uniforme</b>	<b>Método 2 Por frecuencias</b>	<b>Método 3 Aleatoriamente, por frecuencias</b>	<b>Método 4 Por votación aleatoria</b>
Lexemas	2789, $s = 25$	2302	2770	<b>2246</b>
Significados	2266, $s = 28$	1955	2257	<b>1913</b>

**Tabla 3. Número de definidores, con diferentes algoritmos.**

Al aplicar el algoritmo de depuración (algoritmo B) a las dos variantes del grafo (por palabras y por significados), encontramos que sólo aproximadamente unos 10 mil vértices están, posiblemente, involucrados en los ciclos (véase Tabla 2).

Un resultado inesperado fue que este núcleo de unos 10 mil vértices está muy fuertemente interconectado: prácticamente con cualquier palabra del diccionario, están accesibles, a través de alguna ruta, todas las 10 mil palabras de este núcleo. O sea, al sustituir iterativamente las palabras en la definición con sus propias definiciones, la definición de casi cualquier palabra del diccionario se

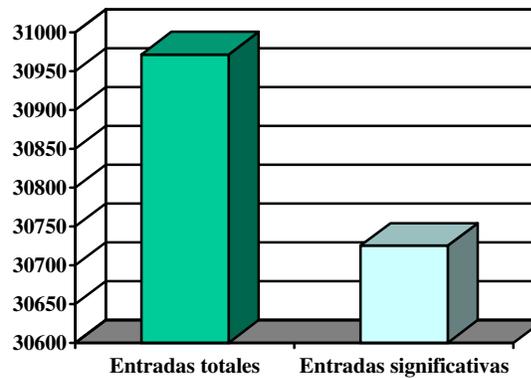
---

<sup>15</sup> En realidad, nuestro algoritmo para la detección de los significados no es perfecto, pero sus errores no son estadísticamente significativos

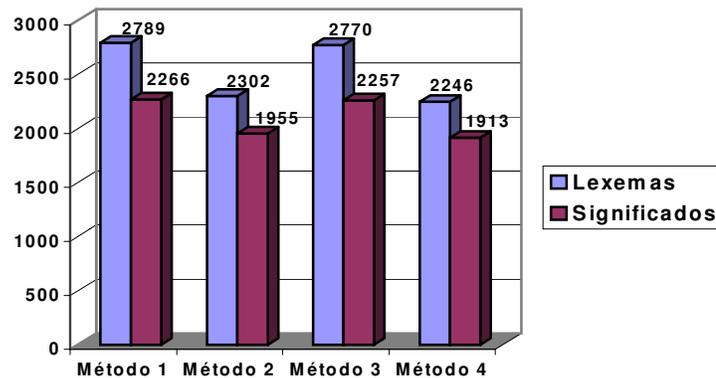
extiende a las mismas 10 mil palabras. Entonces, la tarea de selección del conjunto definidor en un grafo tan fuertemente interconectado es computacionalmente difícil.

A estos dos conjuntos de 10 mil palabras, aplicamos el algoritmo A con las 4 variantes de ordenamiento. Los resultados se presentan en la Tabla 3, mostrándose los tamaños de los conjuntos definidores obtenidos.

Enseguida tenemos podemos observar gráficamente estos resultados:



Gráfica 1. Entradas totales contra entradas significativas



Gráfica 2. Resultados obtenidos por grafo y por método.

Para el método 1, se muestra el promedio de los 20 experimentos y la desviación cuadrática promediada  $s$  ( $s^2$  es la dispersión; se sabe que un 67% de los casos se desvían del promedio no más que una  $s$  y 99% no más que  $3s$ ). Atrae la atención que la desviación es muy baja, lo que significa que con el método 1, los tamaños de los conjuntos obtenidos son diferentes pero muy parecidos.

Para el método 3, sólo mostramos el resultado de un experimento. Fue sorprendente que estos resultados casi no difieran de los obtenidos con el método 1, a pesar de que las probabilidades en este caso correspondieron a las frecuencias del método 2. Este último mostró un muy buen desempeño produciendo los conjuntos definidores mucho más pequeños. Sin embargo, el método heurístico – método 4– produjo los conjuntos más pequeños que hemos obtenido.

Aunque creemos que con métodos más sofisticados, se puede obtener conjuntos aun más pequeños, no esperamos que el tamaño mínimo del conjunto definidor sea mucho menor que los que hemos obtenido, de aproximadamente 2000 palabras. Esto se debe a las siguientes consideraciones lingüísticas: según la opinión común, se supone que 2 mil es el número de las palabras primitivas, por lo menos en el sentido común, suficientes para definir todas las demás palabras del vocabulario general. Por ejemplo, éste es el número de palabras en el vocabulario definidor de Longman (*Longman defining vocabulary*). También, según nuestro conocimiento, éste es el número de los jeroglíficos en el vocabulario chino básico. Creemos que el hecho de que el tamaño de nuestro conjunto definidor tan exactamente corresponda a la cifra prevista –2 mil– es muy significativo.

Consideremos unos ejemplos de las palabras elegidas con el Método 4, Grafo 1. Aquí están las 20 palabras con la mayor frecuencia de las flechas entrantes (las mejores):

*cosa, persona, acción, hacer, efecto, tener, parte, no, conjunto, dar, forma, cierto, cuerpo, relativo, nombre, poder, uno, formar, producir, animal, común, general, determinado, poner, estado, tiempo, decir, planta, obra, etc.*

Estas palabras son buenas candidatas a primitivas semánticas. También hay palabras que entraron en el conjunto definidor porque están involucradas en ciclos cortos, aunque tienen una frecuencia baja (las peores):

*almuerzo, almuédano, almanaque, alinear, algarroba, alarmar, ahíto, etc.*

Se supone que estas palabras son buenos indicadores de la necesidad de cambiar las definiciones en el diccionario para que se excluyan de la lista de las definidoras.

Finalmente, algunas palabras tienen que estar en cualquier conjunto definidor pues tienen lazos. Éstas indican que hay algún problema con el algoritmo de la identificación de las palabras, o que hay una mala definición en el diccionario; en nuestro caso son 47:

*ático, borgoña, lapón, etc.*

Es interesante investigar la longitud de los ciclos en los cuales estaban involucradas las palabras definidoras si fuesen insertadas en el diccionario, es decir, estos ciclos no contienen las demás palabras del conjunto definidor elegido. Ejemplos de tales ciclos son:

1: *ático* → *ático*

2: *premura* → *prisa* → *premura*

3: *grano* → *cereal* → *centeno* → *grano*

etc. Las longitudes de los ciclos en nuestro ejemplo se distribuyeron como sigue (donde LC es la longitud del ciclo más corto y N es el número de las palabras en tales ciclos):

L	N	L	N	L	N
1	47	7	53	13	19
2	1496	8	58	14	9
3	177	9	45	15	8
4	67	10	38	16	12
5	47	11	29	17	11
6	72	12	32	18	3

**Tabla 4. Longitudes de los ciclos más cortos (ejemplo).**

Sólo mostramos aquí los primeros 18 elementos. El ciclo más largo fue de longitud 52.

Al final tenemos cuatro conjuntos de "primitivas semánticas" obtenidos con los distintos métodos (aleatorio, por frecuencias, aleatorio por frecuencias y por votación aleatoria) para cada grafo. En la Tabla 5 podemos ver que 478 palabras aparecieron en los cuatro conjuntos obtenidos para el Grafo 2 (por significados), las características de estas palabras se muestran en la Tabla 8 y en la Tabla 9 .

EN 0 CONJUNTOS	8819
EN 1 CONJUNTO	1762
EN 2 CONJUNTOS	1152
EN 3 CONJUNTOS	805
EN 4 CONJUNTOS	478
<b>TOTAL</b>	<b>13016</b>

**Tabla 5. Número de palabras que aparecen en los distintos conjuntos.**

En el Grafo 2 depurado (sin las palabras con lazos, y sin palabras sin flechas de entrada o flechas de salida) tenemos la composición mostrada en la Tabla 6.

Adverbios	115
Adjetivos	2563
Verbos	2868
Sustantivos	7470
<b>TOTAL</b>	<b>13016</b>

**Tabla 6. Tipos de palabras del grafo depurado.**

En la Tabla 7 tenemos los tipos de palabras de cada conjunto obtenido para el Grafo 2. Podemos observar que la diferencia entre estos conjuntos estuvo determinada por los sustantivos.

<b>Método:</b>	<b>1. Aleatorio</b>	<b>2. Por frecuencias</b>	<b>3. Aleatorio por frecuencias</b>	<b>4. Por votación aleatoria</b>
Adverbios	18	20	22	15
Adjetivos	358	280	367	295
Verbos	416	332	421	337
Sustantivos	1476	1323	1447	1266
<b>TOTAL</b>	<b>2268</b>	<b>1955</b>	<b>2257</b>	<b>1913</b>

**Tabla 7. Tipos de palabras de cada conjunto de primitivas semánticas obtenidas.**

Adverbios	7
Adjetivos	72
Verbos	51
Sustantivos	348
<b>TOTAL</b>	<b>478</b>

**Tabla 8. Tipos de las 478 palabras que están en los 4 conjuntos obtenidos.**

<b>LC</b>	<b>Número de palabras</b>
1	140
2	289
3	11
4	1
5	4
6	3
7	2
8	2
9	5
10	6
11	3
12	3
13	3
14	1
15	2
16	1
17	1
36	1
<b>TOTAL</b>	<b>478</b>

**Tabla 9. Número de palabras del conjunto de primitivas comunes a los 4 conjuntos ordenados la longitud de uno de sus ciclos (LC).**

Conviene mencionar que la mayoría de estas 478 palabras comunes, están involucradas en ciclos de longitud uno o dos, como se muestra en la Tabla 9.

Por último, tenemos un ejemplo de lo que sucede con algunas palabras que denotan parentesco entre personas. Seleccionamos algunas palabras del grafo por significados, ya depurado (palabras como "hijo", "hermano", "madre", etc., las cuales suponemos - y comprobamos - que son palabras con ciclos cortos). En la Tabla 10, que se encuentra a continuación, tenemos una lista de estas palabras. El método 4, como ya se explicó en la sección 4.1 , consiste en obtener con el método 1 (método aleatorio), 20 conjuntos de primitivas y después determinar el número de veces que cada palabra del grafo aparece en estos conjuntos, ordenando de

menor a mayor este criterio, y enseguida tomar como segundo criterio de ordenamiento la frecuencia de las palabras en el grafo (sus flechas entrantes), también de menor a mayor. Esto permite que las palabras que más hayan aparecido en los distintos conjuntos de primitivas obtenidas por el método aleatorio, seguidas de las que tienen una frecuencia alta de uso en las definiciones, son las que tenderán en mayor medida a quedar en el nuevo conjunto de primitivas obtenidas por el método 4. Veamos cuales son las palabras que más aparecen en los conjuntos P obtenidos con el método 1, dato seguido por su frecuencia de uso en las definiciones:

<b>Palabra</b>	<b>Número de veces que aparece en 20 conjuntos P obtenidos por el método 1</b>	<b>Frecuencia</b>
Antecesor	0	1
Cónyuge	0	1
Pariente	0	1
Hermano	0	1
Marido	0	2
Antepasado	0	3
Descendiente	0	5
Novio	1	1
Sobrino	1	1
Tío	2	2
Esposo	3	1
Primo	3	2
Parentela	5	2
Padre	9	4
Parentesco	9	6
Mujer	10	12
Madre	12	1
Familia	12	4
Hijo	13	4

**Tabla 10. Algunos ejemplos de datos de palabras relacionadas y sus frecuencias.**

Aplicando el método 4, es decir, ordenando el procesamiento de las palabras del diccionario primero por el número de veces que aparecen en 20

conjuntos P obtenidos con el método 1 y luego por la frecuencia de la palabra en las definiciones del diccionario, obtenemos un conjunto de 1913 primitivas entre las que se encuentran:

---

---

Parentesco  
Padre  
Mujer  
Madre  
Familia  
Hijo

---

---

**Tabla 11. Algunas primitivas halladas por el método 4.**

Como se ve en esta lista, las palabras que más veces aparecen en los conjuntos P generados por el método aleatorio son las que tienden a entrar en nuestro conjunto final de primitivas generadas con el método 4.

## 5 Conclusiones

---

Presentamos un método y el algoritmo correspondiente, del conjunto definidor formado por palabras a través de las cuales se pueden definir todas las demás palabras en un diccionario explicativo; a un conjunto con tal característica lo llamamos un conjunto definidor minimal.

La construcción de tal conjunto se necesita para la conversión del diccionario tradicional en un diccionario semántico computacional (“para la máquina”), siendo un rasgo de éste último, que no se permiten los ciclos en las definiciones.

Se logró obtener una red semántica a partir de un diccionario tradicional, la cual fue depurada, y se desarrolló un algoritmo que encuentra conjuntos definidores, por medio de la aplicación de diversos métodos. Nuestro algoritmo sirve para detectar problemas y defectos en las definiciones del diccionario, relacionados con la presencia de los ciclos, por lo que puede ser de útil para que el lexicógrafo evalúe las definiciones.

También creemos que el método puede ser útil para la discusión teórica del tema de las primitivas semánticas, pero la profundización en las posibles interpretaciones e implicaciones lingüísticas del conjunto **P** obtenido son el tema de investigación futura.

