

Reconocimiento de definiciones asociadas a frases predicativas en contextos definitorios

Recognition of definitions associated to predicative phrases in definitional contexts

César Aguilar

Facultad de Lenguas y Letras,
Universidad Autónoma de Querétaro
Centro Universitario, s/n C.P. 76010
Querétaro, México
Caguilar@iingen.unam.mx

Gerardo Sierra

Grupo de Ingeniería Lingüística,
Instituto de Ingeniería, UNAM
Cubículo 3, Basamento, Torre de Ingeniería
C.U., C.P. 04510, México D.F.
GsierraM@iingen.unam.mx

Resumen: En este trabajo se exponen los resultados de un experimento para el reconocimiento semiautomático de definiciones, considerando su relación con frases predicativas. Tal experimento considera el uso combinado de la llamada *teoría de la predicación* para delimitar patrones sintácticos, junto con el cálculo de probabilidades condicionales con miras a determinar la frecuencia de asociación entre definiciones y predicaciones.

Palabras clave: Definición, predicación verbal, contexto definitorio, teoría de la predicación, probabilidades condicionales.

Abstract: This paper exposes the results obtained from an experiment to recognise semi-automatically definitions, taking in mind its association with predicative phrases. This experiment considers the combined use of the called *theory of predication*, with the calculus of conditional probabilities for determining the frequency of these associations between definitions and predicative phrases.

Keywords: Definition, predicative phrase, definitional context, theory of predication, conditional probabilities.

1 Planteamiento del problema

Hoy en día, reconocer y extraer definiciones a partir de su codificación en patrones lingüísticos es una tarea importante para la lexicografía computacional y la terminología. Entre los patrones considerados para realizar esta tarea, sobresalen predicaciones verbales que recurrentemente ligan a las definiciones con sus correspondientes términos.

Si bien estas predicaciones han sido reportadas por varios autores para el inglés (Alshawi, 1987; Pearson, 1999; Meyer, 2001), o el francés (Auger, 1997; Rebeyrolle, 2000; Malaisé *et al.*, 2005) —por citar algunos casos—, estos autores no atienden las siguientes cuestiones:

- ¿Es posible explicar formalmente qué relación existe entre predicaciones y definiciones?

- ¿Puede determinarse, a partir de métodos estadísticos, cuál es la probabilidad de identificar una definición, dada una predicación?

Lo que se desarrolla aquí es un método semiautomático de extracción de definiciones, el cual atiende estas dos cuestiones. Por una parte, se propone una descripción de patrones sintácticos asociados a definiciones y predicaciones, con miras a explicar cuál es la relación que subyace entre ambas unidades.

Por otro lado, se presenta un análisis estadístico, basado en el uso de un método de Chi cuadrada y probabilidades condicionales, para evaluar si, dada una predicación, es posible reconocer una definición. La evaluación estadística que se hace del modelo sintáctico brinda un sustento empírico para validar la posibilidad de extraer definiciones en corpus a través de patrones sintácticos.

2 Definiciones en contextos definitorios

El sustento teórico y metodológico de este trabajo se ubica en el marco de la extracción de contextos definitorios en español (Sierra *et al.*, 2006). Por contexto definitorio (CD) se entiende cualquier fragmento textual en donde se introduzca y asocie un término con una definición. Para fijar tal asociación, se emplean conectores lingüísticos como predicaciones verbales, patrones pragmáticos y marcas tipográficas, entre otros (Alarcón, 2006).

Términos y definiciones son los elementos básicos de construcción de un CD. Una definición consiste en la descripción lingüística de un concepto asociado a un término, formada por dos elementos: un género próximo y una diferencia específica.

El género próximo equivale a un descriptor que indica la clase a la que pertenece una entidad, mientras que la diferencia específica configura una serie de rasgos que distinguen a dicha entidad de otras, agrupados por la clase que es referida por el género próximo.

