CORPUS DE CONTEXTOS DEFINITORIOS: UNA HERRAMIENTA PARA LA LEXICOGRAFÍA Y LA TERMINOLOGÍA

Gerardo Sierra, Rodrigo Alarcón, César Aguilar, Alberto Barrón, Valeria Benítez e Itzia Baca

{gsierram,ralarconm,caguilar,lbarronc,vbenitezr,ibacai}@ii.unam.mx
Instituto de Ingeniería, UNAM
Torre de Ingeniería, Basamento
Ciudad Universitaria
México 04510, D.F.

RESUMEN

Un corpus de contextos definitorios (CCDs), más allá de ser concebido como un mero repositorio de documentos, es una herramienta valiosa para la terminología y la lexicografía, ya que puede facilitar el proceso de extracción de unidades tales como términos y definiciones. Así, la propuesta que aquí se presenta expone el diseño y desarrollo del CCDs que contiene tales unidades, las cuales han sido obtenidas de corpus de textos técnicos de diversas áreas temáticas. Del mismo modo, se explica la metodología empleada y el desarrollo de las herramientas y aplicaciones previstas para este corpus.

1 Introducción

Los corpus textuales son ampliamente utilizados por la terminología, la lexicografía y algunas otras áreas de investigación, en cuestiones tales como frecuencia y conteo de palabras, análisis de concordancias y otras labores similares que son útiles para distinguir, por ejemplo, las diferencias del sentido de una palabra. Con todo, en algunos casos el tratamiento de dichos corpus no resulta suficiente para ubicar la relación que se da entre términos y definiciones dentro de un texto especializado, de modo que pueda clarificarse el significado de tal término en un determinado contexto.

En años recientes, se han formulado diversas investigaciones orientadas al reconocimiento y clasificación de términos y definiciones —de manera manual o automática—, entre las cuales se pueden mencionar la búsqueda sistemática de contextos ricos en conocimiento (ing. *Knowledge-Rich Contexts*) por parte de Pearson [1998] y Meyer [2001], el trabajo sobre la extracción de información metalingüística por parte de Rodríguez [2004], así como el uso de enunciados definitorios (fr. *énoncés définitoires*) realizadas por Auger [1997] y Rebeyrolle [2000].

Siguiendo esta línea de trabajo, el Grupo de Ingeniería Lingüística de la UNAM (GIL) ha formulado la noción de *contexto definitorio* (CD), la cual servirá como base para la edificación de un *Corpus de Contextos Definitorios* (CCDs). Dicho corpus es una herramienta valiosa para la terminología, la lexicografía, la extracción de información y la minería de textos, ya que apoya a la construcción de diccionarios electrónicos onomasiológicos y semasiológicos, la elaboración de bancos terminológicos, el diseño de ontologías, además de agilizar la búsqueda automática de términos y definiciones, por sólo mencionar algunas aplicaciones.

Este trabajo presenta la siguiente distribución: en primer lugar se definirá la noción de CD y se describirán los elementos que lo conforman. En segundo lugar, se explicará la metodología empleada para la extracción de CDs del Corpus Técnico del IULA.

Finalmente, se presentarán las etiquetas y herramientas que se han aplicado para el diseño y construcción del CCDs.

2 Contextos definitorios

Como se sabe, dentro de la terminología actual, una de las tareas que ha tenido gran éxito es el reconocimiento y extracción automática de términos [Cabré *et al.* 2001]. Empero, en el caso de la extracción de definiciones nos enfrentamos a un asunto complejo que tiene que ver con las diferentes posturas acerca de los límites y características de una definición. Por un lado existe una tendencia normativa reflejada, por ejemplo, en los diccionarios de lengua, mientras en otra postura se retoman aspectos comunicativos y cognitivos subyacentes a la configuración de conceptos [Lara 2004, Sager y Ndi-Kimbi 1995, Vossen y Copestake 1993].

Por lo anterior, el análisis lingüístico de una definición no es un asunto trivial. Una manera de destacar la riqueza de relaciones que puede mantener una definición respecto a un término, es considerar que ambos se configuran en una estructura llamada *contexto definitorio* (CD), entendido éste como un fragmento textual donde se introduce un *término* y su correspondiente *definición* [Alarcón y Sierra 2003]. Este trabajo se enfoca en la presencia de patrones lingüísticos y no-lingüísticos constitutivos de CDs, así como en el uso de dichos patrones para identificar y etiquetar de manera automática términos y definiciones en textos especializados.