## 5.1 Resultados obtenidos

Basándonos en el algoritmo arriba descrito, desarrollamos una herramienta que permite al lexicógrafo investigar la estructura del diccionario, proporcionándole la siguiente información:

- Muestra varias características de la palabra, tales como su frecuencia (la cantidad de las flechas entrantes), el tamaño de la definición (la cantidad de las flechas salientes), el largo mínimo de uno de los ciclos en que está involucrada y su descripción.
- Genera diferentes conjuntos definidores minimales permitiéndole al usuario seleccionar las variantes del algoritmo (las cuales están en función del ordenamiento de las palabras que se analizan).
- Permite al lexicógrafo cambiar manualmente el conjunto generado y verifica que el conjunto cambiado todavía es un conjunto definidor y si éste es minimal.
- Permite al lexicógrafo cambiar las definiciones de algunas palabras e investigar el impacto en los conjuntos definidores que se generan.
- Dado un conjunto de palabras que el lexicógrafo quiere que sean *no* definidoras, verifica si son compatibles. Si lo son, genera un conjunto definidor (o varios) que no contenga estas palabras. Si no lo son, muestra qué ciclos hay en este conjunto.
- Dado un conjunto de palabras que el lexicógrafo quiere que sean definidoras, genera un conjunto definidor (o varios) que contenga estas palabras. Si este conjunto no puede ser minimal, sugiere eliminar ciertas palabras de éste.
- Dado un conjunto definidor minimal, la herramienta puede:

- Para una palabra no definidora, mostrar su definición expandida a las palabras definidoras, es decir, que consiste sólo de las palabras definidoras.
- Para una palabra definidora, mostrar los ciclos (más cortos o todos) que esta palabra causaría si formase parte del diccionario, dada su definición actual.
- La interpretación de los resultados se discute en la sección 4, donde se menciona que en un buen diccionario, el conjunto definidor no debe tener muchas palabras con frecuencia baja, lo que ayuda al lexicógrafo a detectar y corregir ciertos defectos cambiando las definiciones.

## 5.2 Trabajo futuro

Las tareas principales a investigar se agrupan en dos clases de problemas:

- El problema lingüístico: profundizar en la interpretación lingüística de los resultados y
- El problema técnico: el diseño del algoritmo que encuentre un conjunto P óptimo en un sentido (técnico y/o lingüístico) dado.
- Con más detalle, las tareas futuras específicas son las siguientes:
  1. Dar una interpretación lingüística clara al conjunto P obtenido.
  2. Elaborar criterios lingüísticos que permitan preferencias en el proceso de inclusión de las palabras en el conjunto definidor (“querer” que una palabra sea o no sea primitiva).
  3. Ponderar dichas preferencias (pues no cualquier conjunto de palabras se puede incluir, o no, en un conjunto definidor minimal) y realizar un algoritmo para la selección del conjunto P en cierto sentido óptimo.

4. Desarrollar un algoritmo, exacto o aproximado, para la construcción de un conjunto  $P$  del tamaño más pequeño (mínimo), y analizar su complejidad.

## 6 Bibliografía

---

1. Allport, D. A. (1985), Distributed Memory, Modular Subsystems and Dysphasia, in *Current Perspectives in Disphasia*, Newman, S., Epstein, R. (Eds.). Edinburgh: Churchill Livingstone.
2. Apresjan, J. D. (1974), Regular polysemy, *Linguistics*, 142: 5-32.
3. Apresjan, J. D. (1995), Selected works. Moscow, V 1, 472 p., V 2, 768 p.
4. Bar-Hillel, Y. (1964), ' *A demonstration of the nonfeasibility of fully automatic high quality machine translation* ', in *Language and Information: Selected Essays an Their Theory and Application*. Reading: Addison-Wesley.
5. Budanitsky, A. (1999), *Lexical semantics relatedness and its application in natural language processing*. Technical report CSRG-390, University of Toronto, Toronto, 146 p.
6. Butterworth, B. (1980), Evidences from Pauses in Speech, *Language Production: Speech and Talk, Vol. I*. London:Academic Press, pp. 155-176.
7. Collins, A. M., Quillian, M.R. (1972), How to Make a Language User, in *Organization of Memory*, Tulving, Donaldson (Eds.), New York: Academic Press, pp. 309-351.

8. Collins, A. M, Loftus, E. F. (1975), A spreading activation theory of semantic processing. *Psychological Review*, 82(6):407-428.
9. Conrad, C. (1972), "Cognitive Economy in Semantic Memory", *Journal of Experimental Psychology* , Volume 92, pg 149-154.
10. Diccionario Anaya de la lengua, Grupo Anaya, Internet, Marzo 1997.
11. Dolan, W., Vanderwende, L., Richardson, S. (1999), Polysemy in a Broad-Coverage Natural Language Processing System. In *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press.
12. Evens, M. N. (ed.) (1988), *Relational models of lexicon: Representing knowledge in semantic network*. Cambridge: Cambridge University Press.
13. Fellbaum, C. (1990), The English verb lexicon as a semantic net. *International Journal of Lexicography* 3: 278-301.
14. Fellbaum, C. (1998), *WordNet, An Electronic Lexical Database*. Cambridge, Mass.: MIT Press, 423 p.
15. Fillmore, J. (1982), Towards a descriptive framework for spatial deixis, in R.J. Jarvella and W. Klein (Eds.), *Speech, Place and Action*. New York: Wiley.
16. Forster, K.I. (1976), Accesing the Mental Lexicon, in *New Approaches to Language Mechanisms*, R.J. Wales and Walker (Eds.), Amsterdam: North Holland, pp. 257-287.
17. Gale, W., Curch, K.W. Yarowsky, D. (1992), A method for disambiguating word senses in a large corpus, *Computers and Humanities*, 26:416-39.
18. Garrett, M.F. (1982), Production of Speech: Observations from Normal and Pathological Language Use, in *Normality and Pathology in Cognitive Functions*, Ellis, A. W. (Ed.), London: Academy Press.

19. Gelbukh A., Sidorov G. (2002), Selección automática del vocabulario definidor en un diccionario explicativo. Accepted to: *J. Procesamiento de Lenguaje Natural*, No 29, September 2002. Sociedad Española para el Procesamiento de Lenguaje Natural (SEPLN).
20. Goddard, Cliff. (1999), Polisemy: a problem of definition. In Yael Ravin and Claudia Leacock (Eds.) *Polysemy: Theoretical and computational approaches*. Oxford: Oxford University Press, pp.129-151.
21. Guthrie, J. A., Guthrie, L., Wilks, Y., Aidinejad, H. (1991), Subject-dependent co-occurrence and word sense disambiguation. In *Proceedings of the 29<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Morristown, NJ: ACL, 146-52.
22. Guthrie L., Pustejovsky J., Wilks Y., Sator B.M. (1996), The role of lexicons in natural language processing, in *Communications of the ACM*, 39(1): pp. 63-72, January 1996.
23. Hirst G, St-Onge, D. (1998), Lexical chains as representations of context for the detection and correction of malapropisms. In Christiane Fellbaum, editor, *WordNet: An Electronic Lexical Database*, chapter 13, pages 305-332. The MIT Press. Cambridge, MA.
24. Katz, J. J. (1972), *Semantic Theory*. New York. Harper & Row.
25. Katz, J. J., Fodor, J.A. (1963), "The structure of a semantic theory", *Language*, 39; pp. 179-210.
26. Katz, J. J., Leacock, C. and Ravin, Y. (1985), A decompositional approach to modification, in E. LaPore and B. MacLaughlin (Eds.), *Action and Events*. Oxford: Blackwell.
27. Kozima, H. and Furugori, T. (1993), Similarity between words computed by spreading activation on an English dictionary. In: *Proceedings of the 6 conference of the European chapter of ACL*, pp. 232-239.

28. Kozima, H. and Ito, A. (1997), Context-sensitive word distance by adaptive scaling of a semantic space. In: Mitkov, R. and Nicolov, N. (Eds.) *Recent Advances in Natural Language Proceedings: Selected Papers from RANLP'95*, pp. 111-124.
29. Leacock, G., Chodorow, M.S., and Miller, G.A. (1998), Using corpus estatistic and WordNet relations for sense identification. In *Computational Linguistics*, 24.1:47-65.
30. Lesk, M. (1986), ' Automatic sense disambiguation: how to tell a pine cone form an ice cream cone' , in *Proceedings of the 1986 SIGDOC Conference*. New York: Association for Computing Machinery, 24-6.
31. Marslen-Winslow, W. (1984), Function and Process in *Spoken Word Recognition: A Tutotial Review*, in Bouma and Bouwhuis (Eds.), pp. 125-150.
32. Mel' cuk, I. and Zholkovsky, A. (1988), The explanatory combinatorial dictionary. In M. W. Enens (ed.), *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*. Cambridge: Cambridge University Press.
33. Miller, G.A., Johnson-Laird, P.N. (1976), *Language and Perception*, Cambridge, Mass.: The Belknap Press of Harvard University Press.
34. Morris J. and Hirst G. (1991), Lexical cohesion computed by thesaural relations as an indicator of the structure of text. In *Computational Linguistics*, 17:21-48.
35. Morton, J. (1982), *Disintegrating the Lexicon: An Information Processing Approach*, in Mehler, Walker and Garrett (Eds.), pp.89-109.
36. Osgood, C.E. (1952), The nature and measurement of meaning. *Psychological Bulletin*, 49:197-237.
37. Pustejovsky, J. (1991), The Generative Lexicon, in *Computational Linguistics*, vol, 17 no. 4, pp. 409-441.

38. Quilliant, M. Ross. (1968), Semantic memory. In M. Minsky, editor, *Semantic Information Processing*. MIT Press, Cambridge, MA.
39. Ravin, Y. and Leacock, C. (2000), Polysemy: An Overview. In Yael Ravin and Claudia Leacock (Eds.) *Polysemy: Theoretical and computational approaches*. Oxford: Oxford University Press, 227 p.
40. Resnik, Phillip (1995), Using information content to evaluate semantic similarity. In *Proceedings of the 14<sup>th</sup> International Conference on Artificial Intelligence*, pages 448-453, Montreal, Canada, August 1995.
41. Rumelhart, D. E., Smolensky, J.L., McClelland, J.L., Hinton, G.E. (1986), Schemata and Sequential Thought Processes in PDP models, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume I; Foundations*, Rumelhart, D. E., McClelland, J.L. (Eds.), Cambridge, Mass.: MIT Press.
42. Saint-Dizier, P. and Viegas, E. (Eds.) (1995), *Computational lexical semantics*. Cambridge: Cambridge University Press, 447 p.
43. Sidorov, G., Gelbukh, A. (2002), Desambiguación de los sentidos de las palabras en un diccionario explicativo. Proc. CICLing-2002, 3rd International Conference on Intelligent Text Processing and Computational Linguistics, February 17–23, 2002, Mexico City. Lecture Notes in Computer Science, N 2276, Springer-Verlag, pp. 300–303.
44. Towell, G., Voorhees, E.M. (1998), Disambiguating highly ambiguous words, *Computational Linguistics*, 24.1:125-46.
45. West M. P. (1953), *A general service list of english words, with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Longman, Harlow, Sussex.
46. Wierzbicka, A. (1980), *Lingua Mentalis: The semantics of natural language*. New York: Academic Press.

47. Wierzbicka, A. (1990), ' Prototypes save: on the uses and abuses of the notion of "prototypes" in linguistics and related fields' , in S. L. Tsohatzidis (ed.), *Meanings and Prototypes: Studies in Linguistic Categorization*. London: Routledge & Kegan Paul.
48. Wierzbicka, A. (1996), *Semantics: Primes and Universals*. Oxford: Oxford University Press.
49. Wittgestein, L. (1953), *Philosophical Investigations*. Oxford: Basil Blackwell & Mott.

## Anexo 1. Formato del archivo de entrada

---

En este anexo se muestra una parte del contenido del archivo net1.vis, mismo que lista cada palabra del grafo seguido por las palabras empleadas en su definición, un salto de línea, y las palabras que la usan en su definición.

```
0:ábaco
380:alambre, 860:arquitectura, 1246:base, 1349:bola, 1673:capitel,
2468:contar, 2605:correr, 3495:dórico, 3932:equino, 4630:forma,
4953:griego, 6347:marco, 7126:orden, 7291:paralelepípedo,
7293:paralelo, 9121:servir, 9392:sujeto,

3495:dórico, 3932:equino, 9468:tablero,

-----
1:abajo
1170:bajo, 3353:dirección, 5521:inferior, 6181:lugar,

1169:bajar, 1170:bajo, 1525:caer, 2861:declinar, 4412:falda,
4656:foso, 4797:gancho, 5771:invertir, 6032:levantar, 6390:mate,
7256:palo, 7332:partido, 8210:quebrado,

-----
2:abandonado
3088:descuidado, 5114:higiénico, 7741:poco, 9367:sucio,

3088:descuidado, 6203:madeja,

-----
3:abandonar
171:actividad, 274:afecto, 279:afición, 436:algo, 742:apoyar,
868:arrastrar, 1525:caer, 2253:completo, 2637:cosa, 2905:dejar,
3020:desamparar, 3089:descuidar, 3152:desistir, 3691:empezar,
3878:entregar, 5702:interés, 5789:ir, 6181:lugar, 6448:meditar,
6988:obligación, 7354:pasión, 7553:persona, 8594:renunciar,
9974:uno, 9991:usar, 10201:vicio, 10302:voluntad,
```

4:abandono, 354:ahorcar, 482:alta, 2151:colgar, 2798:cuña,  
3019:desalojar, 3438:disuadir, 3664:emigrante, 3665:emigrar,  
4254:evacuar, 4384:extraviar, 7696:plantar, 7934:presionar,

-----  
4:abandono  
3:abandonar, 98:acción, 2637:cosa, 2904:dejadez, 3017:desaliño,  
3565:efecto, 7553:persona,

3663:emigración, 8593:renuncia,

-----  
5:abanico  
359:aire, 1742:carretera, 1973:ciclismo, 2011:circular,  
2600:corredor, 2626:cortar, 3413:disposición, 4536:fila,  
4630:forma, 5024:hacer, 5665:instrumento, 5923:lado, 6732:mover,  
8894:sable, 9012:sector, 9604:tener, 10212:viento,