A partir de estos elementos, Sierra *et al.* (2003) proponen una tipología basada en la presencia o ausencia del género próximo y/o la diferencia específica. En concreto, se consideran 4 tipos de definiciones:

- **Definición analítica o aristotélica:** es aquella que manifiesta un género próximo y una diferencia específica explícitas.
- **Definición sinonímica:** es aquella que presenta únicamente el género próximo, fijando una equivalencia conceptual con el término.
- **Definición funcional:** es aquella que solo indica la diferencia específica, la cual ofrece la descripción conceptual de una entidad a partir de su uso en una situación dada.
- **Definición extensional:** es aquella que hace explícita la diferencia específica, la cual enumera los componentes que conforman la entidad referida por el término.

Una forma de representar cómo se relacionan estos 4 tipos de definiciones es:

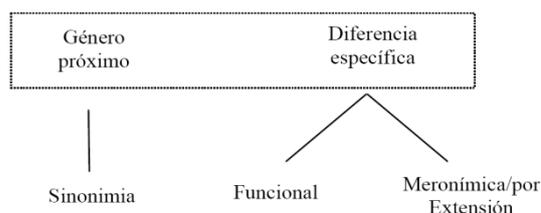


Figura 1: Tipología de definiciones (tomado de Sierra *et al.*, 2008: 81).

Estos 4 tipos, además de ser derivables a partir de la presencia del género próximo y/o la diferencia específica, mantienen un vínculo con el verbo que configura la predicación a la cual aparecen asociadas.

3 Predicaciones verbales

En un plano de estructuración sintáctica de un CD, una predicación organiza en qué posiciones pueden situarse el término y la definición en torno al verbo que opera como núcleo de dicha predicación.

Siguiendo el análisis de Sierra *et al.* (2008), se observan dos patrones regulares en las predicaciones asociadas a definiciones:

- I. Patrón término + verbo + definición, donde el término equivale al sujeto, el verbo funge como núcleo, y la definición es representada por el predicado que se asocia al sujeto, p. e.: *un error de programación es un fallo en la semántica de un programa.*
- II. Patrón autor + término + verbo + definición, donde el sujeto indica al autor de una definición, el término equivale al objeto de la predicación, el verbo opera como núcleo, y la definición es introducida por el predicado, p. e.: *Turing definió la inteligencia artificial como aquella inteligencia exhibida por artefactos creados por humanos.*

Estas distinciones entre los dos patrones puede explicarse en términos de predicaciones primarias y secundarias (Mallén 1991), las cuales son derivables de lo que se conoce como frase predicativa (PRED"). En términos formales, una frase predicativa se configura en torno a un núcleo funcional, el cual rige complementos que pueden ir desde una palabra hasta toda una oración (Bowers, 1993, 2001). Gráficamente, PRED" se describe así:

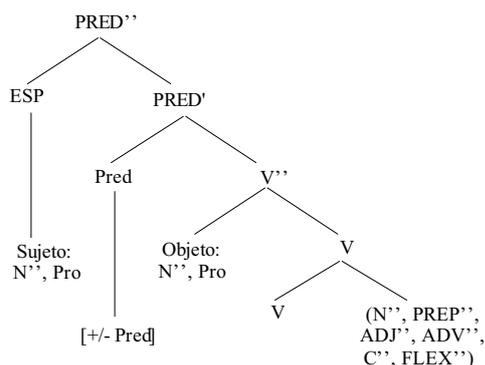


Figura 2: Esquema arbóreo de PRED'' (tomado de Bowers, 1993: 601)

Este esquema permite explicar casos como:

- 1) El contenedor refrigerado [es [una forma especializada de transporte de perecederos PRED''] V'']
- 2) En realidad, GPRS [es [similar a UMTS PRED''] V'']
- 3) El spyware [incluye [un gestor de Pop-ups y una función de búsqueda en Web PRED''] V''] FLEX''
- 4) La cuarentena VPN [permite [poner en marcha un tipo diferente de restricciones sobre los intentos de conexión PRED''] V'']
- 5) Carlos Godino [define la arquitectura naval [como la ciencia que trata de los conocimientos necesarios para la construcción de los buques PRED''] V'']