2.1 Elementos constitutivos de los CDs

Un CD presenta estructuras diversas, sin embargo, se ha podido observar la presencia recurrente de elementos como: término, definición, predicación verbal definitoria, marcadores reformulativos definitorios, marcadores tipográficos definitorios, patrones pragmáticos y relaciones de correferencias y anáforas que se establecen entre el término definido y estas unidades enlistadas. La identificación de dichos componentes permite distinguir patrones de comportamiento sintáctico y semántico que facilitan la obtención de datos para la búsqueda y clasificación de las definiciones.

Los dos componentes mínimos y básicos de un CD son el término y la definición del mismo. Se ha observado que ambos elementos establecen ligas a partir de predicaciones verbales definitorias (PVD) del tipo *X* ***se define como*** *Y, X* ***comprende*** *Y, X* ***es*** *Y*, entre otras. Las PVDs mantienen una estructura reconocible en patrones determinados, llamados patrones definitorios (PD), que permiten establecer un catálogo de verbos asociados a tipos de definiciones específicas. La búsqueda y establecimiento de estos patrones verbales ha sido fundamental en el análisis de CDs.

Si bien estos tres elementos configuran una estructura canónica de un CD, no siempre ocurre así, ya que la mayoría de las veces los CDs presentan otros PDs que también ligan a términos y definiciones, los cuales son:

- Marcadores reformulativos definitorios (MRDs): estructuras sintácticas relacionadas con un proceso metalingüístico. Estas construcciones ayudan a reinterpretar o retomar algún elemento discursivo para presentarlo de otra forma, p. e.: *es decir, por ejemplo, esto es*, etc.
- Marcadores tipográficos definitorios (MTDs): cualquier signo de puntuación o marca tipográfica que tenga como función, por un lado, ligar a un término con su definición, sustituyendo o complementando la función de la PVD (viñetas,

paréntesis, guiones, comillas, etc.); y por otro lado, aquellas marcas que recalquen la presencia tanto de un término como de su definición (negritas, subrayado, cursivas, comillas, etc.).

- Patrones pragmáticos (PPs): proporcionan información sobre el uso de los términos. Se consideran tres variantes: (a) autoría, cuando es un autor o una institución quien define el termino; (b) patrones pragmáticos temporales, en el caso en que se proporcione alguna fecha en la que se definió o se introdujo el término; y (c) patrones instruccionales, referidos a estructuras sintácticas que dan matices diferentes para introducir la definición: *desde el punto de vista, de manera general, desde la perspectiva, científicamente,* etc.

Finalmente se plantea la presencia de correferencias y anáforas que, aunque no son componentes de un CD, implican relaciones referenciales a partir de las cuales se puede deducir si un término está ligado a otros términos o entidades lingüísticas alejadas del CD. Correferencias y anáforas tienen un contenido conceptual valioso para determinar el sentido del término definido. A continuación vemos un ejemplo con algunos de estos componentes:

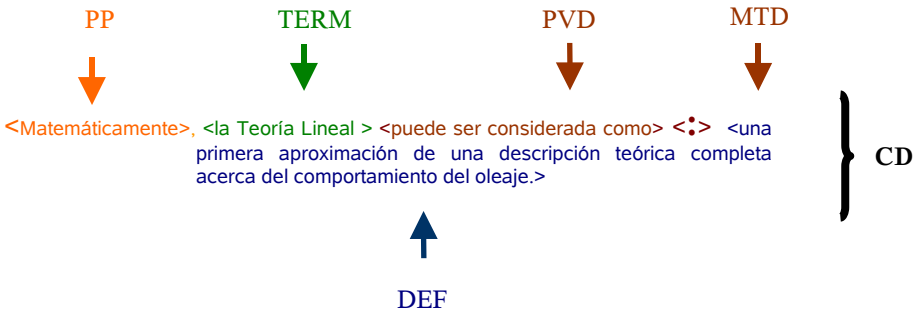

Figura 1: Ejemplo de un CD

2.2 Tipos de definiciones y CDs

Para los fines de este trabajo, se concibe una definición como la descripción lingüística de un concepto representada por un término. Dicha descripción establece relaciones con otros términos para delimitar el significado de un concepto. A partir de la figura 2 se muestran 5 tipos de definiciones:

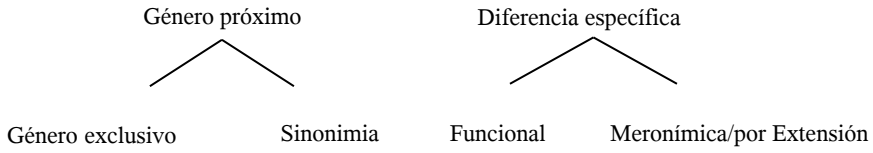

Figura 2: Tipología de definiciones a partir de un modelo Género próximo y Diferencia específica

- Definición analítica: señala el Género próximo y la Diferencia específica.
- Género exclusivo: no provee una Diferencia específica.
- Definición sinonímica: indica una fuerte relación semántica con el Género próximo.
- Definición funcional: incluye la Diferencia específica que indica la función del concepto.
- Definición meronímica/por extensión: incluye la Diferencia específica que enumera las partes que componen el concepto.

A partir de esto, es posible identificar patrones específicos y estructuras asociadas con la representación de conceptos en lenguaje natural. Con base en trabajos anteriores [Sierra *et al* 2003], se ha visto que las definiciones con un patrón canónico se asocian a estructuras lingüísticas determinadas, como ocurre con las de tipo analítico o aristotélico que presentan rasgos lingüísticos particulares:

- La mayoría de las veces inician con frases nominales, las cuales pueden contener un cuantificador (*todos, algunos, cada uno, ninguno*), determinantes (*un, una, unos, unas, el, la, los, las*), o demostrativos (*este, esta, estos, estas*).
- El Género próximo puede configurarse a partir de un conjunto de frases adjetivas o prepositivas que siguen a la frase nominal inicial.
- La Diferencia específica se inserta a través de oraciones subordinadas compuestas por un sustantivo, un adjetivo o frases preposicionales.

En general, existe una estrecha relación entre el tipo de definición y la estructura verbal dentro del CD. Se ha observado que una PVD determina el tipo de definición que se introducirá en un CD [Aguilar *et al* 2004]. De este modo, la relación recurrente que se establece entre cierta clase de definiciones y las estructuras lingüísticas presentes en el CD, permite establecer patrones que facilitan el reconocimiento y clasificación de CDs dentro del corpus.

3 Metodología para la identificación de CDs

En este rubro presentamos la metodología que empleamos para la extracción de CDs. Se utilizaron principalmente dos fuentes:

- Corpus Lingüístico de Ingeniería (CLI) [Medina *et al*. 2004] desarrollado por el Grupo de Ingeniería Lingüística (GIL). Dicho corpus está conformado por documentos tipo TXT con un total aproximado de 500,000 palabras. El CLI es una recopilación integrada por documentos técnicos (artículos, informes, tesis, etc.) en diversas áreas temáticas de ingeniería.
- Corpus Técnico del IULA (http://bwananet.iula.upf.edu/indexes.htm) [Cabré y Vivaldi 1997], desarrollado por el Instituto Universitario de Lingüística Aplicada (IULA), al cual puede accederse a través del motor de búsqueda *BwanaNet* que permite hacer consultas terminológicas.

Para ambos casos se emplearon patrones lingüísticos en la extracción de CDs. En este trabajo se describe únicamente el proceso que se siguió para el Corpus Técnico del IULA, ya que en una investigación previa se abordó la metodología para la extracción de patrones del CLI [Aguilar *et al*. 2004].

A continuación se describe brevemente el proceso empleado para la extracción de CDs y la herramienta usada para su ubicación. Como parte inicial de dicho proceso se identificaron los verbos que están ligados a definiciones del tipo Género próximo y Diferencia específica, de los cuales encontramos *comprender, concebir, conocer, considerar, definir, denominar, entender* e *identificar*. La idea de esto fue extraer de manera automática posibles candidatos a CDs, y construir un repositorio que permitió establecer una clasificación de CDs, así como de los patrones que los conforman.

3.1 Extracción de CDs

El primer paso consistió en rastrear las ocurrencias de lemas de verbos definitorios. La búsqueda se limitó a 500 ocurrencias de cada lema para obtener los contextos al azar. Una vez recuperados los contextos que contenían verbos definitorios, se analizaron manualmente para distinguir candidatos a CDs. Posteriormente se seleccionó y clasificó cada CD según la flexión verbal, lo que permitió identificar patrones verbales recurrentes en las definiciones.