4975:guarda, 7253:palmito, 10060:varilla, 10061:varillaje,

-----  
6:abaratar  
436:algo, 7856:precio, 8335:rebajar,

1169:bajar,

-----  
7:abarca  
1587:calzado, 2431:consistente, 2594:correa, 2767:cuerda,  
8345:reborde, 8884:rústico, 9375:suela, 9391:sujetar,

2174:comarca, 4763:gacela, 5675:integral, 6135:llave,  
6445:mediodía, 6772:mundial, 6824:narrativa, 7891:prehistoria,  
9046:semana,

-----  
8:abarcar  
89:acaparar, 404:alcanzar, 1015:atender, 1409:brazo, 1824:caza,  
2268:comprender, 2474:contener, 2637:cosa, 8822:rodear,  
9311:sorprender, 9604:tener, 9645:terreno, 9670:tiempo,  
10265:vista,

2117:coger, 3800:englobar,

-----  
9:abastecer  
2637:cosa, 6846:necesario, 8116:proveer,

8115:proveedor, 8618:repostar, 9401:suministrar,

-----  
10:abatido  
623:ánimo, 4726:fuerza, 6900:no, 9604:tener,

732:apocado, 1170:bajo,

-----  
11:abatimiento  
12:abatir, 98:acción, 3144:desilusión, 3565:efecto, 4569:físico,  
6691:moral, 7828:postración, 9902:tristeza,

307:agobiar, 2845:decaimiento,

-----

12:abatir

436:algo, 1116:ave, 1169:bajar, 2637:cosa, 2995:derribar,  
3015:desalentar, 3022:desanimar, 3029:desarmar, 3058:descender,  
3164:desmontar, 3539:echar, 5024:hacer, 5239:humillar, 5789:ir,  
7779:poner, 9601:tendido, 9675:tierra, 10176:vertical,  
10291:volar,

11:abatimiento, 723:aplastar, 1055:atropellar, 5239:humillar,  
5246:hundir,

-----

13:abdomen

684:aparato, 913:artrópodo, 2474:contener, 2771:cuerpo,  
3324:digestivo, 6728:motor, 7321:parte, 7371:pata, 7824:posterior,  
8633:reproductor, 9986:urinario, 10213:vientre, 10252:víscera,

14:abdominal, 342:aguijón, 787:arácnido, 1888:cerco,  
2718:crustáceo, 4010:escorpión, 4775:galera, 5112:hígado,  
5626:insecto, 7063:oído, 9741:tórax, 10252:víscera,

-----

14:abdominal

13:abdomen, 8544:relativo,

790:araña, 5737:intestino, 6367:marsupial,

-----

15:abecedario

1751:cartel, 2393:conjunto, 3695:emplear, 3849:enseñar,  
5267:idioma, 5991:leer, 6012:lenguaje, 6024:letra, 6319>manual,  
7128:ordenada, 9109:serie, 9146:signo, 9309:sordo,  
9569:telegráfico, 9991:usar,

426:alfabeto,

-----

16:abeja

329:agrupar, 1693:caracterizar, 1880:cera, 1983:cierto,  
2284:común, 2771:cuerpo, 3443:diverso, 4079:especie, 5459:indicar,  
5498:industrial, 5626:insecto, 6561:miel, 6851:néctar,  
6914:nombre, 7128:ordenada, 7553:persona, 7758:polen,  
8014:producción, 8377:recibir, 9031:segregar, 9228:social,  
9604:tener, 9779:trabajador, 10003:utilizar,

1133:avispa, 1172:bala, 1880:cera, 2155:colmena, 6561:miel,  
6851:néctar, 7263:panal, 7769:pollo, 8523:reina,

-----

17:abertura

41:abrir, 346:agujero, 359:aire, 585:ancho, 586:anchura,  
900:articulación, 2293:conceder, 3283:diámetro, 3565:efecto,  
4681:franqueza, 4954:grieta, 5068:hendidura, 6014:lente,  
6677:montaña, 7143:órgano, 7359:paso, 9069:sencillez,  
9196:sistema, 9858:trato, 9999:útil, 10038:valle,

346:agujero, 1093:automático, 1308:bigote, 1337:boca,  
1365:boquilla, 1483:buzón, 1960:chorro, 3104:desembocadura,  
3105:desembocar, 3331:dilatación, 4344:explosión, 4954:grieta,

6443:medio, 6743:mucosa, 7017:obturador, 7018:obturar, 7067:ojo,  
7151:orificio, 7803:portilla, 7804:portillo, 8196:pupila,  
8229:quiebra, 8281:ranura, 8509:registro, 8695:resquicio,  
9152:silbo, 9433:surco, 9723:toma, 10044:válvula, 10129:ventana,  
10130:ventilación, 10131:ventilador, 10133:ventosa, 10294:volcán,  
10323:vulva,

-----  
18:abeto

128:acicular, 386:alba, 496:alto, 745:apreciar, 799:árbol,  
1182:baleares, 1325:blanco, 1326:blancura, 2284:común, 2553:copa,  
2984:derecho, 3443:diverso, 4079:especie, 4363:extender,  
4834:género, 5152:hoja, 5185:horizontal, 6204:madera, 6914:nombre,  
7551:persistente, 8273:rama, 8377:recibir, 9501:tamaño,  
9912:tronco,

7640:piña,

-----  
19:abierto

19:abierto, 41:abrir, 183:acuerdo, 359:aire, 730:aplicar,  
900:articulación, 1197:bancario, 1902:cerrado, 1971:cicatrizar,  
2044:claro, 2241:competición, 2753:cubierta, 2765:cuenta,  
2973:deportivo, 3332:dilatado, 3635:embarcación, 4120:espontáneo,  
4268:evolucionar, 4311:exigencia, 4322:expansivo, 4623:fonético,  
4679:franco, 5492:indudable, 6131:llano, 6784:muro, 6900:no,  
7143:órgano, 7323:participante, 7359:paso, 7374:patente,  
7553:persona, 7744:poder, 8064:pronunciar, 8296:raso,  
8623:representar, 8712:resultar, 8822:rodear, 8922:saldar,  
9179:sincero, 9294:sonido, 9604:tener, 9670:tiempo, 10036:valla,  
10287:vocal,

19:abierto, 565:amplio, 825:argolla, 1150:azote, 1339:bocado,  
1346:bofetada, 1364:boquiabierto, 1616:canal, 1632:canguro,  
4809:gasa, 4946:grave, 5926:ladrón, 6042:liberal, 6051:libre,  
6160:logia, 6390:mate, 6443:medio, 7281:parábola, 7714:plaza,  
7825:postigo, 8307:ratón, 8851:roseta, 8960:sandalia,  
9303:soportal, 9517:tapar, 9928:tubo,

-----  
20:ablandar

1327:blando, 1831:ceder, 2398:conmover, 2637:cosa, 3776:enfado,  
4701:frío, 4752:furia, 5689:intensidad, 6144:lluvia, 6608:mitigar,  
7779:poner, 8784:rigor, 9331:suavizar, 10212:viento,

520:amansar, 8070:propiciar,

-----  
21:abnegación

169:actitud, 436:algo, 1288:beneficiar, 2737:cualidad, 2941:demás,  
5024:hacer, 7985:privación, 8593:renuncia, 8906:sacrificio,  
9304:soportar, 9388:sufrir, 9974:uno, 10303:voluntario,

8906:sacrificio,

-----  
22:abogado

2067:cliente, 2881:defender, 3298:dictamen, 3670:emitir,  
5702:interés, 5890:jurídico, 5894:justicia, 6429:media,  
7553:persona, 7698:plantear, 7997:problema, 8100:protector,  
8974:santo, 9890:tribunal,

## Anexo I. Formato del archivo de entrada

---

---

416:alegar, 1317:birrete, 2806:curia, 2882:defensa, 5547:informar,  
5550:informe, 6025:letrado,

-----  
23:abolengo  
649:antepasado, 922:ascendencia, 4828:general, 5077:herencia,  
5291:ilustre, 7390:patrimonio, 7553:persona, 8000:procedente,  
  
9247:solar,

-----  
24:abolición  
25:abolir, 98:acción, 673:anular, 2653:costumbre, 3565:efecto,  
6037:ley,  
  
2172:coma,

-----  
25:abolir  
2653:costumbre, 5761:invalidar, 6037:ley,  
  
24:abolición, 2991:derogar, 4151:estamento,

-----  
26:abonar  
163:acreditar, 436:algo, 969:asistir, 971:asociación, 1107:avaluar,  
1455:buen, 1564:calificar, 1656:cantidad, 2637:cosa, 3539: echar,  
4499:fertilizante, 5623:inscribir, 6181:lugar, 6392:materia,  
7226:pagar, 7229:pago, 7553:persona, 7744:poder, 8377:recibir,  
9119:servicio, 9675:tierra, 10039:valor,  
  
27:abono, 163:acreditar, 4500:fertilizar, 7342:pasaje,

-----  
27:abono  
26:abonar, 98:acción, 572:añadir, 969:asistir, 1656:cantidad,  
2439:constar, 2984:derecho, 3470:documento, 3565:efecto, 4036:ese,  
5623:inscribir, 6181:lugar, 6453:mejor, 8377:recibir,  
9119:servicio, 9447:sustancia, 9675:tierra,  
  
4178:estiércol, 4499:fertilizante,

-----  
28:abordar  
436:algo, 1001:asunto, 1037:atracar, 2774:cuestión, 1954:chocar,  
3314:difícil, 3635:embarcación, 3636:embarcadero, 3691:empezar,  
3697:emprender, 6839:nave, 9857:tratar,  
  
3644:embestida, 8911:sadismo,

-----  
29:aborigen  
4365:extensión, 4630:forma, 5015:habitante, 7156:originario,  
7230:país, 7553:persona, 7970:primitivo, 9377:suelo,  
  
7970:primitivo,

-----  
30:abortar

## Anexo I. Formato del archivo de entrada

---

---

436:algo, 3634:embarazo, 4663:fracasar, 5024:hacer,  
5728:interrupción, 8124:provocar, 9388:sufrir,

4508:feto,

-----  
31:abrasar

305:agitar, 436:algo, 816:ardor, 1401:brasa, 1561:caliente,  
1580:calor, 2943:demasiado, 3477:dolor, 4011:escozor, 4701:frío,  
7354:pasión, 7620:picor, 7694:planta, 8015:producir, 8221:quemar,  
8459:reducir, 9002:secar, 9018:sed, 9094:sentir, 9447:sustancia,  
10284:vivo,

813:arder, 921:asar, 1700:carbonizar, 5385:incendio, 8221:quemar,

-----  
32:abrasión

98:acción, 864:arrancar, 3129:desgastar, 3565:efecto,  
3773:enérgico, 4698:fricción, 5809:irritación, 6469:membrana,  
8015:producir, 8199:purgante, 9408:superficial,

33:abrasivo, 3531:dureza,

-----  
33:abrasivo

32:abrasión, 282:afilar, 450:alisar, 1327:blando, 8015:producir,  
8167:pulir, 8297:raspar, 8544:relativo, 9121:servir,  
9447:sustancia,

3282:diamante, 8432:rectificar,

-----  
34:abrazadera

35:abrazar, 937:asegurar, 2637:cosa, 7628:pieza, 9121:servir,

1365:boquilla, 4219:estribo,

-----  
35:abrazar

115:aceptar, 138:acoger, 274:afecto, 1409:brazo, 1859:ceñir,  
3466:doctrina, 4209:estrechar, 8822:rodear, 9066:señal,

34:abrazadera,

-----  
36:abreviado

2634:corto, 3988:escaso, 8714:resumir,

38:abreviatura,

-----  
37:abreviar

107:acelerar, 157:acortar, 4059:espacio, 9670:tiempo,

200:adelantar, 8714:resumir,

-----  
38:abreviatura

36:abreviado, 4019:escritura, 7235:palabra, 8621:representación,

## Anexo I. Formato del archivo de entrada

---

---

1213:bar, 1869:centímetro, 1985:cifra, 4240:etcétera, 6554:micra,  
6571:milicia, 9072:seno, 9166:símbolo,

-----  
39:abrigar

274:afecto, 559:amparar, 5254:idea, 6233:mal, 7102:opinión,  
8661:resguardar, 9604:tener, 9670:tiempo,

40:abrigo, 4896:gorro, 6293:manguito, 9517:tapar, 10183:vestido,

-----  
40:abrigo

39:abrigar, 560:amparo, 1822:cavidad, 2637:cosa, 2645:costa,  
4367:exterior, 4701:frío, 4828:general, 5956:largo, 6181:lugar,  
6291:manga, 6393:material, 6830:natural, 6839:nave, 7903:prenda,  
8099:protección, 8101:proteger, 8670:resistente, 8810:roca,  
9563:tela, 10003:utilizar, 10212:viento,

1163:bahía, 1668:capilla, 1676:capote, 1682:capucha, 2763:cuello,  
4695:fresco, 4762:gabardina, 4804:garita, 5575:inhóspito,  
9877:trenca, 10152:verdugo,

-----  
41:abrir

297:afuera, 436:algo, 699:apartar, 1500:cabeza, 1531:cajón,  
1684:capullo, 1902:cerrado, 1906:cerrojo, 2210:comienzo,  
2320:concreto, 2449:construir, 2605:correr, 2626:cortar,  
2637:cosa, 2815:curva, 2830:dar, 3030:desarrollar, 3078:descorrer,  
3086:descubrir, 3184:despegar, 3503:dos, 4059:espacio,  
4363:extender, 4367:exterior, 4865:girar, 4900:gozne, 4994:guía,  
5024:hacer, 5377:inaugurar, 5708:interior, 5789:ir, 5923:lado,  
6053:libro, 6181:lugar, 6347:marco, 6410:mayor, 7040:oculto,  
7139:organismo, 7359:paso, 7534:permitir, 7553:persona,  
7594:pestillo, 7595:pétalo, 7725:pliego, 8149:público,  
8154:puerta, 8214:quedar, 8270:rajarse, 8293:rasgar, 8838:romper,  
9053:semejante, 9096:separar, 9124:sesión, 9179:sincero,  
9670:tiempo, 9724:tomar, 10089:vehículo, 10137:ver,

17:abertura, 19:abierto, 713:aperitivo, 714:apertura, 1178:balcón,  
1283:bellota, 1679:cápsula, 1904:cerrar, 2598:corredera, 2830:dar,  
3439:diurno, 3979:escape, 4136:establecer, 5152:hoja,  
5377:inaugurar, 5397:incisivo, 6004:legumbre, 6127:llamar,  
6135:llave, 6298:manilla, 7132:ordenanza, 7614:picaporte,  
7825:postigo, 7842:pozo, 8270:rajarse, 8963:sangría, 9698:tirador,  
9895:trinchera, 10044:válvula, 10131:ventilador,

## Anexo 2. Lista de las primitivas encontradas

---

Resultados obtenidos corriendo el programa **visualizer** (modo de lista). El archivo de entrada a este módulo tiene solamente los identificadores de las primitivas. El archivo de salida, que es el que mostramos aquí, tiene el número de flechas salientes y entrantes a cada primitiva -FS y FE respectivamente-, seguido de la palabra, y su identificador entre <>. Aquí mostramos sólo un fragmento el archivo de salida obtenido con el método cuatro (votación aleatoria).

