

CRITERIOS SINTÁCTICOS APLICADOS A UN PROGRAMA RECUPERADOR DE INFORMACIÓN

SONIA MORETT ÁLVAREZ
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Resumen: *Una alternativa para expandir las posibilidades de búsqueda dentro de un diccionario onomasiológico electrónico es la consideración de agrupamientos semánticos, esto es, grupos de palabras que en contextos determinados pueden funcionar como sinónimos. Clustering es un programa computacional que genera automáticamente dichos agrupamientos. Bajo la hipótesis de que ciertas deficiencias sistemáticas en la localización de agrupamientos semánticos obedecen a que el sistema (que se basa en la comparación paralela de las cadenas de palabras que componen diferentes definiciones sobre un mismo término) alinea las secuencias palabra por palabra -sin considerar su función dentro de un sintagma; aplicamos un análisis sintáctico funcional a las definiciones de un banco terminológico. Siguiendo la metodología de la lingüística de corpus, a partir de la observación -sustentada en un aparato crítico- de los alineamientos se perfilaron los aspectos que deben considerarse para optimizar el programa: Perífrasis gramaticales, adjetivos, adverbios, sintagmas enfáticos y determinantes.*

Palabras clave: Agrupamientos semánticos, Clustering, pares semánticos, pares iguales, pares correspondientes, pares vinculados, pares nulos, pares semi nulos, pares semi iguales.

Abstract: *An alternative in order to expand the search possibilities within an electronic onomasiological dictionary is the taking into account semantic clusters, which are groups of words that may act as synonyms in certain contexts. Clustering is a system software that generates this groupings automatically. The system is based on a parallel comparison between the word chains that make up different definitions of one and the same term, so it lines up the word sequences word for word –without considering the function it plays within the syntagma. Considering the hypothesis which maintains that certain systematic faults regarding the location of [semantic groupings] have to do with the system, we have carried out a functional and syntactic analysis of a terminology bank definitions. Following the linguistic of corpus methodology, and from a based on a critical apparatus observation of the alignments, we have outlined the features that should be taken into account in order to optimize the software: grammatical periphrasis, adjectives, adverbs and stressing syntagmas.*

Keywords: semantic clusters, Clustering, semantic pairs, equal pairs, matched pairs, bindings, null pairs, semi-null pairs, semi-equal pairs.

1. Introducción

El tema de la presente comunicación surge en el seno del Grupo de Ingeniería Lingüística (GIL) de la UNAM y está inscrito en esta área interdisciplinaria de investigación aplicada, la cual se sirve de conocimientos lingüísticos para mejorar el funcionamiento de los sistemas informáticos.

A partir del análisis del funcionamiento de un programa computacional denominado Clustering, se busca construir propuestas, desde la sintaxis, para mejorar el desempeño del mismo programa y, en general, el de los sistemas recuperadores de información.

2. Antecedentes

Debido a que este trabajo consiste en una propuesta de mejoramiento para el programa de alineamiento de pares semánticos empezaremos por conocerlo; con tal propósito, presentamos el proyecto del que forma parte, el Diccionario Electrónico de Búsqueda Onomasiológica (DEBO) (2.1), lo cual es necesario para entender la utilidad de Clustering; así como el banco terminológico del que se nutre (2.2), para llegar a describir el programa (2.3).

2.1. Diccionario Electrónico de Búsqueda Onomasiológica

Hasta este momento, el DEBO constituye el desarrollo más importante del GIL. Sus características principales están contenidas en el propio nombre: electrónico y onomasiológico.

Con *electrónico* se quiere decir que la colección de voces y sus definiciones se presenta en un formato digital y que la información se recupera mediante un sistema computacional.

Lo *onomasiológico* se refiere al modo de búsqueda. Mientras que los diccionarios semasiológicos parten de un término para acceder a los significados que le corresponden, en los onomasiológicos la incógnita es el término, pues se puede describir el concepto pero se desconoce el significante que lo denota.