3.2 Patrones verbales definitorios

Empleando esta serie de patrones se emprendió una nueva búsqueda en la que éstos fueron codificados en expresiones regulares, de acuerdo con los criterios formulados por las etiquetas EAGLES usadas en el Corpus Técnico del IULA. Dichos patrones se componen de un verbo definitorio (VD) y de elementos gramaticales tales como el adverbio *como* y el clítico *se*.

Encontramos que los lemas de los VDs, así como los patrones sintácticos constituidos a partir de éstos, tienen estructuras similares. Sin embargo, algunos patrones presentan particularidades dependiendo del verbo. A continuación mostramos un patrón posible y ejemplos de estructuras de PVDs localizadas con *BwanaNet* usando los verbos *definir, entender* y *denominar*:

- [word="se"] [pos="R.*"]{0,1} [lemma="definir|entender|denominar" & pos="V[^IGC]...."] [word!="como"]{0,15} [word="como"]

El patrón verbal está formado por el clítico *se*, señalado por la etiqueta "R", el lema de los VDs con su etiqueta "V", cualquier palabra opcional que aparezca en un rango de quince palabras hasta antes del adverbio *como*, y finalmente este adverbio. Con dicho patrón se recuperaron contextos como el siguiente, en donde vemos que el término *hemólisis* está dentro de la distancia opcional:

- Estudio de la hemólisis: <search pattern>*se define* **la hemólisis** *como*</search pattern> la reducción de la duración de la vida de los hematíes, que normalmente es de 110 + 10 días.

3.3 Resultados

Se muestran a continuación algunos ejemplos de PDs buscados con BwanaNet:

- ***Definir***:
  [pos="Z"] [lemma="definir" & pos="HMS"] [word!="como"]{0,15}[word="como"]
- ***Entender***:
  [pos="V.*"]{1,2} [lemma="entender" & pos="HMS"] [word!="como"]{0,15} [word="como"]
- ***Denominar***:
  [pos="N.*"] [pos="J.*"]{0,2} [lemma="denominar" & pos="HMS"]

En el primer ejemplo, la etiqueta "Z" representa cualquier signo de puntuación. En el segundo, consideramos la ocurrencia de un verbo auxiliar. Finalmente, en el tercero las etiquetas "N" y "J" representan sustantivo y adjetivo respectivamente, que son elementos comunes en términos especializados.

Empleando estos patrones y la metodología descrita arriba, recuperamos un total de 10.589 contextos. En la primera etapa se extrajeron lemas del VD, y se encontraron 4.352 contextos, de los cuales 363 eran buenos CDs, mientras que 3.989 no lo fueron. En la segunda etapa de búsqueda con los patrones verbales, encontramos 3.095 buenos CDs y 3.142 malos. Hay una diferencia substancial entre el número de CDs que se obtuvieron buscando lemas de verbos definitorios, y aquellos que se encontraron por medio de la detección de PDs. Con éste último recurso podemos obtener mejores resultados en *Precision* y *Recall*.

Por último, medimos *Precision* en cada PD. Esta medida, utilizada en la recuperación y extracción de información, se obtuvo dividiendo el número encontrado de CDs entre el número total de contextos recuperados. Entre más cercano a 1 es el valor de *Precision,* el resultado es mejor*,* es decir, existe menos ruido en la recuperación de buenos candidatos a CDs. Así pues, se observó que los valores de *Precision* de cada verbo en las definiciones del tipo Género próximo y Diferencia específica se pueden ordenar de manera descendiente: *denominar*, 0.720; *conocer*, 0.5183; *concebir*, 0.5116; *entender*, 0.4559; *definir*, 0.4483; *identificar*, 0.2133; *comprender*, 0.1580; y *considerar*, 0.1267**.**

4. Etiquetas para el CCDs

Después de la extracción y clasificación de los CDs, se diseñó un sistema de etiquetas en XML que facilitara la identificación de las partes de un CD. Las etiquetas delimitan tanto al CD de forma global, como los elementos particulares inmersos en él. Las etiquetas permiten configurar un archivo de XML que cuenta con cabeza y un cuerpo. Los campos declarados en la cabeza se muestran en la Tabla 1.