FS	FE	Palabra<Id>	FS	FE	Palabra<Id>
9	2	ábaco<0>	6	59	acción<117>
1	1	abandonado<3>	9	147	acción<119>
21	4	abeja<19>	11	11	aceite<126>
3	13	abertura<20>	2	3	aceituna<130>
6	2	abrir<48>	14	8	acento<135>
10	1	abscisa<57>	15	3	acepción<141>
13	3	absoluto<58>	4	11	ácido<160>
6	1	aburrido<79>	1	1	acompañante<180>
3	4	aburrimiento<80>	3	20	acompañar<182>
1	1	acabar<88>	2	11	acontecimiento<187>
4	1	acabar<90>	1	1	acorde<193>
22	3	acacia<92>	7	15	actitud<207>
5	1	acanto<104>	3	23	actividad<210>
4	4	acceder<107>	6	33	actividad<211>
3	1	accesorio<109>	2	12	activo<213>
2	1	accidental<111>	4	4	activo<216>
5	1	accidente<112>	5	8	acto<218>
7	68	acción<115>	5	2	acuñaar<246>

Anexo 2. Lista de las primitivas encontradas

FS	FE	Palabra<Id>	FS	FE	Palabra<Id>
2	1	acusar<249>	16	1	ajo<469>
3	1	acústica<251>	4	8	ajustar<473>
5	24	adecuado<256>	14	13	ala<478>
5	1	adelante<262>	1	5	alabar<483>
2	1	aderezar<267>	4	1	alambique<484>
5	2	adivinanza<283>	4	1	alambre<486>
9	14	adjetivo<286>	8	2	albedrío<493>
9	9	adjetivo<287>	7	1	alcalde<500>
9	11	administrativo<297>	5	2	alcance<505>
10	3	admiración<300>	18	5	alcohol<515>
4	15	admitir<304>	5	7	alcohólico<519>
5	1	adoquín<311>	11	1	alcornoque<520>
12	1	adormidera<317>	4	2	alegría<528>
19	12	adverbio<330>	3	2	alfiler<543>
2	3	advertir<337>	13	1	algarrobo<546>
1	1	aeroplano<343>	12	4	algodón<550>
3	6	afectivo<353>	2	1	algodón<551>
5	30	afecto<355>	1	1	aligerar<556>
2	3	afectuoso<357>	5	7	alimentar<561>
4	2	afeitar<358>	17	3	almeja<581>
4	7	afición<360>	7	1	almendra<582>
5	7	afilar<364>	3	1	almendra<583>
5	7	afirmar<370>	3	1	almíbar<585>
2	6	aflicción<372>	9	2	almidón<587>
4	8	agarrar<378>	8	1	alminar<588>
2	1	agasajar<381>	1	1	almohada<589>
3	2	ágil<387>	3	5	alojamiento<593>
5	6	agitar<388>	3	22	alrededor<600>
5	5	agradar<392>	3	1	altanero<604>
3	3	agricultura<404>	4	3	altar<606>
9	1	agrimonia<406>	3	1	altisonante<618>
3	5	agrupación<410>	4	1	alto<626>
29	92	agua<414>	5	2	alto9<632>
3	1	agua<416>	4	7	altura<638>
4	2	agua<418>	15	2	aluminio<644>
5	4	aguantar<420>	9	9	alumno<646>
4	17	aguar<423>	4	2	alveolo<647>
8	5	agudo<434>	3	1	ama<649>
6	13	agujero<447>	3	4	amable<650>
4	2	ahora<449>	4	1	amaestrar<652>
17	35	aire<452>	4	2	amanecer<654>
3	10	aislado<458>	2	1	amansar<655>
11	2	ajedrez<461>	6	2	amenaza<677>

Anexo 2. Lista de las primitivas encontradas

FS	FE	Palabra<Id>	FS	FE	Palabra<Id>
3	2	amistad<686>	9	24	menor<7998>
9	1	macero<7663>	4	24	menos<8003>
1	1	mácula<7669>	6	1	mentol<8014>
4	1	madre<7680>	6	7	mérito<8027>
3	1	maleza<7721>	4	1	mesa<8034>
4	1	malla<7726>	7	1	meteorito<8052>
8	3	mama<7736>	5	14	método<8064>
1	4	manda<7752>	4	1	mezclar<8076>
9	1	mandarino<7758>	3	1	miau<8081>
5	1	manecilla<7766>	11	2	microscopio<8087>
6	3	manejable<7767>	6	21	miembro<8091>
3	24	manifestar<7784>	2	3	militar<8105>
7	11	mano<7790>	2	1	mina<8112>
5	7	mano<7791>	6	25	mineral<8113>
3	1	mañoso<7799>	6	4	ministerio<8116>
7	1	manubrio<7815>	6	3	ministro<8118>
10	4	manuscrito<7816>	2	9	minucioso<8121>
10	1	manzano<7819>	2	3	miramiento<8128>
10	34	máquina<7822>	5	2	mirar<8129>
8	29	mar<7824>	8	6	misión<8137>
4	5	marco<7844>	3	4	moda<8149>
13	4	marea<7846>	5	13	modelo<8153>
12	1	margarita<7849>	4	1	modelo<8156>
9	2	marqués<7862>	3	4	moderar<8160>
14	32	masa<7867>	14	22	modo<8167>
5	35	materia<7883>	3	3	mojar<8175>
6	11	materia<7884>	4	1	mojón<8176>
5	51	materia<7885>	14	19	molécula<8181>
4	7	materia<7886>	3	7	molestar<8186>
7	5	máxima<7904>	11	11	molusco<8192>
13	48	mayor<7907>	8	25	moneda<8201>
2	1	mayor<7909>	3	1	montacargas<8212>
4	9	mayoría<7910>	1	2	montar<8216>
2	1	mazorca<7914>	3	1	mordisquear<8240>
4	2	media<7931>	2	1	morral<8250>
2	3	mediar<7936>	7	1	morse<8251>
6	19	medicamento<7938>	3	1	mortificar<8255>
9	14	medio6<7957>	6	1	mostaza<8261>
3	57	medio9<7958>	2	13	motivo<8271>
7	10	medir<7962>	2	29	mover<8280>
2	1	melena<7978>	2	5	movimiento<8296>
1	1	melocotón<7982>	3	1	mucosidad<8302>
5	6	memoria<7990>	7	15	mueble<8306>

Anexo 2. Lista de las primitivas encontradas

FS	FE	Palabra<Id>	FS	FE	Palabra<Id>
4	1	muela<8309>	6	1	tragaluz<12327>
6	2	muelle<8310>	3	2	traje<12332>
9	1	muelle<8311>	3	4	trámite<12334>
2	2	muestra<8320>	5	1	trampa<12337>
2	1	mugre<8321>	7	4	tranquilo<12341>
3	12	mujer<8322>	3	7	transcurrir<12345>
9	2	multa<8327>	4	3	transmitir<12361>
3	2	multiplicación<8329>	8	2	transparencia<12363>
3	1	mundo<8343>	17	3	tráquea<12377>
2	1	municipio<8347>	1	1	trasero<12379>
6	2	muralla<8349>	3	7	trasladar<12381>
10	10	música<8360>	6	3	trastornar<12385>
5	3	músico<8367>	2	3	tren<12422>
4	9	nacer<8375>	9	1	trenca<12425>
13	3	naranja<8395>	7	2	trenza<12426>
2	2	naranja<8396>	6	11	tribunal<12439>
3	5	narrar<8402>	19	7	trigo<12441>
4	2	natural<8416>	5	1	trío<12444>
4	2	natural<8417>	3	7	tristeza<12450>
3	1	necedad<8438>	4	1	triturar<12451>
4	5	necesario<8440>	5	4	tropa<12463>
3	14	tipo<12198>	10	4	tubérculo<12470>
12	1	tiranía<12206>	11	2	tulipa<12481>
1	1	tiro<12214>	3	2	turbación<12489>
5	8	título<12220>	3	1	ubicuidad<12499>
4	19	título<12221>	8	44	unidad<12512>
4	8	tocar<12230>	5	10	unidad<12513>
2	26	tomar<12251>	3	6	universo<12534>
3	4	tomar7<12258>	9	159	uno<12535>
6	2	tónica<12262>	4	29	uno4<12540>
16	11	tono<12263>	4	56	usar<12552>
4	1	tonsura<12266>	2	36	uso<12556>
8	1	tordo<12276>	3	17	utensilio<12560>
5	3	tornillo<12280>	2	9	útil<12563>
21	20	toro<12284>	10	5	uva<12569>
3	3	trabajador<12305>	1	1	vacuno<12580>
2	3	trabajar<12309>	5	3	vallado<12601>
7	10	trabajo<12313>	6	25	valor<12602>
13	6	tradición<12317>	5	8	valor<12605>
2	2	tradicional<12319>	4	20	valor<12608>
1	1	traer<12325>	4	1	vanagloriarse<12614>
3	5	tráfico<12326>	6	3	variante<12630>

*Anexo 2. Lista de las primitivas encontradas*

FS	FE	Palabra<Id>	FS	FE	Palabra<Id>
2	6	variedad<12633>	2	3	Vestir<12790>
5	1	varillaje<12637>	23	1	Víbora<12806>
9	7	vaso<12645>	3	2	vibrar<12809>
3	3	vaso<12646>	5	2	víctima<12815>
2	1	vecindad<12651>	5	14	vida<12817>
2	2	vectorial<12657>	9	4	vigente<12839>
6	4	vela<12677>	5	16	vino<12852>
10	2	velar<12682>	5	5	virtud<12879>
3	1	velero<12684>	5	1	víscera<12884>
4	1	vello<12685>	1	1	vistazo<12902>
4	2	veloz<12692>	3	8	viveza<12913>
3	1	vencimiento<12699>	1	1	vivir<12919>
8	1	venda<12700>	3	14	volumen<12944>
4	4	vender<12703>	5	23	voluntad<12951>
4	11	venenoso<12707>	8	17	volver<12959>
10	1	ventrículo<12726>	7	15	voz<12974>
22	13	verbal<12736>	11	2	vulgar<12991>
21	33	verbo<12737>	7	1	yeyuno<12997>
5	2	verja<12759>	15	1	zarza<13004>
23	25	verso<12766>	7	10	zona<13008>
7	6	vértebra<12768>	6	1	zurdo<13013>
9	1	Vértigo<12777>	2	1	zurra<13014>
8	1	Vestíbulo<12779>			
4	2	Vestimenta<12786>			

## Anexo 3. Lista de los ciclos encontrados

---

Este es un ejemplo de los resultados obtenidos corriendo el programa **visualizer** (modo normal). El archivo de entrada es el que contiene la lista de identificadores de las primitivas y el archivo de salida tiene la longitud del ciclo (LC) asociado a cada primitiva, número de flechas salientes (FS), número de flechas entrantes (FE), y por último uno de los ciclos más cortos. Aquí presentamos sólo un fragmento de un archivo de salida obtenido con el método 4 (método de votación aleatoria).