En la era digital, los diccionarios onomasiológicos han adquirido mayor relevancia gracias a la enorme capacidad de almacenamiento de los sistemas computacionales y a la flexibilidad que ofrece la automatización de procesos para el acceso a la información contenida en ellos.

El hecho de consultar un diccionario a través de una computadora no hace más eficientes las búsquedas onomasiológicas automáticamente. Para que las posibilidades de éxito para estas búsquedas se multipliquen, es necesario que las definiciones contenidas en el diccionario reciban algún tipo de preprocesamiento y así puedan ser explotadas con diferentes herramientas para el tratamiento de bases de datos léxicos.

Una posibilidad para expandir las búsquedas consiste en la determinación de *paradigmas semánticos*, formados por «el conjunto de *palabras clave*¹ que presentan rasgos comunes y que pueden ser utilizadas con el mismo sentido en el contexto del término final al que corresponden, esto es, puede ser sustituido cualquier miembro del paradigma en los elementos correspondientes del sintagma sin cambiar el significado del mismo.» (G. SIERRA 1996)

Por ejemplo, *realizar, producir, hacer y efectuar* constituyen un paradigma semántico, pues en las siguientes definiciones de *energía* podemos sustituir la palabra correspondiente del paradigma por cualquiera de las otras tres sin alterar el sentido de la frase:

- Capacidad de un sistema físico para ***realizar*** trabajo. (*Encarta 98*)
- Capacidad de un sistema material para ***producir*** trabajo, con las propiedades de la conservación y la interconvertibilidad. (*Océano Uno Color Diccionario Enciclopédico*)
- Propiedad de un sistema -su capacidad para ***hacer*** un trabajo. (Definición traducida de John Daintith, *Dictionary of physics*)
- Es la capacidad de ***efectuar*** trabajo. (*Física Weber*)

2.2. Banco terminológico

El DEBO se nutre con la información léxica contenida en un banco terminológico implementado en el GIL para este propósito², el cual contiene definiciones provenientes de diferentes fuentes para cada uno de los términos (ordenados por área de especialidad) que maneja el mismo banco.

Además de la información contenida en textos especializados y en diccionarios, para los propósitos de una búsqueda onomasiológica no debe descuidarse la conceptualización de los hablantes. Por esto, encontramos en el banco terminológico muchas definiciones cuya fuente es *personal*, las cuales enriquecen de manera importante la información léxica con la que se sustenta Clustering.

2.3. Descripción de Clustering

2.3.1. Algoritmo básico de agrupamiento semántico

Bajo el título de Clustering, Gerardo Sierra³ y John McNaught⁴ desarrollaron entre 1999 y 2000 el programa de alineamiento de pares semánticos que, a partir de una revisión en el año 2002, se identificó con el nombre de *Algoritmo básico de agrupamiento semántico*. Dicho programa es un sistema de recuperación de información que tiene como finalidad agrupar paradigmas semánticos.

En una primera etapa, el sistema alinea dos definiciones de un mismo término; el procedimiento se repite hasta agotar todas las posibles combinaciones de pares de definiciones para cada término.

Una vez con los alineamientos, se localizan los *pares iguales* (pares de palabras que comparten posición y lema -no necesariamente flexiones-), los *pares nulos* (Se consideran así las palabras que dentro de un alineamiento no tienen una paralela en la otra definición) y los *pares correspondientes* (parejas de palabras diferentes que ocupan una posición paralela dentro del alineamiento).

Véase en un ejemplo (1):

Def. 1	Energía	Capacidad	de	un	sistema	material	para	producir	trabajo	con	las	...
Def. 2	Energía	Capacidad	de	un	sistema	físico	para	realizar	trabajo			
Tipo	par igual	par igual	par igual	par igual	par igual	par correspondiente	par igual	par correspondiente	Par igual	par nulo	par nulo	pares nulos

Posteriormente y mediante el cálculo de un coeficiente de similitud, denominado *Longest Collocation Couple (LCC)*, se determina si los pares correspondientes se promueven a *pares vinculados*. Éstos son los que el sistema identifica como posibles *pares semánticos* (pares de formas que en un contexto determinado pueden intercambiarse sin alterar el significado de la estructura que los engloba).