En los ejemplos que van de (1-4), se observa que verbos como *es*, *incluye* y *permite* proyectan frases verbales V'', cuyo argumento externo representa a un término, en tanto que PRED'' se liga a dicho patrón verbal, de tal suerte que los atributo aludidos por PRED'' configuran la definición de dicho término. En estos 4 casos, el patrón estructural que siguen las predicaciones es de tipo primario.

Por su parte, en el caso de (5), el verbo *define* configura dos argumentos: uno externo, que representa al autor de una definición (*Carlos Godino*), y otro interno, que equivale al término a definir (*arquitectura naval*). (5) también introduce una frase PRED'', la cual se liga al argumento interno, dando lugar a un patrón del tipo predicación secundaria. Ambos patrones, primario y secundario, son entonces explicables a partir de la propuesta de Bowers.

4 Relaciones entre definiciones y frases predicativas

Con base en la estructuración gramatical que siguen las PRED'', puede hacerse tangible una relación entre predicaciones y definiciones. Tal relación presenta los siguientes rasgos:

- i. Existen dos tipos de PRED'': primarias, cuyo patrón estructural es sujeto + predicado; y secundarias, cuyo patrón estructura es sujeto+ objeto + predicado.
- ii. Respecto a las PRED'' primarias, éstas ligan a un término con su definición del siguiente modo: sujeto = término; predicado = definición.
- iii. Respecto a las PRED'' secundarias, éstas ligan al término con su definición del siguiente modo: sujeto = autor de la definición (explícito o implícito); objeto=término; predicado=definición.
- iv. Tanto para las PRED'' primarias y secundarias, el género próximo de la definición es introducido vía una frase nominal acompañada, opcionalmente, por alguna frase prepositiva, frase adjetiva o frase adverbial.
- v. Por su parte, la diferencia específica se vinculada, en algunos casos, a través de una oración relativa o complementaria; mientras que en otros casos lo hace a partir de una frase prepositiva, una frase adverbial o una frase adjetiva.

Tomando en cuenta estos rasgos, es posible distinguir patrones sintácticos que configuran predicaciones ligadas a definiciones. Dichos patrones son proyectados por los siguientes verbos, los cuales son expuestos en la Tabla 1.

Lo que se muestra dicha tabla son aquellos patrones sintácticos que, de acuerdo con Sierra *et al.* (2008), se derivan de verbos que operan como ligas entre términos con definiciones conforme a la tipología del apartado 2.

Definición	Verbo	Adverbio o preposición	Predicación
Analítica (género próximo/ diferencia específica)	Referir Representar Ser Significar	<i>A</i>	<i>Primaria</i>
	Caracterizar Comprender Concebir Conocer Considerar Definir Describir Entender Identificar Visualizar	<i>Como</i> <i>Por</i>	<i>Secundaria</i>
Sinonímica (género próximo)	Denominar Equivaler Llamar Nombrar Ser	<i>También</i> <i>A</i> <i>Igual a</i> <i>Similar a</i>	<i>Primaria</i>
Funcional (diferencia específica)	Emplear(se) Encargar Funcionar Ocupar Permitir Servir Usar Utilizar	<i>De</i> <i>Para</i>	<i>Primaria</i>
Extensional (diferencia específica)	Componer Comprender Consistir Constar Contar Constituir Contener Incluir Integrar Es/son parte Es / son + : (dos puntos)	<i>De</i> <i>Por</i> <i>Con</i>	<i>Primaria</i>

Tabla 1: Tipos de definiciones asociadas a predicaciones primarias y secundarias

5 Propuesta de análisis

Una vez establecidos los patrones anteriores, la siguiente fase fue evaluar su frecuencia de uso en textos. Para lograr esto, se desarrolla la siguiente metodología:

- i. Selección y análisis de dos corpus lingüísticos de carácter técnico para la obtención de datos.
- ii. Empleo de una herramienta automática de búsqueda para identificar los patrones considerados.
- iii. Implementación de técnicas estadísticas para establecer, por una parte, las frecuencias de asociación

entre predicaciones y definiciones; así como determinar, por otra parte, hasta qué grado puede reconocerse un tipo de definición dada una predicación.