---

<i>LC</i>	<i>FS</i>	<i>FE</i>	<i>Ciclo</i>
2	9	2	ábaco<0> -> equino<4830> -> ábaco<0>
2	1	1	abandonado<3> -> descuidado<3839> -> abandonado<3>
2	21	4	abeja<19> -> cera<2259> -> abeja<19>
2	3	13	abertura<20> -> grieta<6110> -> abertura<20>
3	6	2	abrir<48> -> gozne<6024> -> hoja<6357> -> abrir<48>
1	10	1	abscisa<57> -> abscisa<57>
1	13	3	absoluto<58> -> absoluto<58>
1	6	1	aburrido<79> -> aburrido<79>
2	3	4	aburrimiento<80> -> aburrido<78> -> aburrimiento<80>
2	1	1	acabar<88> -> rematar<10751> -> acabar<88>
1	4	1	acabar<90> -> acabar<90>
1	22	3	acacia<92> -> acacia<92>
1	5	1	acanto<104> -> acanto<104>
22	4	4	acceder<107> -> querer<10341> -> aceptar<146> -> aprobar<945> -> asignatura<1192> -> académico<96> -> clásico<2470> -> griego<6109> -> helénico<6254> ->

---

LC	FS	FE	Ciclo
			estado<5077> -> anterior<808> -> indicativo<6731> -> acción<120> -> integrar<6959> -> entrar<4768> -> academia<93> -> reconocido<10558> -> reconocer<10556> -> asentir<1171> -> acuerdo<242> -> asentimiento<1170> -> consentimiento<2969> -> acceder<107>
2	3	1	accesorio<109> -> secundario<11297> -> accesorio<109>
2	2	1	accidental<111> -> ocasional<8697> -> accidental<111>
2	5	1	accidente<112> -> suceso<11775> -> accidente<112>
11	7	68	acción<115> -> elemento<4509> -> noción<8524> -> conocimiento<2946> -> intelectual<6966> -> manual<7814> -> resumir<10950> -> sintetizar<11556> -> metódico<8063> -> seguir<11322> -> movimiento<8292> -> acción<115>
6	6	59	acción<117> -> impacto<6563> -> impresión<6601> -> cuerpo<3445> -> objeto<8643> -> ejercicio<4458> -> acción<117>
2	9	147	acción<119> -> obra<8666> -> acción<119>
2	11	11	aceite<126> -> bacalao<1418> -> aceite<126>
2	2	3	aceituna<130> -> olivo<8775> -> aceituna<130>
3	14	8	acento<135> -> intensidad<6973> -> acentuación<138> -> acento<135>
1	15	3	acepción<141> -> acepción<141>
2	4	11	ácido<160> -> agrio<407> -> ácido<160>
2	1	1	acompañante<180> -> séquito<11438> -> acompañante<180>
2	3	20	acompañar<182> -> acompañamiento<178> -> acompañar<182>
2	2	11	acontecimiento<187> -> suceso<11772> -> acontecimiento<187>
2	1	1	acorde<193> -> conforme<2894> -> acorde<193>
2	7	15	actitud<207> -> postura<9800> -> actitud<207>
8	3	23	actividad<210> -> prontitud<10103> -> velocidad<12688> -> poco<9672> -> empleado<4586> -> oficina<8743> -> trabajo<12312> -> trabajar<12307> -> actividad<210>
3	6	33	actividad<211> -> tarea<11981> -> trabajo<12312> -> actividad<211>
2	2	12	activo<213> -> diligente<4106> -> activo<213>
1	4	4	activo<216> -> activo<216>
28	5	8	acto<218> -> suceso<11775> -> delito<3677> -> culpa<3464> -> perjudicar<9398> -> moral<8233> -> espíritu<5024> -> inteligente<6969> -> culto<3476> -> adoración<313> -> apasionado<872> ->

LC	FS	FE	Ciclo
			vehemente<12666> -> viveza<12912> -> prontitud<10103> -> velocidad<12688> -> poco<9672> -> corta<3226> -> tala<11933> -> cortar<3231> -> confección<2871> -> vestir<12787> -> ropa<11088> -> servir<11463> -> bebida<1550> -> beber<1546> -> boca<1617> -> entrada<4754> -> asistir<1207> -> acto<218>
3	5	2	acuñar<246> -> cuño<3491> -> medalla<7928> -> acuñar<246>
2	2	1	acusar<249> -> delatar<3651> -> acusar<249>
2	3	1	acústica<251> -> sonoro<11692> -> acústica<251>
2	5	24	adecuado<256> -> conveniente<3114> -> adecuado<256>
18	5	1	adelante<262> -> posponer<9786> -> menos<8000> -> cuantitativo<3387> -> relativo<10724> -> regular<10685> -> pasable<9144> -> aceptable<142> -> aceptar<145> -> voluntario<12955> -> prestar<9936> -> dejar<3636> -> soltar<11660> -> desenvoltura<3861> -> facilidad<5409> - > excesivo<5248> -> sobrepasar<11590> -> adelantar<258> -> adelante<262>
3	2	1	aderezar<267> -> guisar<6150> -> salsa<11200> -> aderezar<267>
2	5	2	adivinanza<283> -> acertijo<156> -> adivinanza<283>
2	9	14	adjetivo<286> -> calificar<1853> -> adjetivo<286>
1	9	9	adjetivo<287> -> adjetivo<287>
8	9	11	administrativo<297> -> ciudadano<2433> -> derecho2<3739> -> comunidad<2784> -> junta<7217> -> tratar<12394> -> gestionar<5960> -> diligencia<4105> -> administrativo<297>
1	10	3	admiración<300> -> admiración<300>
6	4	15	admitir<304> -> entrada<4755> -> plato<9634> -> recipiente<10512> -> general<5920> -> común<2767> -> admitir<304>
1	5	1	adoquín<311> -> adoquín<311>
2	12	1	adormidera<317> -> opio<8809> -> adormidera<317>
2	19	12	adverbio<330> -> invariable<7066> -> adverbio<330>
2	2	3	advertir<337> -> aconsejar<185> -> advertir<337>
2	1	1	aeroplano<343> -> avión<1388> -> aeroplano<343>
7	3	6	afectivo<353> -> sensible<11391> -> natural<8413> -> sencillo<11376> -> sincero<11541> -> hipocresía<6334> - > sentimiento<11414> -> afectivo<353>
2	5	30	afecto<355> -> cariño<2070> -> afecto<355>
2	2	3	afectuoso<357> -> cariñoso<2072> -> afectuoso<357>

LC	FS	FE	Ciclo
2	4	2	afeitar<358> -> rasurar<10420> -> afeitar<358>
2	4	7	afición<360> -> interés<6987> -> afición<360>
10	5	7	afilarse<364> -> lapicero<7301> -> mina<8111> -> lápiz<7302> -> escribir<4906> -> trazar<12415> -> diseñar<4173> -> esbozo<4856> -> posterior<9794> -> boca<1620> -> afilar<364>
2	5	7	afirmar<370> -> asegurar<1162> -> afirmar<370>
3	2	6	aflicción<372> -> congoja<2908> -> angustia<750> -> aflicción<372>
2	4	8	agarrar<378> -> asir<1199> -> agarrar<378>
2	2	1	agasajar<381> -> obsequiar<8672> -> agasajar<381>
2	3	2	ágil<387> -> ligero<7472> -> ágil<387>
2	5	6	agitar<388> -> revolver<11001> -> agitar<388>
3	5	5	agradar<392> -> grato<6094> -> agradable<390> -> agradar<392>
19	3	3	agricultura<404> -> cultivar<3471> -> científico<2362> -> método<8065> -> ejercicio<4459> -> teoría<12099> -> síntesis<11551> -> exposición<5325> -> cultura<3479> -> social<11605> -> negocio<8464> -> negociar<8461> -> comerciar<2643> -> lucrativo<7635> -> utilidad<12565> -> cosa<3255> -> posesión<9768> -> hacienda<6212> -> agrícola<402> -> agricultura<404>
1	9	1	agrimonia<406> -> agrimonia<406>
2	3	5	agrupación<410> -> agrupar<413> -> agrupación<410>
2	29	92	agua<414> -> fuente<5796> -> agua<414>
1	3	1	agua<416> -> agua<416>
1	4	2	agua<418> -> agua<418>
2	5	4	aguantar<420> -> resistir<10888> -> aguantar<420>
17	4	17	aguar<423> -> mezclar<8078> -> deshacer<3875> -> quitar<10357> -> arrebatar<1075> -> fuego<5794> -> combustión<2624> -> oxidación<8938> -> electrón<4489> -> entorno<4752> -> intervalo<7034> -> punto<10293> -> convencional<3110> -> costumbre<3278> -> temperamento<12053> -> práctico<9825> -> navegación<8435> -> aguar<423>
2	8	5	agudo<434> -> alto6<630> -> agudo<434>
3	6	13	agujero<447> -> perforar<9379> -> agujerear<446> -> agujero<447>
2	4	2	ahora<449> -> enseguida<4722> -> ahora<449>
2	17	35	aire<452> -> argón<1011> -> aire<452>
2	3	10	aislado<458> -> suelto<11792> -> aislado<458>
2	11	2	ajedrez<461> -> casilla<2137> -> ajedrez<461>
1	16	1	ajo<469> -> ajo<469>

LC	FS	FE	Ciclo
2	4	8	ajustar<473> -> acoplar<189> -> ajustar<473>
2	14	13	ala<478> -> insecto<6901> -> ala<478>
2	1	5	alabar<483> -> elogiar<4524> -> alabar<483>
3	4	1	alambique<484> -> destilación<3957> -> destilar<3959> -> alambique<484>
1	4	1	alambre<486> -> alambre<486>
1	8	2	albedrío<493> -> albedrío<493>
1	7	1	alcalde<500> -> alcalde<500>
2	5	2	alcance<505> -> tiro<12215> -> alcance<505>
2	18	5	alcohol<515> -> etílico<5195> -> alcohol<515>
1	5	7	alcohólico<519> -> alcohólico<519>
2	11	1	alcornoque<520> -> corcho<3149> -> alcornoque<520>
1	4	2	alegría<528> -> alegría<528>
1	3	2	alfiler<543> -> alfiler<543>
2	13	1	algarrobo<546> -> algarroba<545> -> algarrobo<546>
2	12	4	algodón<550> -> borra<1657> -> algodón<550>
2	2	1	algodón<551> -> hilado<6326> -> algodón<551>
2	1	1	aligerar<556> -> aliviar<572> -> aligerar<556>
2	5	7	alimentar<561> -> mantener<7804> -> alimentar<561>
1	17	3	almeja<581> -> almeja<581>
2	7	1	almendra<582> -> almendro<584> -> almendra<582>
1	3	1	almendra<583> -> almendra<583>
1	3	1	almíbar<585> -> almíbar<585>
2	9	2	almidón<587> -> apresto<934> -> almidón<587>
2	8	1	alminar<588> -> almuédano<591> -> alminar<588>
2	1	1	almohada<589> -> almohadón<590> -> almohada<589>
2	3	5	alojamiento<593> -> alojar<595> -> alojamiento<593>
2	3	22	alrededor<600> -> rodear<11066> -> alrededor<600>
3	3	1	altanero<604> -> altivo<621> -> arrogante<1090> -> altanero<604>
2	4	3	altar<606> -> monumento<8226> -> altar<606>
2	3	1	altisonante<618> -> ampuloso<706> -> altisonante<618>
8	4	1	alto<626> -> gran<6069> -> grande<6073> -> importancia<6586> -> conveniente<3113> -> positivo<9783> -> práctico<9824> -> muy<8374> -> alto<626>
1	5	2	alto9<632> -> alto9<632>
2	4	7	altura<638> -> altitud<620> -> altura<638>
2	15	2	aluminio<644> -> alúmina<643> -> aluminio<644>
10	9	9	alumno<646> -> examen<5234> -> estudio<5179> -> pintor<9562> -> ventana<12720> -> muro<8352> -> tapia<11971> -> pared<9079> -> superficie<11838> -> externo<5362> -> alumno<646>

LC	FS	FE	Ciclo
2	4	2	alveolo<647> -> raíz<10386> -> alveolo<647>
1	3	1	ama<649> -> ama<649>
16	3	4	amable<650> -> complaciente<2705> -> propenso<10115> -> aficionado<362> -> profesional<10058> -> deporte<3716> -> competitivo<2700> -> competencia<2692> -> aspirar<1231> -> aspiración<1227> -> soplo<11697> -> patológico<9222> -> patología<9221> -> estudiar<5176> -> seguir<11322> -> atento<1271> -> amable<650>
2	4	1	amaestrar<652> -> adiestrar<281> -> amaestrar<652>
1	4	2	amanecer<654> -> amanecer<654>
2	2	1	amansar<655> -> domesticar<4299> -> amansar<655>
1	6	2	amenaza<677> -> amenaza<677>
1	3	2	amistad<686> -> amistad<686>
2	9	1	macero<7663> -> maza<7912> -> macero<7663>
2	1	1	mácula<7669> -> mancha<7749> -> mácula<7669>
2	4	1	madre<7680> -> maternidad<7892> -> madre<7680>
3	3	1	maleza<7721> -> espesura<5013> -> matorral<7896> -> maleza<7721>
1	4	1	malla<7726> -> malla<7726>
2	8	3	mama<7736> -> leche<7340> -> mama<7736>
10	1	4	manda<7752> -> oferta<8737> -> regalo<10650> -> dádiva<3514> -> gratuito<6096> -> arbitrario<982> -> arbitrio<984> -> impuesto<6623> -> obligatorio<8658> -> precepto<9833> -> manda<7752>
2	9	1	mandarino<7758> -> mandarina<7757> -> mandarino<7758>
2	5	1	manecilla<7766> -> aguja<442> -> manecilla<7766>
2	6	3	manejable<7767> -> fácil<5407> -> manejable<7767>
2	3	24	manifestar<7784> -> manifestación<7780> -> manifestar<7784>
2	7	11	mano<7790> -> muñeca<8344> -> mano<7790>
2	5	7	mano<7791> -> dedo<3596> -> mano<7791>
2	3	1	mañoso<7799> -> habilidoso<6165> -> mañoso<7799>
2	7	1	manubrio<7815> -> medusa<7970> -> manubrio<7815>
36	10	4	manuscrito<7816> -> histórico<6343> -> verificar<12758> -> pronosticar<10099> -> predecir<9854> -> futuro<5849> -> enunciación<4793> -> enunciar<4795> -> claro<2459> -> sinceridad<11540> -> franqueza<5753> -> exención<5267> -> privilegio<10001> -> derivar<3748> -> derivada<3744> -> variable<12624> -> original<8887> -> letra<7393> -> escritura<4913> -> caligrafía<1855> -> letra<7398> -> humanístico<6455> -> humanidad<6454> -