En la siguiente tabla (2), se muestran los índices de LCC para el alineamiento de nuestro ejemplo anterior:

Def. 1	Energía	Capacidad	de	un	sistema	material	para	producir	trabajo	con	las	...
Def. 2	Energía	Capacidad	de	un	sistema	físico	para	realizar	trabajo			
Tipo	par igual	par igual	par igual	par igual	par igual	par correspondiente	par igual	par correspondiente	par igual	par nulo	par nulo	pares nulos
LCC	0	0	0	0	0	7	0	3	0	0	0	0
Par vinculado	no	No	no	no	No	sí	no	no	no	no	no	no

Lo que se buscase con este coeficiente es que los pares vinculados coincidan con los pares semánticos que una revisión humana del alineamiento identifica. Su valor se obtiene al contabilizar el número de pares iguales inmediatamente anteriores o posteriores a un par

correspondiente dentro de un alineamiento (hasta la aparición de un par nulo u otro par correspondiente), más una unidad que suma el par correspondiente en examen.

De la fusión de los pares semánticos que comparten palabras se obtienen los paradigmas semánticos. Así es como, a partir de los pares {realizar, producir}, {realizar, hacer} y {hacer, efectuar}, se consiguió el paradigma {realizar, producir, hacer, efectuar}.

2.3.2. Algoritmo flexibilizado de agrupamiento semántico

El éxito del DEBO dependerá, en gran medida, de que los agrupamientos que genere Clustering sean precisos. Con el fin de optimizar el rendimiento del programa y como resultado de la evaluación hecha en 2002 se llevaron a cabo modificaciones al sistema que se integraron en el *Algoritmo flexibilizado de alineamiento semántico*. Las nuevas aplicaciones no se incorporan automáticamente a las búsquedas dentro de Clustering, sino que se presentan como opcionales⁵. Éstas son la consideración de *pares semi nulos*, consideración de *pares semi iguales* e *inversión de dos palabras*.

2.3.2.1. Consideración de pares semi nulos

Se denomina par *semi nulo* a un par nulo formado por una palabra funcional y un espacio vacío en la cadena, debido a que:

Las palabras contenidas en la lista de palabras irrelevantes no aportan información relevante durante el proceso de identificación de pares vinculados, por lo tanto, los pares nulos que agrupen palabras de esta lista, y sólo de esta lista, pueden ser considerados pares iguales, bajo la óptica de que alinear palabras irrelevantes con la cadena vacía (ε) puede ser equivalente, en este contexto, a insertar la palabra involucrada en lugar de ε. (G. CASTILLO 2002: 28).

El propósito de establecer este nuevo tipo de pares es aumentar el valor de LCC de los pares correspondientes que aparezcan --en los alineamientos-- cerca de los semi nulos, ya que éstos, al ser considerados iguales, también lo son para el cálculo de LCC.

2.3.2.2. Consideración de pares semi iguales

Se sustentan en la misma premisa que los pares semi nulos. Un par semi igual es aquél que está formado por dos palabras funcionales y que, para efectos de cómputo de LCC, se considera equivalente a un par igual.

2.3.2.3. Inversión de dos palabras

La finalidad de esta operación es realizar intercambios de formas consecutivas dentro de una definición, sin que se modifique el sentido de la frase, para que sea posible promover a vinculados pares correspondientes que no alcanzaban el valor de LCC necesario en el algoritmo básico. Ejemplo (3):

Resultados obtenidos con la aplicación del algoritmo básico:

Def. 1	Estroboscopio	Aparato	de	alta	velocidad	que	permite	congelar	...
Def. 2	Estroboscopio	Instrumento	de	velocidad	alta	que	permite		...
LCC	0	3	0	2	3	0	0	0	
Tipo*	I	C	I	C	C	I	I	N	

* Tipos de pares identificados: I = par igual; C = par correspondiente; N = par nulo.