5.1 Corpora empleados

Se utilizaron dos corpus técnicos: el primero fue el Corpus Lingüístico de Ingeniería (CLI), conformado por documentos en formato plano (.TXT), orientados hacia distintas áreas de ingeniería, el cual cuenta con alrededor de 500,000 palabras (Medina *et al* 2004).

El segundo fue el Corpus de Informática en Español (CIE), cuyo contenido va dirigido hacia las áreas de informática y ciencias de la computación (L'Homme y Drouin 2006). Contiene cerca de 500,000 palabras y es divisible en 4 sub-corpus:

Sub-Corpus	Abreviatura
PCWLAF (Revista <i>PC Word Latinoamérica</i>)	Corpus UM01
Guía_Computación (Revista <i>Guía Computación</i>)	Corpus UM02
<i>Windows TI Magazine</i>	Corpus UM03
<i>Wikipedia Español</i> (entradas obtenidas de Wikipedia)	Corpus UM04

Tabla 2: Descripción de los sub-corpus que conforman el CIE

5.2 Herramientas de búsqueda

Para realizar esta tarea, se usó una herramienta diseñada en Python, la cual opera con una lista de patrones que son buscados en todo un texto. El resultado es una tabla en XML que presenta el patrón buscado, la muestra obtenida, así como la frecuencia del patrón respecto al total de líneas del archivo. Un ejemplo de esto es:

Ref.	Patrón	Texto
1	coseche	Un conjunto de equipo eléctrico utilizado para un fin determinado se le coseche con el nombre de SUBESTACIÓN ELÉCTRICA.
2	coseche	Interrupción de circuitos en oposición de fases inicialmente, cuando las centrales operaban aisladas, el voltaje que se presentaba entre sus contactos después de una interrupción es el que se coseche como voltaje de restablecimiento, y los interruptores se construyen de tal manera que el arco no se presentaba nuevamente entre los contactos, ya que la separación no lo permitía.
3	coseche	Reenganche rápido Se coseche como reenganche rápido a la operación de cierre de un interruptor después de una falla.
4	coseche	El lapso que permanece el interruptor abierto después de una falla se coseche como tiempo muerto y siempre es recomendable que su duración sea corta.

Figura 3: Ejemplo de extracción

Para delimitar qué clase de predicaciones tienen una mayor probabilidad de ligarse a una definición, se tomaron en cuenta los siguientes rasgos de flexión verbal:

- Tercera persona del singular y plural.
- Participios y gerundios, los cuales estarían asociados a pretéritos perfectos o a formas en pasiva.
- Infinitivos ligados a perífrasis verbales con el auxiliar poder.

5.3 Frecuencias de asociación

Usando esta herramienta de búsqueda, se hizo una exploración tanto a los corpus CLI y CIE. En el caso del CLI, de un total de 12,532 líneas (esto es, oraciones o secuencias de oraciones que inician y terminen entre puntos), se obtuvieron 3,261 candidatos a CDs que contenían el patrón verbal considerado.