LC	FS	FE	Ciclo
			> indulgencia<6768> -> disposición<4211> -> salud<11210> -> espiritual<5030> -> canto<1975> -> cuchillo<3418> -> acero<152> -> aleación<522> -> metal<8046> -> brillante<1701> -> sobresaliente<11594> -> calificación<1850> -> nota<8561> -> manuscrito<7816>
2	10	1	manzano<7819> -> manzana<7817> -> manzano<7819>
2	10	34	máquina<7822> -> pieza<9530> -> máquina<7822>
15	8	29	mar<7824> -> salado<11175> -> tostado<12295> -> tostar<12296> -> fuego<5794> -> combustión<2624> -> oxidación<8938> -> electrón<4489> -> átomo<1282> -> conserva<2971> -> hermético<6281> -> impenetrable<6571> -> penetrar<9310> -> interior<7000> -> costa<3265> -> mar<7824>
2	4	5	marco<7844> -> ventana<12721> -> marco<7844>
2	13	4	marea<7846> -> agua<415> -> marea<7846>
2	12	1	margarita<7849> -> compuesto<2762> -> margarita<7849>
2	9	2	marqués<7862> -> conde<2837> -> marqués<7862>
2	14	32	masa<7867> -> pan<9014> -> masa<7867>
3	5	35	materia<7883> -> corpóreo<3171> -> material<7889> -> materia<7883>
2	6	11	materia<7884> -> espíritu<5027> -> materia<7884>
2	5	51	materia<7885> -> hablar<6184> -> materia<7885>
2	4	7	materia<7886> -> asignatura<1192> -> materia<7886>
10	7	5	máxima<7904> -> personal<9430> -> gramática<6064> -> lenguaje<7376> -> dicha<4049> -> felicidad<5513> -> dicho<4050> -> frase<5755> -> significado<11492> -> oración<8825> -> máxima<7904>
2	13	48	mayor<7907> -> grande<6073> -> mayor<7907>
1	2	1	mayor<7909> -> mayor<7909>
18	4	9	mayoría<7910> -> número<8617> -> representar<10819> -> figura<5576> -> expresión<5334> -> frase<5755> -> significado<11492> -> oración<8825> -> unidad<12510> -> indivisible<6753> -> dividir<4259> -> separar<11429> -> autónomo<1341> -> trabajar<12306> -> esmero<4960> -> gran<6069> -> singular<11545> -> común<2767> -> mayoría<7910>
2	2	1	mazorca<7914> -> maíz<7704> -> mazorca<7914>
2	4	2	media<7931> -> pie<9513> -> media<7931>
2	2	3	mediar<7936> -> interceder<6984> -> mediar<7936>
2	6	19	medicamento<7938> -> administrar<294> -> medicamento<7938>
12	9	14	medio6<7957> -> abertura<21> -> paso1<9188> ->

LC	FS	FE	Ciclo
			castizo<2156> -> puro<10316> -> vicioso<12814> -> vicio<12813> -> interior<7000> -> contraste<3091> -> contrastar<3090> -> contraste<3092> -> televisión<12035> -> medio6<7957>
2	3	57	medio9<7958> -> camino<1906> -> medio9<7958>
2	7	10	medir<7962> -> magnitud<7702> -> medir<7962>
2	2	1	melena<7978> -> león<7387> -> melena<7978>
2	1	1	melocotón<7982> -> melocotonero<7983> -> melocotón<7982>
28	5	6	memoria<7990> -> exposición<5328> -> inicial<6857> -> letra<7394> -> composición<2732> -> aguardiente<426> - > destilación<3958> -> soltar<11659> -> aligerar<555> -> acelerar<133> -> accionar<124> -> gesto<5961> -> costumbre<3278> -> temperamento<12053> -> práctico<9825> -> navegación<8435> -> nave<8434> -> surcar<11873> -> surco<11875> -> cicatriz<2348> -> llaga<7546> -> úlcera<12501> -> mucosa<8301> -> húmedo<6462> -> impregnar<6597> -> empapar<4573> - > absorber<61> -> retener<10954> -> memoria<7990>
7	9	24	menor<7998> -> menos<8000> -> cualitativo<3386> -> relativo<10722> -> condicionado<2850> -> condicional<2851> -> incluir<6672> -> menor<7998>
1	4	24	menos<8003> -> menos<8003>
2	6	1	mentol<8014> -> menta<8008> -> mentol<8014>
2	6	7	mérito<8027> -> digno<4100> -> mérito<8027>
1	4	1	mesa<8034> -> mesa<8034>
2	7	1	meteorito<8052> -> aerolito<341> -> meteorito<8052>
2	5	14	método<8064> -> orden<8836> -> método<8064>
19	4	1	mezclar<8076> -> meter<8056> -> liar<7425> -> atar<1258> -> inmovilizar<6875> -> deje<3643> -> acento<136> -> tono<12265> -> relieve<10731> -> superficie<11838> -> externo<5362> -> clase<2464> -> grado<6044> -> dividir<4259> -> separar<11429> -> autónomo<1341> -> trabajar<12306> -> simple<11533> - > mezcla<8071> -> mezclar<8076>
2	3	1	miau<8081> -> maullido<7901> -> miao<8081>
3	11	2	microscopio<8087> -> sistema<11561> -> objetivo<8638> -> microscopio<8087>
9	6	21	miembro<8091> -> articulado<1110> -> pieza<9534> -> pesca<9465> -> pescar<9467> -> caña<1926> -> bota<1664> -> calzado<1876> -> pierna<9527> -> miembro<8091>
2	2	3	militar<8105> -> milicia<8100> -> militar<8105>

LC	FS	FE	Ciclo
3	2	1	mina<8112> -> artefacto<1102> -> carga<2057> -> mina<8112>
2	6	25	mineral<8113> -> inorgánico<6889> -> mineral<8113>
2	6	4	ministerio<8116> -> departamento<3705> -> ministerio<8116>
10	6	3	ministro<8118> -> país<8974> -> geográfico<5945> -> geografía<5944> -> superficie<11838> -> externo<5362> -> colegio<2569> -> profesión<10055> -> oficio<8746> -> ministerio<8117> -> ministro<8118>
4	2	9	minucioso<8121> -> detallista<3987> -> detalle<3985> -> pormenorizar<9749> -> minucioso<8121>
2	2	3	miramiento<8128> -> respeto<10909> -> miramiento<8128>
8	5	2	mirar<8129> -> vista<12897> -> ojo<8763> -> luminoso<7645> -> emitir<4565> -> lanzar<7298> -> arrojar<1092> -> reparar<10787> -> mirar<8129>
13	8	6	misión<8137> -> gobierno<5998> -> gobernador<5990> -> civil<2436> -> eclesiástico<4382> -> iglesia<6512> -> culto<3477> -> sufrir<11812> -> expiar<5302> -> sacrificio<11161> -> abnegación<27> -> hacer<6205> -> función<5820> -> misión<8137>
14	3	4	moda<8149> -> pasajero<9153> -> pasa<9143> -> secar<11268> -> humedad<6460> -> impregnar<6596> -> combinar<2620> -> concertar<2805> -> pactar<8952> -> condición<2846> -> exigencia<5271> -> reclamación<10519> -> cual<3382> -> introducir<7058> -> moda<8149>
2	5	13	modelo<8153> -> pauta<9238> -> modelo<8153>
2	4	1	modelo<8156> -> maniquí<7788> -> modelo<8156>
2	3	4	moderar<8160> -> templar<12060> -> moderar<8160>
2	14	22	modo<8167> -> habla<6178> -> modo<8167>
2	3	3	mojar<8175> -> humedecer<6461> -> mojar<8175>
2	4	1	mojón<8176> -> señal<11361> -> mojón<8176>
17	14	19	molécula<8181> -> constituyente<3007> -> reformar<10635> -> rehacer<10688> -> mal<7708> -> calamidad<1822> -> gente<5942> -> parentela<9082> -> pariente<9085> -> lazo<7333> -> cinta<2393> -> verde<12746> -> seca<11267> -> sequía<11437> -> sequedad<11436> -> humedad<6460> -> impregnar<6596> -> molécula<8181>
2	3	7	molestar<8186> -> incomodar<6678> -> molestar<8186>
2	11	11	molusco<8192> -> caracol<2024> -> molusco<8192>
2	8	25	moneda<8201> -> pieza<9532> -> moneda<8201>

LC	FS	FE	Ciclo
2	3	1	montacargas<8212> -> ascensor<1152> -> montacargas<8212>
2	1	2	montar<8216> -> cabalgar<1764> -> montar<8216>
2	3	1	mordisquear<8240> -> morder<8238> -> mordisquear<8240>
2	2	1	morral<8250> -> mochila<8147> -> morral<8250>
2	7	1	morse<8251> -> alfabeto<540> -> morse<8251>
2	3	1	mortificar<8255> -> castigar<2153> -> mortificar<8255>
1	6	1	mostaza<8261> -> mostaza<8261>
2	2	13	motivo<8271> -> causa<2187> -> motivo<8271>
2	2	29	mover<8280> -> menear<7995> -> mover<8280>
2	2	5	movimiento<8296> -> ademán<264> -> movimiento<8296>
2	3	1	mucosidad<8302> -> mucoso<8304> -> mucosidad<8302>
2	7	15	mueble<8306> -> mobiliario<8146> -> mueble<8306>
2	4	1	muela<8309> -> piedra<9520> -> muela<8309>
2	6	2	muelle<8310> -> espiral<5023> -> muelle<8310>
2	9	1	muelle<8311> -> andén<732> -> muelle<8311>
2	2	2	muestra<8320> -> prueba<10218> -> muestra<8320>
2	2	1	mugre<8321> -> grasa<6090> -> mugre<8321>
2	3	12	mujer<8322> -> femenino<5517> -> mujer<8322>
2	9	2	multa<8327> -> sanción<11225> -> multa<8327>
30	3	2	multiplicación<8329> -> multiplicar<8331> -> llamada<7551> -> remite<10759> -> indicación<6728> -> informe<6828> -> exposición<5326> -> clase<2465> -> profesor<10060> -> enseña<4723> -> insignia<6909> -> legión<7354> -> milicia<8102> -> gente<5942> -> parentela<9082> -> pariente<9085> -> lazo<7333> -> adorno<320> -> embellecer<4537> -> bello<1554> -> belleza<1553> -> admiración<299> -> admirar<303> -> extraordinario<5376> -> comida<2655> -> alimento<566> -> legal<7351> -> cumplimiento<3483> -> cumplir<3485> -> llevar<7585> -> multiplicación<8329>
2	3	1	mundo<8343> -> secular<11295> -> mundo<8343>
2	2	1	municipio<8347> -> municipal<8345> -> municipio<8347>
2	6	2	muralla<8349> -> muro<8353> -> muralla<8349>
2	10	10	música<8360> -> armonía<1048> -> música<8360>
2	5	3	músico<8367> -> música<8362> -> músico<8367>
2	4	9	nacer<8375> -> salir<11190> -> nacer<8375>
2	13	3	naranja<8395> -> corteza<3243> -> naranja<8395>
2	2	2	naranja<8396> -> anaranjado<717> -> naranja<8396>
2	3	5	narrar<8402> -> contar<3038> -> narrar<8402>

LC	FS	FE	Ciclo
4	4	2	natural<8416> -> espontáneo<5040> -> propio<10125> -> artificial<1120> -> natural<8416>
2	4	2	natural<8417> -> bemol<1557> -> natural<8417>
2	3	1	necedad<8438> -> tontería<12267> -> necedad<8438>
2	4	5	necesario<8440> -> espontáneo<5042> -> necesario<8440>
13	3	14	tipo<12198> -> clase<2466> -> cuenta<3432> -> explicación<5304> -> exposición<5329> -> cardinal<2054> -> numeral<8614> -> designar<3882> -> destinar<3960> -> señalar<11369> -> nombrar<8534> -> designar<3883> -> nombrar<8533> -> tipo<12198>
2	12	1	tiranía<12206> -> tirano<12207> -> tiranía<12206>
2	1	1	tiro<12214> -> disparo<4200> -> tiro<12214>
15	5	8	título<12220> -> certificado<2299> -> certificar<2300> -> correo<3196> -> oficina<8744> -> laboratorio<7256> -> equipar<4832> -> víveres<12911> -> provisión<10190> -> reserva<10856> -> custodia<3512> -> metal<8046> -> brillante<1701> -> sobresaliente<11594> -> examen<5234> -> título<12220>
22	4	19	título<12221> -> nobiliario<8517> -> nobleza<8522> -> noble<8520> -> honrado<6388> -> norma<8545> -> correcto<3183> -> regla<10677> -> ley<7419> -> regir<10662> -> exigir<5273> -> enérgico<4662> -> eficaz<4429> -> idóneo<6511> -> aptitud<957> -> apropiado<948> -> destinar<3960> -> señalar<11369> -> nombrar<8534> -> designar<3883> -> nombrar<8533> -> honorífico<6383> -> título<12221>
2	4	8	tocar<12230> -> llegar<7573> -> tocar<12230>
2	2	26	tomar<12251> -> beber<1547> -> tomar<12251>
2	3	4	tomar7<12258> -> confundir<2901> -> tomar7<12258>
3	6	2	tónica<12262> -> cafeína<1810> -> cola<2555> -> tónica<12262>
10	16	11	tono<12263> -> segundo<11325> -> responsabilidad<10918> -> obligación<8650> -> moral<8231> -> conducta<2859> -> manejar<7770> -> gobernar<5994> -> guiar<6145> -> aconsejar<185> -> tono<12263>
2	4	1	tonsura<12266> -> coronilla<3167> -> tonsura<12266>
2	8	1	tordo<12276> -> mirlo<8131> -> tordo<12276>
2	5	3	tornillo<12280> -> tuerca<12480> -> tornillo<12280>
2	21	20	toro<12284> -> buey<1731> -> toro<12284>
2	3	3	trabajador<12305> -> obrero<8671> -> trabajador<12305>
8	2	3	trabajar<12309> -> cultivar<3471> -> científico<2362> ->