Resultados obtenidos con la opción de inversión de dos palabras:

Def. 1	Estroboscopio	Aparato	de	alta	velocidad	que	permite	congelar	...
Def. 2	Estroboscopio	Instrumento	de	velocidad	alta	que	permite		...
LCC	0	7	0	0	0	0	0	0	
Tipo*	I	C	I	Intercambio	Intercambio	I	I	N	

* Tipos de pares identificados: I = par igual; C = par correspondiente; N = par nulo.

Además, se experimentó con una operación derivada de la anterior que no se había desarrollado previamente en otros sistemas, el *intercambio conjuntivo*, es decir, el de palabras no contiguas vinculadas mediante una conjunción.

2.3.2.4. Evaluación al algoritmo flexibilizado

Después de realizar una evaluación sistemática con 32 combinaciones diferentes de parámetros para Clustering, se concluyó que las alternativas del *Algoritmo flexibilizado* relajan las restricciones del algoritmo original con resultados que incrementan considerablemente los pares vinculados que el sistema identifica, pero que hacen disminuir la correspondencia entre pares vinculados y semánticos.

Como se ha mencionado, la finalidad de Clustering es expandir las posibilidades de búsqueda onomasiológica en un diccionario y para que esto suceda el sistema debe mejorarse continuamente. En cuanto a nuestra área de estudio --la lengua española-- hay mucho que hacer, pues Clustering presenta una serie de deficiencias derivadas de que fue diseñado originalmente para la lengua inglesa.

3. Propuestas

A partir de los resultados obtenidos por el sistema se delimitaron, entre otros, los siguientes aspectos que resultan problemáticos para el sistema y cuya explicación está en la sintaxis.

La instrumentación de las propuestas requiere de un análisis sintáctico previo y de un etiquetado de categorías gramaticales aplicado a las definiciones con las que trabaja Clustering.

3.1. Perífrasis gramaticales

Se refiere a aquellas construcciones gramaticales formadas por más de una palabra gráfica que se comportan como unidad indivisible para efectos semánticos y sintácticos. Actualmente, Clustering agrupa por separado las palabras que hacen parte de una perífrasis. En consecuencia, su presencia desencadena desajustes dentro de los alineamientos.

Consideramos tres tipos de perífrasis: locuciones, términos compuestos y perífrasis verbales.

3.1.1. Locuciones

Se trata de estructuras que semántica y sintácticamente forman una unidad pero aparecen separadas en la escritura. Ejemplo (4):

Los alineamientos que hemos analizado son ricos en construcciones perifrásticas, como son abundantes en español. Observamos que su presencia desencadena el que aparezcan como pares correspondientes palabras con diferente categoría gramatical. Resulta lógico inferir que si consideramos verbo a toda estructura gramatical que funcionalmente actúa como tal, se obtendrán alineamientos más precisos, lo que redundará en mejores pares semánticos. Por tanto, proponemos que se marquen como unidades léxicas no sólo aquellos verbos en los que el proceso que describen se manifiesta con más de una palabra (los tiempos compuestos de la conjugación y los verbos pronominales en donde el pronombre *se* aparece en posición proclítica); sino también las perífrasis verbales. Dicha identificación podría hacerse a partir de información gramatical inserta en el motor de Clustering.

3.1.4. Preposiciones regidas por un verbo

Al buscar pares semánticos manualmente, encontramos otro fenómeno que obstaculiza el correcto alineamiento de las definiciones. Se trata de los pares correspondientes que desencadena la unión de un verbo que rige preposición con otro verbo transitivo en un mismo par correspondiente. Tal es el caso del par {trata, estudia} en el ejemplo (7):

Término: Cinemática
Fuente de la definición 1: Vox
Fuente de la definición 2: Larousse

Def. 1	C	Parte	de	la	mecánica	que	trata	del	movimiento	en	sus	condiciones			de	espacio	y	tiempo		
Def. 2	C	Parte	de	la	mecánica	que	estudia	el	movimiento	de	los	cuerpos	con	independencia	de	las	fuerzas	que	lo	producen
Tipo de par	I	I	I	I	I	I	C	C	I	C	C	C	N	N	I	C	C	C	N	N
LCC	0	0	0	0	0	0	7	2	0	2	1	1	0	0	0	2	1	1	0	0
Par vinculado	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no