Posteriormente, se hizo una revisión manual para determinar cuáles candidatos tenían definiciones acordes con los rasgos planteados en la tipología propuesta. Así, se detectaron 307 CDs con buenas definiciones. La distribución de los patrones es la siguiente:

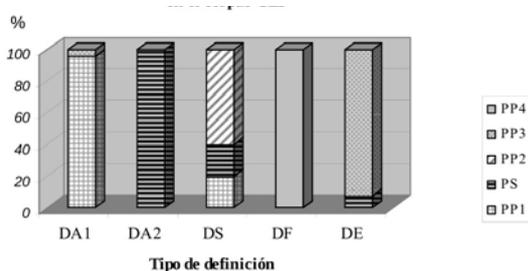


Figura 4: Frecuencias de asociación entre predicaciones verbales con tipos de definiciones en el CLI

En esta gráfica, se indica la frecuencia de asociación entre tipos de definiciones con predicaciones primarias (PP) y secundarias (PS). Como se puede notar, las PP tienen a relacionarse con los 4 tipos de definiciones considerados: analíticas (DA1), sinonímicas (DS), funcionales (DF) y extensionales (DE). En el caso de las PS, éstas se vinculan sobre todo con definiciones analíticas, y en algunas ocasiones con sinonímicas y extensionales.

Con relación al CIE, se seleccionó uno de sus sub-corpus (UM04) el cual cuenta con términos y definiciones tomados de Wikipedia.

Tras seleccionar este sub-corpus, se localizaron y extrajeron 342 líneas, en las cuales se observó la siguiente distribución entre predicaciones y definiciones:

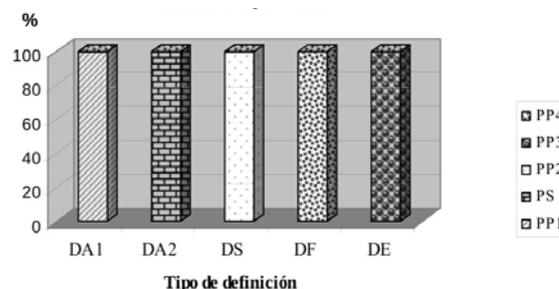


Figura 5: Frecuencias de asociación entre predicaciones verbales con tipos de definiciones en el CIE

Partiendo de las frecuencias reconocidas en el CLI y en el CIE, se puede notar que existe un alto grado de regularidad respecto a la asociación entre definiciones y predicaciones conforme a los patrones delineados en la tipología aquí propuesta, la cual puede verse en la tabla 1.

Comparando la distribución de frecuencias absolutas entre el CLI y el CIE, es posible ver cuán regular es el vínculo que establecen predicaciones con definiciones:

Predicación	Tipo de definición					Total
	Análítica 1	Análítica 2	Sinonímica	Funcional	Extensional	
PP1	111	0	8	0	0	119
PS	0	66	8	0	2	76
PP2	0	0	24	0	0	24
PP3	5	0	0	0	0	31
PP4	0	0	0	57	0	57
Total	116	66	40	57	28	307
Sub-corpus UM04						
PP1	127	0	0	0	0	127
PS	0	11	0	0	0	11
PP2	0	0	25	0	0	25
PP3	0	0	0	0	35	35
PP4	0	0	0	144	0	144
Total	127	11	25	144	35	342

Tabla 3: Distribución de frecuencias absolutas respecto a la asociación entre predicaciones y definiciones en el CLI y en el sub-corpus UM04

Conforme a la tabla, puede verse la distribución porcentual respecto a la clase de predicación asociada a un tipo de definición. Un hecho visible en esta distribución es la concentración porcentual que presenta un tipo de definición asociada a una predicación verbal específica. Así, se observa que las definiciones analíticas se ligan principalmente a predicaciones primarias: para el caso del

conjunto de CDs del CLI, éstas tienen una representación porcentual del 36%; mientras que para el caso del sub-corpus UM04 se corresponden con un 37%.

Por otro lado, la relación que se da entre definiciones funcionales y predicaciones primarias: si bien el CLI arroja un indicador porcentual del 18%, en el caso del sub-corpus UM04 esta relación está representada con un porcentaje mayor de casos (42%). Esto permite identificar un uso preferente de definiciones funcionales en el sub-corpus UM04.