LC	FS	FE	Ciclo
			método<8065> -> ejercicio<4459> -> ayuda<1398> ->
			ayudar<1401> -> cooperar<3130> -> trabajar<12309>
2	7	10	trabajo<12313> -> capital<1999> -> trabajo<12313>
2	13	6	tradicción<12317> -> folclor<5668> -> tradición<12317>
2	2	2	tradicional<12319> -> clásico<2472> ->
			tradicional<12319>
2	1	1	traer<12325> -> atraer<1289> -> traer<12325>
2	3	5	tráfico<12326> -> tránsito<12354> -> tráfico<12326>
2	6	1	tragaluz<12327> -> claraboya<2448> -> tragaluz<12327>
2	3	2	traje<12332> -> gala<5851> -> traje<12332>
2	3	4	trámite<12334> -> diligencia<4105> -> trámite<12334>
2	5	1	trampa<12337> -> cepo<2258> -> trampa<12337>
2	7	4	tranquilo<12341> -> sosegado<11712> ->
			tranquilo<12341>
2	3	7	transcurrir<12345> -> correr<3202> -> transcurrir<12345>
2	4	3	transmitir<12361> -> contagiar<3035> ->
			transmitir<12361>
2	8	2	transparencia<12363> -> retroproyector<10977> ->
			transparencia<12363>
2	17	3	tráquea<12377> -> bronquio<1716> -> tráquea<12377>
2	1	1	trasero<12379> -> nalga<8394> -> trasero<12379>
2	3	7	trasladar<12381> -> llevar<7580> -> trasladar<12381>
24	6	3	trastornar<12385> -> inquietar<6892> -> quitar<10357> ->
			arrebatar<1075> -> fuego<5794> -> combustión<2624> ->
			oxidación<8938> -> electrón<4489> -> largo<7311> ->
			sílaba<11508> -> voz<12978> -> coral<3144> ->
			cantar<1966> -> juego2<7186> -> deporte<3716> ->
			competitivo<2700> -> competencia<2692> ->
			aspirar<1231> -> aspiración<1227> -> soplo<11697> ->
			patológico<9222> -> patología<9221> -> funcional<5826>
			-> trastorno<12387> -> trastornar<12385>
2	2	3	tren<12422> -> ferrocarril<5534> -> tren<12422>
2	9	1	trenca<12425> -> capucha<2016> -> trenca<12425>
2	7	2	trenza<12426> -> ramal<10396> -> trenza<12426>
2	6	11	tribunal<12439> -> juez<7189> -> tribunal<12439>
2	19	7	trigo<12441> -> espiga<5014> -> trigo<12441>
2	5	1	trío<12444> -> tres<12434> -> trío<12444>
3	3	7	tristeza<12450> -> triste<12448> -> melancolía<7976> ->
			tristeza<12450>
2	4	1	triturar<12451> -> moler<8184> -> triturar<12451>
2	5	4	tropa<12463> -> ejército<4468> -> tropa<12463>
2	10	4	tubérculo<12470> -> patata<9211> -> tubérculo<12470>
2	11	2	tulipa<12481> -> tulipán<12482> -> tulipa<12481>

LC	FS	FE	Ciclo
3	3	2	turbación<12489> -> turbar<12492> -> aturdir<1308> -> turbación<12489>
2	3	1	ubicuidad<12499> -> omnipresencia<8785> -> ubicuidad<12499>
11	8	44	unidad<12512> -> comparación<2672> -> cotejar<3282> -> vista<12896> -> distinguir<4230> -> diferenciar<4074> -> dos<4324> -> representar<10819> -> figura<5576> -> expresión<5334> -> frase<5755> -> unidad<12512>
34	5	10	unidad<12513> -> independiente<6721> -> dependencia<3709> -> negocio<8463> -> intentar<6976> -> ejecución<4443> -> pictórico<9507> -> pintura<9567> -> detallar<3983> -> contar<3039> -> merecer<8022> -> hacer<6205> -> desempeñar<3854> -> teatro<12001> -> actor<221> -> protagonista<10161> -> ficción<5554> -> fingir<5615> -> pasar<9165> -> tacha<11927> -> defecto<3601> -> propio<10125> -> contraposición<3083> -> contraponer<3082> -> contrarrestar<3088> -> frente<5765> -> fachada<5404> -> principal<9974> -> planta<9615> -> piso<9575> -> vivienda<12915> -> habitación<6168> -> cuarto<3391> -> familia<5453> -> unidad<12513>
2	3	6	universo<12534> -> universal<12528> -> universo<12534>
12	9	159	uno<12535> -> nombre<8535> -> objeto<8644> -> término<12118> -> extremo<5390> -> llegar<7573> -> turno<12493> -> orden<8835> -> disposición<4213> -> superior<11846> -> comunidad<2786> -> privado<9991> -> uno<12535>
4	4	29	uno4<12540> -> neutro<8486> -> operación<8800> -> aplicación<899> -> uno4<12540>
2	4	56	usar<12552> -> utilizar<12568> -> usar<12552>
2	2	36	uso<12556> -> costumbre<3279> -> uso<12556>
2	3	17	utensilio<12560> -> instrumento<6947> -> utensilio<12560>
2	2	9	útil<12563> -> herramienta<6287> -> útil<12563>
2	10	5	uva<12569> -> racimo<10364> -> uva<12569>
2	1	1	vacuno<12580> -> bovino<1673> -> vacuno<12580>
2	5	3	vallado<12601> -> cerca<2262> -> vallado<12601>
16	6	25	valor<12602> -> pagar<8968> -> ostentación<8924> -> ostentar<8926> -> manifiesto<7786> -> declaración<3579> -> personal<9430> -> gramática<6064> -> lenguaje<7376> -> dicha<4049> -> felicidad<5513> -> bien<1578> -> sano<11237> ->

LC	FS	FE	Ciclo
			salud<11209> -> tanto<11965> -> punto<10292> -> valor<12602>
14	5	8	valor<12605> -> patrón<9230> -> tomar6<12257> -> moho<8173> -> óxido<8940> -> neón<8474> -> frío<5776> -> verde<12745> -> espectro<4996> -> horrible<6410> -> horrendo<6409> -> terror<12145> -> miedo<8088> -> real<10457> -> valor<12605>
2	4	20	valor<12608> -> acción<121> -> valor<12608>
2	4	1	vanagloriarse<12614> -> presumir<9942> -> vanagloriarse<12614>
2	6	3	variante<12630> -> modalidad<8151> -> variante<12630>
2	2	6	variedad<12633> -> diversidad<4252> -> variedad<12633>
2	5	1	varillaje<12637> -> paraguas<9055> -> varillaje<12637>
2	9	7	vaso<12645> -> savia<11262> -> vaso<12645>
2	3	3	vaso<12646> -> casco<2128> -> vaso<12646>
2	2	1	vecindad<12651> -> vecino<12655> -> vecindad<12651>
2	2	2	vectorial<12657> -> vector<12656> -> vectorial<12657>
2	6	4	vela<12677> -> lona<7614> -> vela<12677>
2	10	2	velar<12682> -> posterior<9793> -> velar<12682>
2	3	1	velero<12684> -> goleta<6001> -> velero<12684>
2	4	1	vello<12685> -> pelusa<9291> -> vello<12685>
2	4	2	veloz<12692> -> ágil<386> -> veloz<12692>
2	3	1	vencimiento<12699> -> plazo<9642> -> vencimiento<12699>
2	8	1	venda<12700> -> gasa<5892> -> venda<12700>
2	4	4	vender<12703> -> mercancía<8018> -> vender<12703>
2	4	11	venenoso<12707> -> tóxico<12300> -> venenoso<12707>
2	10	1	ventrículo<12726> -> aurícula<1324> -> ventrículo<12726>
11	22	13	verbal<12736> -> aportar<911> -> extraviar<5380> -> perder<9352> -> nombre<8535> -> objeto<8644> -> son<11683> -> agradable<390> -> satisfacer<11259> -> exigencia<5271> -> reclamación<10519> -> verbal<12736>
2	21	33	verbo<12737> -> voz<12979> -> verbo<12737>
2	5	2	verja<12759> -> enrejado<4711> -> verja<12759>
2	23	25	verso<12766> -> estrofa<5160> -> verso<12766>
2	7	6	vértebra<12768> -> vertebral<12771> -> vértebra<12768>
2	9	1	vértigo<12777> -> mareo<7847> -> vértigo<12777>
2	8	1	vestíbulo<12779> -> entrada<4756> -> vestíbulo<12779>
3	4	2	vestimenta<12786> -> vestir<12787> -> ropa<11088> -> vestimenta<12786>

LC	FS	FE	Ciclo
2	2	3	vestir<12790> -> vestido<12781> -> vestir<12790>
2	23	1	víbora<12806> -> ofidio<8750> -> víbora<12806>
16	3	2	vibrar<12809> -> trémulo<12421> -> temblar<12048> -> involuntario<7091> -> reflejo<10627> -> reflejar<10626> -> choque<2338> -> sangre<11230> -> metabólico<8042> -> metabolismo<8043> -> asimilación<1194> -> asimilar<1197> -> asimilación<1195> -> articular<1116> -> articulación<1108> -> vocal<12931> -> vibrar<12809>
3	5	2	víctima<12815> -> sacrificio<11159> -> expiatorio<5303> -> víctima<12815>
22	5	14	vida<12817> -> funcionamiento<5827> -> funcionar<5829> -> desempeñar<3854> -> teatro<12001> -> actor<221> -> protagonista<10161> -> ficción<5554> -> fingir<5615> -> pasar1<9165> -> tacha<11927> -> falta<5447> -> tara<11977> -> hereditario<6270> -> herencia<6272> -> serie<11445> -> emisión<4556> -> electromagnético<4488> -> fenómeno<5524> -> deforme<3629> -> irregular<7121> -> conducta<2859> -> vida<12817>
2	9	4	vigente<12839> -> regir<10663> -> vigente<12839>
17	5	16	vino<12852> -> fermentación<5529> -> intervención<7037> -> oficina<8744> -> laboratorio<7256> -> equipar<4832> -> dotar<4331> -> buque<1742> -> cubierta<3402> -> suelo<11791> -> casco<2130> -> uña<12506> -> dorsal<4322> -> prueba<10216> -> saborear<11140> -> sabor<11139> -> boca<1621> -> vino<12852>
2	5	5	virtud<12879> -> virtuoso<12881> -> virtud<12879>
2	5	1	víscera<12884> -> abdomen<16> -> víscera<12884>
2	1	1	vistazo<12902> -> ojeada<8761> -> vistazo<12902>
2	3	8	viveza<12913> -> calor<1869> -> viveza<12913>
2	1	1	vivir<12919> -> habitar<6171> -> vivir<12919>
2	3	14	volumen<12944> -> bulto<1739> -> volumen<12944>
2	5	23	voluntad<12951> -> gana<5868> -> voluntad<12951>
2	8	17	volver<12959> -> convertir<3123> -> volver<12959>
7	7	15	voz<12974> -> vibrar<12808> -> elástico<4477> -> caucho<2180> -> látex<7320> -> vegetal<12663> -> llamar<7552> -> voz<12974>
2	11	2	vulgar<12991> -> vulgarismo<12992> -> vulgar<12991>
2	7	1	yeyuno<12997> -> duodeno<4360> -> yeyuno<12997>
2	15	1	zarza<13004> -> zarzamora<13005> -> zarza<13004>
2	7	10	zona<13008> -> círculo<2413> -> zona<13008>

*Anexo 3. Lista de los ciclos encontrados*

---

---

<b>LC</b>	<b>FS</b>	<b>FE</b>	<b>Ciclo</b>
2	6	1	zurdo<13013> -> izquierdo<7142> -> zurdo<13013>
2	2	1	zurra<13014> -> azotaina<1407> -> zurra<13014>

---

## Anexo 4. Ejemplos de ciclos en el diccionario.

---

En este anexo se podrán ver las definiciones reales de algunas palabras<sup>16</sup>. Después tenemos un cuadro con las palabras que se usan en su definición (palabras a las que apuntan las flechas salientes de dicha palabra) y las palabras que la usan en su definición (palabras de las que provienen las flechas entrantes a la palabra en cuestión). Antes de cada palabra en el cuadro aparece el identificador de ésta. Entre paréntesis tenemos la longitud de uno de los ciclos en los que está involucrada la palabra. En algunos casos se ilustran los ciclos con gráficas<sup>17</sup>.

### AVENA (1)

**avena** I. Del lat. *avena* 1. (sustantivo femenino). Nombre por el que se conocen otras plantas del género *avena*, como la ballueca, que crece en los prados y se confunde fácilmente con la *avena*.

---

<sup>16</sup> Estas definiciones están tomadas del diccionario Anaya (1997).

<sup>17</sup> Las flechas punteadas en las gráficas indican que no aparecen las palabras a las que se dirigen, ya que no tienen que ver con el ciclo que se quiere ilustrar, para sintetizar.

1380:avena	Palabras que se usan en su definición (flechas salientes)
	1381:avena, 2899:confundir, 2940:conocer, 5407:fácil, 5487:fatuo, 5931:género, 8537:nombre, 9613:planta, 9826:prado
1380:avena	Palabras que la usan en su definición (flechas entrantes)
	1380:avena

## CÁMARA (1)

**cámara** I. Del lat. vulg. Camara, del gr. kamara = bóveda, cuarto abovedado 1.(femenino, sustantivo masculino). Quien maneja la cámara de cine o televisión.

1886:cámara	Palabras que se usan en su definición (flechas salientes)
	1886:cámara, 2383:cine, 7770:manejar 12036:televisión
1886:cámara	Palabras que la usan en su definición (flechas entrantes)
	1886:cámara, 3090:contrastar

## MOSTAZA (1)

**mostaza** I. De mosto 1.(sustantivo femenino). Salsa preparada con la semilla de la mostaza, de sabor picante.

8261:mostaza	Palabras que se usan en su definición (flechas salientes) 8261:mostaza, 9497:picante, 9896:preparado, 11138:sabor, 11200:salsa, 11353:semilla
	Palabras que la usan en su definición (flechas entrantes) 8261:mostaza

## PIRATA (1)

**pirata** I. Del lat. pirata, del gr. peirates = bandido, peirao = yo me aventuro 1.(sustantivo masculino). Ladrón de los mares que saqueaba los barcos y navegaba sin licencia de ningún gobierno. 2.(adjetivo, -a). Relativo o perteneciente al pirata.