En el ejemplo anterior el suplemento de la primera definición y el implemento de la segunda desempeñan la misma función transitiva. Uno de los argumentos con los que se sostiene la relación de identidad semántica entre los complementos oracionales mencionados es que en los suplementos hay una pérdida progresiva del significado propio de la preposición. Por lo tanto, si ésta carece de significado (salvo el de indicar relación), no queda nada que diferencie, en el plano del contenido, suplemento de implemento.

Considerando todo lo anterior, vemos que obtendríamos más pares semánticos acertados si se tomara la preposición como un incremento del verbo. Tendría, entonces, que incorporarse al sistema la información que permitiera a la preposición que sigue a verbos como *tratar* aparecer como un elemento constitutivo suyo. Esto es factible en tanto son contados los verbos que rigen preposición y cada uno de ellos determina una específica. Además, los verbos que rigen preposición aparecen enlistados en gramáticas. Dichas listas podrían introducirse como archivo oculto a Clustering, a pesar de que no todos sus elementos están representados en el corpus, pues esto evitaría etiquetar manualmente cada ocurrencia.

3.2. Adjetivos

Entendemos por adjetivo toda unidad gramatical, independientemente de su clase sintagmática, que desempeña la función de modificador nominal. Ejemplo (8):

Término: Primera ley de Newton

Fuente de la definición 1: Física universitaria Sears

Fuente de la definición 2: El mundo de la Física 1

Def. 1	P	l	d	N	Todo	cuerpo	continúa	en	su	estado	de	reposo	o	de	movimiento	rectilíneo	uniforme				a	...
Def. 2	P	l	d	N	Todo	cuerpo	continúa	en	su	estado	de	reposo	o	de	movimiento		uniforme	en	línea	recta	a	...
Tipo de par	I	I	I	I	I	I	I	I	I	I	I	I	I	N	I	N	I	N	N	N	I	
LCC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Par vinculado	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no

Def. 1	...	menos	que	sea	impelido	a	cambiar	dicho	estado	por	fuerzas	que	actúen	sobre	él
Def. 2	...	menos	que	sea	obligado	a	cambiar	este	estado	por	fuerzas	impresas		sobre	él
Tipo de par	I	I	I	I	C	I	I	C	I	I	I	C	N	I	I
LCC	0	0	0	0	7	0	0	6	0	0	0	4	0	0	0
Par vinculado	no	no	no	no	sí	no	no	sí	no	no	no	no	no	no	no

La propuesta es considerar adjetivos, frases adjetivas, frases preposicionales en función de complemento adnominal y oraciones adjetivas como equivalentes en los alineamientos.

Para los alineamientos que contienen pares formados por un adjetivo y uno de los sintagmas mencionados, o bien, dos de ellos, esto puede realizarse --una vez etiquetadas las categorías gramaticales en las definiciones— con la aplicación de un etiquetador sintáctico parcial (denominado *chunking*). El cual es capaz de localizar frases y determinar con qué categoría gramatical se corresponden. Sin embargo, el procedimiento parece más complicado para los casos que involucran oraciones, por lo que sugerimos que este aspecto se considere a largo plazo.

3.3. Adverbios

Se considera adverbio a toda unidad gramatical, independientemente de su clase sintagmática, que desempeña la función de adyacente circunstancial. Ejemplo (9):

Término: Péndulo

Fuente de la definición 1: Clave

Fuente de la definición 2: Larousse

Def. 1	P	Cuerpo	que	suspendido	de	un	punto	que	está	por	encima	de	su	centro	de	gravedad	puede	oscilar	libremente	alrededor	de	dicho	punto	...	
Def. 2	P	Cuerpo	rigido				que											oscila	con	libertad	alrededor	de	un	eje	...
Tipo de par	I	I	C	N	N	N	I	N	N	N	N	N	N	N	N	N	N	I	C	N	I	C	C		
LCC	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3	1	
Par vinculado	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	

El procedimiento sugerido para detectar pares de adjetivos y estructuras análogas serviría también para los adverbios: Aplicar un *chunking* a las definiciones --después de haber llevado a cabo el etiquetado de las categorías gramaticales (POST)— permitiría al sistema reconocer de una sola vez frases adjetivas, adverbiales, sustantivas y preposicionales, para ser tratadas del mismo modo que las palabras con las que comparten la categoría gramatical que en conjunto desempeñan.