5.4 Probabilidades condicionales

Con base en los datos obtenidos, se observa una relación estrecha entre predicaciones verbales y definiciones, de modo que la presencia de una predicación condiona la aparición de un tipo de definición.

Partiendo de este vínculo resulta viable determinar el grado de dependencia entre predicaciones y definiciones a partir de una prueba del tipo χ^2 , con miras a resolver si la relación entre ambos factores (predicaciones y definiciones) es un hecho fortuito o no.

La aplicación de esta prueba toma en cuenta dos hipótesis:

- **Hipótesis de investigación:** la asociación entre los factores predicación verbal y tipo de definición es dependiente entre sí. Un indicio de tal dependencia lo aporta la probabilidad de que si ocurre una predicación en un CD, es altamente probable de que esté asociada a una definición.
- **Hipótesis nula:** la relación entre los factores predicación y tipo de definición es independiente de la presencia de alguno de estos dos, por lo que se trata de un hecho fortuito en un CD.

Retomando los datos de la tabla 3, es posible plantear la siguiente fórmula:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Fórmula 1: cálculo de Chi cuadrada

En donde O_{ij} denota las frecuencias identificadas conforme a los casos indicados

por las filas (predicaciones) y las columnas (definiciones).

En contraste, E_{ij} denota las frecuencias esperadas, en otras palabras, el número de casos hipotéticos correspondientes a cada fila y columna, en el caso de que las variables predicación y definición fuesen independientes.

Los resultados que arroja la prueba de χ^2 permiten deducir valores altos para el estadístico de prueba: 977.22 en el caso del CLI, mientras que para el caso del sub-corpus UM04 se tiene un valor de 1368, con 16 grados de libertad para ambos corpus.

Estos resultados dan pie a un cálculo de probabilidades condicionales, con miras a delimitar qué tan viable es reconocer un tipo de definición específica dada la ocurrencia de una predicación. Tal probabilidad es representada en la siguiente fórmula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Fórmula 2: cálculo de probabilidades condicionales

En la Fórmula 2, B representa una predicación verbal, mientras que A refiere a un tipo de definición. Lo que se intenta deducir es, dada una predicación B, cuál es la probabilidad de que ésta se asocie a un tipo de definición A.

A continuación se presentan el total de casos de predicaciones asociadas a definiciones, considerando como factores condicionales el verbo que rige a la predicación, junto con el tipo de definición. Respecto al CLI, se observa lo siguiente:

Predicaciones y verbos	Definiciones					
	Analítica 1	Analítica 2	Sinonímica	Funcional	Extensional	Total
Predicación primaria						
Ser	93%	0	7%	0	0	100%
Predicación secundaria						
Caracterizar	0	73%	9%	0	18%	100%
Concebir	0	100%	0	0	0	100%
Conocer	0	69%	31%	0	0	100%
Considerar	0	93%	7%	0	0	100%
Definir	0	95%	5%	0	0	100%
Describir	0	100%	0	0	0	100%
Entender	0	100%	0	0	0	100%
Identificar	0	100%	0	0	0	100%
Visualizar	0	100%	0	0	0	100%
Predicación primaria						
Denominar	0	0	100%	0	0	100%
Equivaler (a)	0	0	100%	0	0	100%
Llamar	0	0	100%	0	0	100%
Nombrar	0	0	100%	0	0	100%
Predicación primaria						
Componer	0	0	0	0	100%	100%
Consistir	0	0	0	0	100%	100%
Constar	0	0	0	0	100%	100%
Contar	0	0	0	0	100%	100%
Constituir	100%	0	0	0	0	100%
Incluir	0	0	0	0	100%	100%
Ser (:)	0	0	0	0	100%	100%
Predicación primaria						
Determinar	0	0	0	100%	0	100%
Emplear(se)	0	0	0	100%	0	100%
Encargar(se)	0	0	0	100%	0	100%
Permitir	0	0	0	100%	0	100%
Servir	0	0	0	100%	0	100%
Usar	0	0	0	100%	0	100%
Utilizar	0	0	0	100%	0	100%

Tabla 4: Porcentaje de probabilidades condicionales entre predicaciones y definiciones en el CLI

De acuerdo con la tabla 4, se observan los porcentajes de asociación entre predicaciones y definiciones conforme a su probabilidad condicional. Dependiendo del verbo que opere como núcleo, ésta puede aparecer ligada a uno o varios tipos de definición con un porcentaje determinado.