9571:pirata	Palabras que se usan en su definición (flechas salientes) 9571:pirata, 10722:relativo,
	Palabras que la usan en su definición (flechas entrantes) 9571:pirata

## FAMILIA(2)

**familia** I. Del lat. familia 2.(sustantivo femenino). En un linaje, conjunto de descendientes, ascendientes, colaterales y afines.

**linaje** I. Del cat. llinatge cat. ant. llinyatge llinya lat. linea = línea 1. (sustantivo masculino). Línea de antepasados o descendientes de una familia. 2. (sustantivo masculino) figurado,

-a, figuradamente Clase o condición de una cosa. 3. (plural) antiguo, -a, anticuado, -a, antiguamente. Personas de la nobleza.

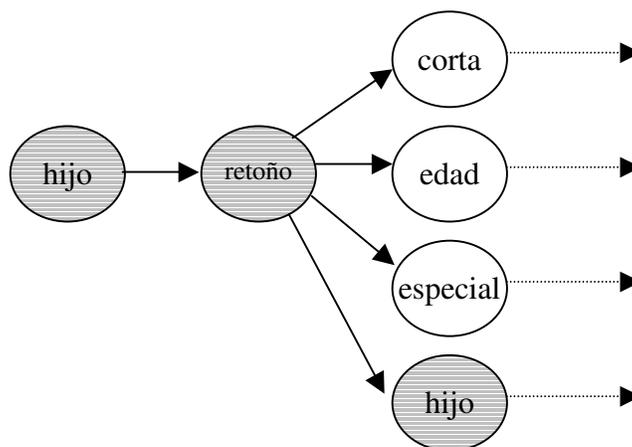
<b>5454: familia</b>	<b>Palabras que se usan en su definición (flechas salientes)</b>  366:afín, 1149:ascendiente, 2563:colateral, 2925:conjunto, 3808:descendiente, <b>7500:linaje</b> ,
	<b>Palabras que la usan en su definición (flechas entrantes)</b>  6937:institución, 7324:latín, <b>7500:linaje</b> , 8883:origen,

## **HIJO(2)**

**hijo**, -a I. Del lat. filius 1.(sustantivo masculino) Persona o animal, respecto de sus padres. 2. (femenino, sustantivo masculino). El yerno o la nuera, respecto a los suegros. 3.(femenino, sustantivo masculino). Expresión cariñosa. 4. (sustantivo masculino). Retoño. 5. (femenino, sustantivo masculino) figurado, -a, figuradamente. Producto de la inteligencia. 6.(femenino, sustantivo masculino). Cualquier persona, respecto al país o comarca donde ha nacido. 7.(femenino, sustantivo masculino). Religioso, respecto al fundador de su orden. 8. (sustantivo masculino, plural). Descendientes. 9. H. adoptivo. El que lo es por adopción.

**retoño** I. De re- + otoño 1.(sustantivo masculino). Tallo que echa de nuevo la planta. 2. (sustantivo masculino) familiar. Hijo, en especial el de corta edad. FAM. Retoñar.

6325:hijo	Palabras que se usan en su definición (flechas salientes)
	10964:retoño
	Palabras que la usan en su definición (flechas entrantes)
	308:adoptar, 9087:parir, 10964:retoño, 12227:toca,



## PADRE(2)

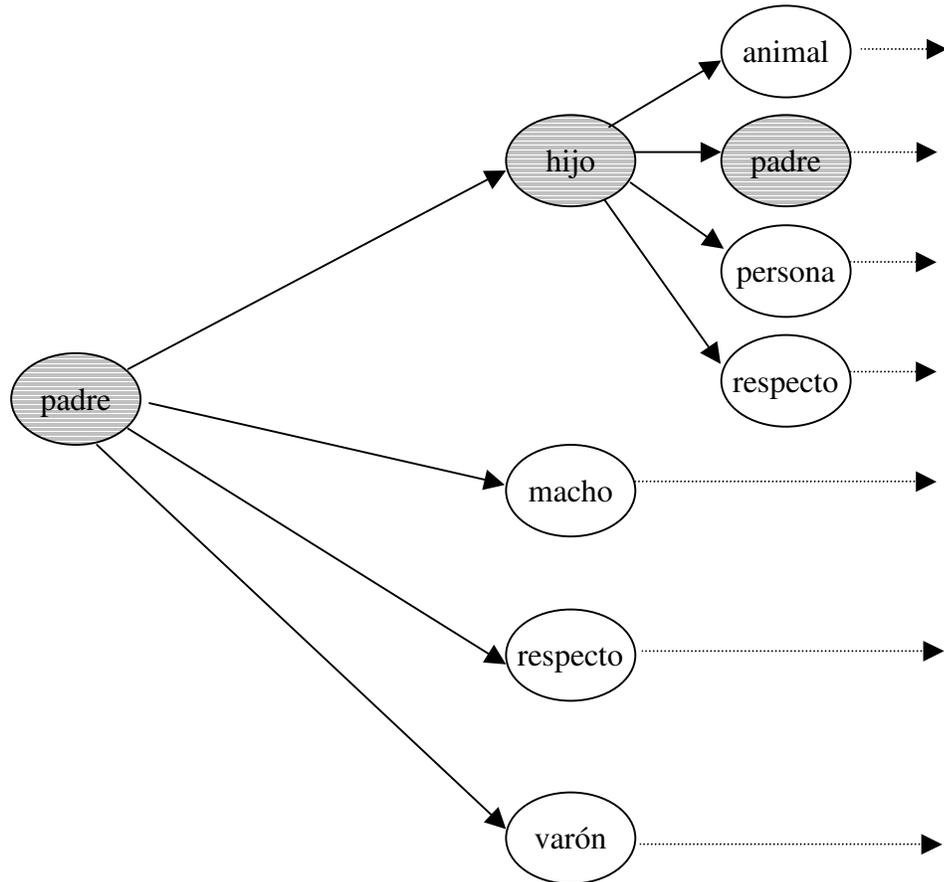
**padre** I. Del lat. pater, -tris 1.(sustantivo masculino). Varón o macho que ha engendrado. 2.(sustantivo masculino). Varón o macho respecto de sus hijos. 3.(sustantivo masculino). Semental. 4.(sustantivo masculino). Título dado a los sacerdotes o religiosos. 5. (plural). El padre y la madre. 6.(plural) figurado, -a, figuradamente. Antepasados.

**hijo, -a** I. Del lat. filius 1.(sustantivo masculino) Persona o animal, respecto de sus padres. 2. (femenino, sustantivo masculino). El yerno o la nuera, respecto a los suegros. 3.(femenino, sustantivo masculino). Expresión cariñosa. 4. (sustantivo masculino). Retoño. 5. (femenino, sustantivo masculino) figurado, -a, figuradamente. Producto de la

*Anexo 4. Ejemplos de ciclos en el diccionario.*

inteligencia. 6. (femenino, sustantivo masculino). Cualquier persona, respecto al país o comarca donde ha nacido. 7. (femenino, sustantivo masculino). Religioso, respecto al fundador de su orden. 8. (sustantivo masculino, plural). Descendientes. 9. H. adoptivo. El que lo es por adopción.

<b>8958:padre</b>	<b>Palabras que se usan en su definición (flechas salientes)</b>
	6323:hijo, 7666:macho, 10905:respecto, 12639:varón,
	<b>Palabras que la usan en su definición (flechas entrantes)</b>
	876:apellido, 6272:herencia, 6323:hijo, 9425:persona,



### CONJETURA(3)

**conjetura** I. Del lat. coniectura < conicere = opinar 1. (sustantivo femenino). Opinión basada en indicios, no en **pruebas**. FAM. Conjetural, conjeturar. SIN. Suposición, presunción, hipótesis, sospecha.

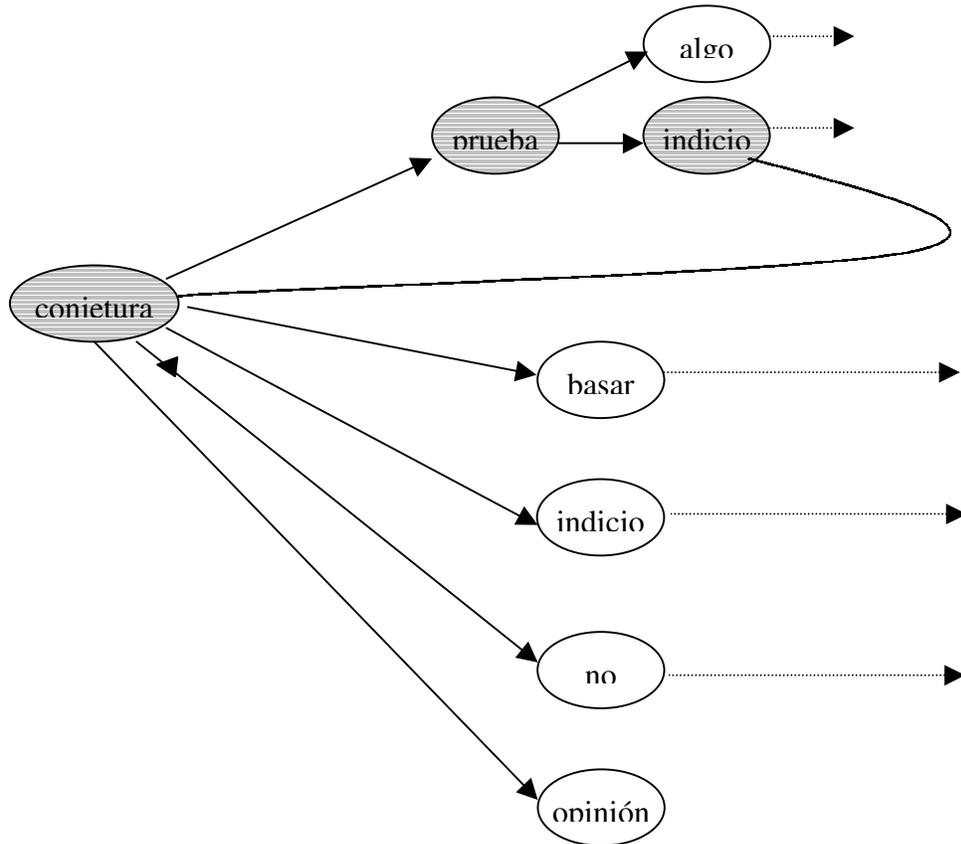
**prueba** I. De probar 1. (sustantivo femenino). Acción y efecto de probar. 2. (sustantivo femenino). Razón o testimonio para demostrar la verdad o falsedad de una cosa. 3. (sustantivo femenino). **Indicio** de algo 4. (sustantivo femenino). Competición, carrera. 5. (sustantivo femenino). Ensayo, experiencia 6. (sustantivo femenino). Operación para comprobar la exactitud de otra anterior 7. (sustantivo femenino). Justificación de la verdad de los hechos controvertidos en un juicio. 8. (sustantivo femenino). Pequeña cantidad de un producto comestible que se saborea para ver si es bueno o malo. 9. (sustantivo femenino). E: Cuestionario, test. 10. (sustantivo femenino). Primera muestra de la composición tipográfica, que se saca para corregir las erratas que tiene. 11. (sustantivo femenino). Muestras del grabado y de la fotografía. 12. A p. de. (locución adverbial). Hecho a la perfección a p. de bombas 13. A p. (locución adverbial). Que se tiene para comprobar su estado o condiciones 14. En p. de. (locución adverbial). Como señal de.

**indicio** I. Del lat. indicium = revelación < index, -icis 1. (sustantivo masculino). **Conjetura** o señal que posibilita el conocimiento de algo que ha existido o que va a ocurrir 2. (sustantivo masculino). Asomo, cantidad muy pequeña o primera manifestación de algo 3. (sustantivo masculino). Clase de signo que mantiene con el objeto significado una relación de dependencia, pero no de semejanza SIN. 1. Vestigio. 1 y 3. Síntoma, signo, manifestación.

Anexo 4. Ejemplos de ciclos en el diccionario.

2916:conjetura	Palabras que se usan en su definición (flechas salientes) 1512:basar, 6736:indicio, 8514:no, 8808:opinión, 10212:prueba,
	Palabras que la usan en su definición (flechas entrantes)  1832:cálculo, 6337:hipótesis, 6735:indicio, 11719:sospechar,

10212:prueba,	Palabras que se usan en su definición (flechas salientes) 549:algo, 6735:indicio,
	Palabras que la usan en su definición (flechas entrantes)  2916:conjetura



### IMAGINARIO(3)

**imaginario**, -a I. De imaginar 1. (adjetivo, -a). Que sólo existe en la imaginación; aparente, **irreal**. 2. (adjetivo, -a). En matemáticas, dicese del número complejo cuya parte real es nula. FAM. Imaginaria, imaginariamente.

**irreal** I. De in-, con sentido negativo + real 1. (adjetivo, -a). No real, **ideal**, imaginado, fantástico, ilusorio. FAM. Irrealidad.

**ideal** I. De idea 1. (adjetivo, -a). Perfecto. *Ha realizado un trabajo i.* 2. (adjetivo, -a). **Imaginario**. *La perfección absoluta es i.* 3. (adjetivo, -a). Relativo a la idea. 4. (sustantivo masculino). Prototipo. *El i. de belleza.* 5. (plural, sustantivo masculino, singular). Serie de ideas que se presentan como meta individual o colectiva, y a las que se tiende por considerarlas el mayor bien. *Es importante actuar movidos por un i.* FAM. Idealismo, idealista, idealizar. SIN. 1. Sublime, excelente, fantástico. 2. Irreal, inmaterial. 4. Modelo, arquetipo, canon. 5. Ilusión, ambición, deseo, sueño.