3.4. Sintagmas enfáticos

Con *sintagmas enfáticos* nos referiremos a todas aquellas estructuras (independientemente de su categoría sintagmática) cuyo valor es enfático o redundante en un contexto dado, ya que no aportan información semántica relevante a las definiciones en el proceso de alineamiento. Los algoritmos básico y flexibilizado los alinean con palabras que cumplen otras funciones. Ejemplo (10):

Término: Energía

Fuente de la definición 1: CIm1

Fuente de la definición 2: Física Weber

Def. 1	E	Capacidad	que	tiene	un	sistema	físico	de	realizar	trabajo
Def. 2	E	Es	la	capacidad				de	efectuar	trabajo
Tipo de par	I	C	C	C	N	N	N	I	C	I
LCC	0	2	1	1	0	0	0	0	3	0
Par vinculado	no	no	no	no	no	no	no	no	sí	no

Consideramos que los sintagmas enfáticos que tuvieran como correspondiente un espacio vacío deberían tratarse como semi nulos. El reconocimiento de este tipo de pares, sin embargo, obedece más a criterios semánticos que a sintácticos; entonces, resulta muy importante la intervención humana para decidir cuándo un adyacente es innecesario.

3.5. Determinantes

La determinación es una categoría funcional que abarca las operaciones déicticas, anafóricas, referenciales y de cuantificación. Las categorías gramaticales que cumplen estas funciones son artículos, demostrativos, pronombres, posesivos e indefinidos.

El algoritmo alinea palabras pertenecientes a esta categoría con palabras de cualquier otra categoría. Cuando dos palabras de este tipo se unen en un par, el algoritmo básico lo puede considerar vinculado; el flexibilizado, le da el tratamiento de par semi igual pero indiscriminadamente. Ejemplo (11):

Término: Dirección de un vector

Fuente de la definición 1: Gran enciclopedia Larousse

Fuente de la definición 2: Teoría y problemas en ciencias físicas

Def. 1	D	d	u	v	La	recta	a	que	pertenece	la	longitud	del	segmento	de	un	vector		es	su	dirección	
Def. 2	D	d	u	v	Es	una	flecha	cuya	longitud	y	orientación	representa	el	módulo	y	la	dirección	y	sentido	del	vector
Tipo de par	I	I	I	I	C	C	C	C	C	C	C	C	C	C	C	N	N	I	C	C	
LCC	0	0	0	0	5	1	1	1	1	1	1	1	1	1	1	1	0	0	0	2	1
Par vinculado	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no

En el ejemplo anterior el *la* que precede a *recta* y el *una* que hace lo propio con *flecha* coinciden en ser artículos con género femenino y número singular; el único rasgo que los diferencia es el de la *definitud*: El hablante –en este caso, el lexicógrafo— opta por el artículo definido cuando en el contexto de uso se cumple la *condición de unicidad* (es decir, que el referente sea el único objeto que satisfaga la descripción aportada), pero cuando la *flecha* no es la única posible se opta por el indefinido. Si *la* y *una* comparten la mayoría de sus propiedades esenciales, pertenecen a la misma categoría gramatical y desempeñan las mismas funciones sintácticas y semánticas, no hay razón para tratarlas como palabras diferentes, sino que deben implementarse los mecanismos para que el programa los identifique como variantes de una sola expresión. El fenómeno se repite en el alineamiento: determinando a *vector*, encontramos en la primera definición, el artículo indefinido, *un*, mientras que en la segunda, esta función la cumple su correspondiente definido *el*.