En relación con las definiciones analíticas vinculadas con predicaciones primarias (Analítica 1) o secundarias (Analítica 2), los verbos *concebir*, *describir*, *entender*, *identificar* y *visualizar* muestran una probabilidad del 100% para ligarse a una definición de este tipo.

En contraparte, *ser* (93%), *caracterizar* (73%), *conocer* (69%), *considerar* (93%) y *definir* (95%), si bien tienden a mantener ligas principalmente con definiciones analíticas, también pueden estar relacionados con definiciones sinonímicas o extensionales. Los verbos que tienen nexos con definiciones sinonímicas son *ser* (7%), *caracterizar* (9%), *conocer* (31%), *considerar* (7%) y *definir* (5%).

En cuanto a las definiciones extensionales, *caracterizar* es el único verbo que se vincula con este tipo de definiciones (18%). Para los demás casos, se observa una probabilidad del 100% que se dé una asociación entre predicaciones primarias con algún otro tipo de definición.

Para el caso del sub-corpus UM04, los porcentajes obtenidos son:

Predicaciones y verbos	Definiciones					
	Analítica 1	Analítica 2	Sinonímica	Funcional	Extensional	Total
Predicación primaria						
Referir (a)	100%	0	0	0	0	100%
Representar	100%	0	0	0	0	100%
Significar	100%	0	0	0	0	100%
Ser	100%	0	0	0	0	100%
Predicación secundaria						
Caracterizar	0	100%	0	0	0	100%
Considerar	0	100%	0	0	0	100%
Definir	0	100%	0	0	0	100%
Describir	0	100%	0	0	0	100%
Entender	0	100%	0	0	0	100%
Predicación primaria						
Llamar	0	0	100%	0	0	100%
Ser (similar a)	0	0	100%	0	0	100%
Predicación primaria						
Componer	0	0	0	0	100%	100%
Contar	0	0	0	0	100%	100%
Contener	0	0	0	0	100%	100%
Constituir	0	0	0	0	100%	100%
Incluir	0	0	0	0	100%	100%
Integrar	0	0	0	0	100%	100%
Ser (:)	0	0	0	0	100%	100%
Predicación primaria						
Emplear(se)	0	0	0	100%	0	100%
Encargar(se)	0	0	0	100%	0	100%
Funcionar	0	0	0	100%	0	100%
Ocupar	0	0	0	100%	0	100%
Permitir	0	0	0	100%	0	100%
Servir	0	0	0	100%	0	100%
Utilizar	0	0	0	100%	0	100%

Tabla 5: Porcentaje de probabilidades condicionales entre predicaciones y definiciones en el sub-corpus UM04

Como se puede notar, UM04 presenta un porcentaje de probabilidad del 100% en todos los casos donde estén asociadas una predicación con una definición. En contraste con el CLI, en este sub-corpus no se observan variaciones entre el tipo de definición que selecciona una predicación.

Los datos que arrojan tanto la aplicación de la prueba de χ^2 , así como la delimitación de probabilidades condicionales, configuran un soporte empírico para argumentar a favor de la hipótesis de que una predicación mantiene un vínculo con una definición, de modo que la predicación influye significativamente en la elección del tipo de definición que puede ser introducida en un CD.

6 Consideraciones finales

En este trabajo se ha planteado un método semiautomático de identificación de definiciones en textos, tomando en cuenta la relación sintáctica que establecen dichas definiciones con patrones asociados a predicaciones verbales. Para deducir estos patrones, se hizo uso de la llamada teoría de la predicación.