En el mismo alineamiento aparece una muestra más de diferentes manifestaciones de determinantes que semánticamente cumplen una sola función. Se trata del posesivo *su* y el artículo *la*, que en la primera y segunda definición, respectivamente, afectan a *dirección*. En el primer caso se justifica el empleo del posesivo porque el verbo copulativo indica que la frase *su dirección* se refiere a *vector*; mientras que en la segunda definición esa relación de posesión se expresa con el genitivo *del* (de + el) *vector*, que afecta simultáneamente a *dirección* y *sentido*.

En lo que se refiere a la determinación, proponemos que los pares correspondientes de este tipo sean considerados semi iguales por el sistema siempre y cuando los dos miembros del par se identifiquen con la misma categoría gramatical, lo cual es perfectamente viable, puesto

que los determinantes constituyen paradigmas cerrados cuya posición sintáctica es estable, por lo que pueden ser fácilmente identificados por el sistema de manera automática.

4. Conclusiones

Con esta investigación hemos respondido a la necesidad que muy claramente plantea Gabriel Castillo (2002: 80) sobre aplicar un análisis de corte lingüístico a las definiciones con que trabaja el banco terminológico del Grupo de Ingeniería Lingüística de la UNAM. Con lo que se da continuidad a una línea de investigación en proceso, lo cual constituye una aportación para la instrumentación del Diccionario Electrónico de Búsqueda Onomasiológica.

En cuanto a las expectativas planteadas, se han delimitado los aspectos sintácticos que deben tomarse en consideración para las posteriores modificaciones al programa de alineamiento de pares semánticos. Incluso hemos propuesto, en muchos de los casos, alternativas específicas sobre cómo podría aprovecharse el análisis que hacemos sin realizar cambios profundos al sistema.

Por otra parte, debemos señalar que el trabajo realizado consistió en un estudio sintáctico acotado a la lexicografía en el dominio de Física, la cual incorpora características que le son exclusivas; pero las iniciativas aportadas para una mejor recuperación de pares semánticos puede extenderse a solucionar problemas en la extracción de información relevante en otras bases de datos léxicos.

Notas

-
- 1 Una palabra clave es cualquier voz que pueda considerarse importante dentro de una definición.
 - 2 <http://iling.torreingenieria.unam.mx/diccionarios>
 - 3 Grupo de Ingeniería Lingüística, Instituto de Ingeniería, Universidad Nacional Autónoma de México
 - 4 Departamento de Ciencia y tecnología, UMIST, Manchester.
 - 5 <http://tabasco.torreingenieria.unam.mx/scripts/clusters.exe/Alineamiento>

Bibliografía

- CASTILLO HERNÁNDEZ, Gabriel, *Algoritmo revisado para la extracción automática de agrupamientos semánticos*, México, UNAM, 2002.
- CASTILLO, Gabriel y Gerardo SIERRA, «Algoritmo flexibilizado de agrupamiento semántico», *Estudios de Lingüística Aplicada*, núm. 38, año 22, México, diciembre de 2003.
- SIERRA MARTÍNEZ, Gerardo, «Avances en el desarrollo del Diccionario Electrónico de Búsqueda Onomasiológica», en *Actas del V Simposio iberoamericano de terminología RITerm*, Ciudad de México, del 3 al 8 de noviembre de 1996.
- SIERRA MARTÍNEZ, Gerardo y John MCNAUGHT, "Serendipitous Wording of POS Tags to Extract Semantic Pairs of Words from Dictionary Definitions", en OKATAN, A., *Computational Intelligence (actas de la Conferencia Internacional en Inteligencia Computacional 2004)*, Estambul, diciembre de 2004.

Cualquier consulta sobre el banco terminológico se puede realizar en la dirección:

<http://iling.torreingenieria.unam.mx/diccionarios>

Si se desea observar los alineamientos con cualquiera de las versiones del programa, consúltese:

<http://tabasco.torreingenieria.unam.mx/scripts/clusters.exe/Alineamiento>