El uso de esta teoría permitió distinguir dos tipos de configuración sintáctica que asumían estos patrones: predicaciones primarias y secundarias. Como se ha mostrado, estas dos clases de predicaciones influyen en la selección del tipo de definición.

Para evaluar hasta qué grado una predicación afecta la selección de un tipo de definición, se realizó un experimento de reconocimiento de predicaciones y definiciones en dos corpus: el CLI y el CIE. Dicho experimento consideró un análisis estadístico para determinar la frecuencia con la que una predicación aparece vinculada con una definición.

Los resultados arrojados ofrecen una base empírica a favor de la relación entre predicaciones y definiciones en CDs, haciendo viable el empleo de predicaciones para identificar, a través de un sistema de búsqueda automática, distintos tipos de definiciones en grandes repositorios de documentos especializados.

Este trabajo fue realizado gracias al apoyo del CONACYT y de la DGAPA-UNAM.

Bibliografía

- Alarcón, R. 2006. Extracción automática de contextos definitorios en textos especializados. Protocolo de Investigación Doctoral, Barcelona, Universidad Pompeu Fabra.
- Alshawi, H. 1987. Processing Dictionary Definitions with Phrasal Pattern Hierarchies, *Computational Linguistics*, 13(3-4): 195-202.
- Auger, A. 1997. Repérage des énonces d'intérêt définitoire dans les bases de données textuelles, Thèse de Doctorat, Neuchâtel, Suisse, Université de Neuchâtel.
- Bowers, J. 1993. The Syntax of Predication. *Linguistic Inquiry*, 24(4): 591-636.
- Bowers, J. 2001. Predication. En Baltin, M. y Collins, C. (eds.), *The Handbook of Contemporary Syntactic Theory*, Oxford, Blackwell: 299-333.
- L'Homme, M. C. y Drouin, P. 2006. *Corpus de Informática en español*, OLST, Université de Montreal: www.olst.umontreal.ca.
- Malaisé, V. Zweigenbaum, P. y Bachimont, B. 2005. Mining defining contexts to help structuring differential ontologies. *Terminology*, 11(1): 21-53.
- Mallén, E. 1991. A Syntactic Analysis of Secondary Predication in Spanish. *Journal of Linguistics*, 27, 375-403.
- Medina, A. Sierra, G. Garduño, G. Méndez, C. y Saldaña, R. 2004. CLI: An Open Linguistic Corpus for Engineering. En *Proceedings of IX Ibero-American Workshop on Artificial Intelligence*, páginas 203-208, BUAP (Puebla, México).
- Meyer, I. 2001. Extracting a knowledge-rich contexts for terminography: A conceptual and methodological frame-work. En Bourigault D., Jaquemin C., y L'Homme M.C. (eds.), *Recent Advances in Computational Terminology*, John Benjamins, Amsterdam/Philadelphia: 279-302.
- Rebeyrolle, J. 2000. Forme et fonction de la définition en discours, Thèse de Doctorat, Toulouse, France, Université Toulouse-Le Mirail.
- Pearson, J. 2004. Terms in Contexts. John Benjamins, Amsterdam.
- Sierra, G. Alarcón, R. y Aguilar, C. 2006. Extracción automática de contextos definitorios en textos especializados. En Inchaurrede, C. y Ibarretxe, I. (eds.), *Memorias del XXII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, páginas 351-352, Universidad de Zaragoza (Zaragoza, España).
- Sierra, G. Alarcón, R. Medina, A. y Aguilar, C. 2003. Definitional Contexts Extraction from Specialised Texts. En *PALC 2003 Proceedings: Language, Corpora and E-Learning*, Frankfurt, Peter Lang Publish: 21-31.
- Sierra, G. Alarcón, R. Aguilar, C. y Bach, C. 2008. Definitional Verbal Patterns for Semantic Relation Extraction, *Terminology*, 14(1): 74-98.