Tabla 1: Etiquetas de la cabeza del documento XML

ETIQUETAS	FUNCIÓN
Cabeza	Dentro de este bloque se encuentra la información del documento como nombre, de qué verbo se trata, la fuente, fecha, recopilador, etc.
Fuente	Indica el nombre del corpus que se está etiquetando. Es muy importante tener localizada la fuente de origen de los documentos.
Fecha	Fecha en la que fue recopilado y etiquetado el documento.
Nombre	Contiene el nombre de la recopilación hallada en el documento, por ejemplo, puede contener “verbo ser”.
Verbo	Cuando en el documento únicamente es analizado un verbo definitorio en cualquiera de sus predicaciones tiene que señalarse el nombre del verbo.
Tipo	Existen varios tipos de definiciones: analítica, funcional, etc. En el caso en que el criterio de clasificación del documento sea el tipo de *definición* se tendrá que indicar.
Recopilador	Nombre de la persona que recopiló el documento y lo consignó al Corpus de CD.

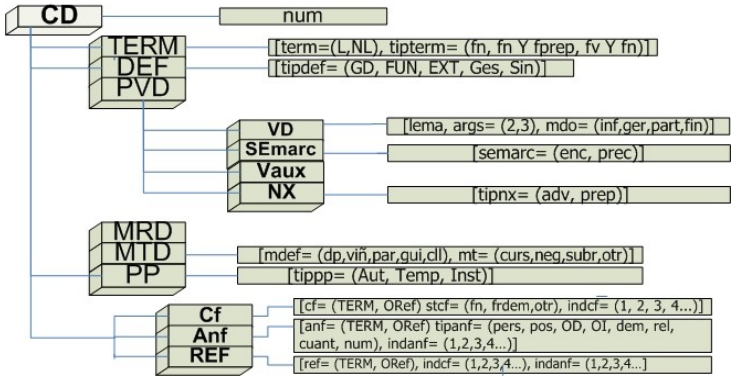

Figura 3: Árbol de etiquetas para el cuerpo de CDs XML (Con atributos)

En la Figura 3 ilustramos las etiquetas y atributos empleados para el cuerpo de los CD y su significado se explica en la Tabla 2.

Tabla 2: Etiquetas para el cuerpo de un CD en documento XML

Etiqueta	Significado	Función
CD	Contexto definitorio	Indica los elementos que constituyen al CD dentro de ellos se encuentran el término, su definición, la predicación verbal y relaciones de correferencia.
TERM	Término	En sus atributos se marca si se trata de un término lingüístico o de uno no lingüístico (cifras, símbolos). Se toman en cuenta tres tipos de frases: *fn* (frase nominal), *fn Y fprep* (frase nominal seguida de frase prepositiva) y *fv Y fn* (frase verbal seguida de frase nominal).
DEF	Definición	En ella se debe omitir cualquier texto complementario que de manera estricta no forme parte de dicha definición. Existen cinco tipos: *GD* (Género próximo/Diferencia específica), *FUN* (Funcional), *EXT* (Meronímia/Extensional), *Gex* (Género exclusivo) y *Sin* (Sinonímica) que se marcan en los atributos.
PVD	Predicación verbal definitoria	Contiene todos los componentes de una PVD: VD, clítico *se*, verbo auxiliar, verbo definitorio y nexo.
VD	Verbo definitorio	Cuenta con los atributos *lema, args* (marca los argumentos del verbo); *mod* (indica el modo verbal: infinitivo *inf,* gerundio *ger,* participio *part,* formas finitas o verbo conjugado *fin*).
SEmarc	Clítico SE	Se indica su posición respecto al verbo. El atributo distingue entre *enclítico* (*enc*) cuado *se* es parte de la morfología verbal y está en posición final y *preclítico* (*prec*) cuando el clítico está en posición preverbal.
Vaux	Verbo auxiliar	Contiene cualquier verbo auxiliar dentro de la PVD (p. e., *se* ***puede*** *considerar como*, se ***ha*** definido, se ***debe*** concebir como…)
NX	Nexo	Señala la función que cumple un adverbio o preposición entre el verbo y la definición.
MRD	Marcadores reformulativos definitorios	Abarcan estructuras sintácticas con la función de explicar el propio lenguaje. Son frases que retoman algún elemento discursivo para reintroducirlo al discurso de otra forma, p.e.: *es decir, por ejemplo, esto es*, etc.
MTD	Marcadores tipográficos definitorios	Señala cualquier signo de puntuación o *marcadores tipográficos definitorios* (MTD). Se distingue en dos tipos: 1) *marcadores definitorios* (mdef): unen a un término con su definición, sustituyendo o complementando la función de la PVD. En los atributos se señalan como ***mdef****= dp, viñ, par, gui, cll, otr* 2) *marcadores tipográficos* (mt): indicación de negritas, cursivas, subrayado y otras marcas que dan prominencia al término definido o a la definición, este caso se marca ***mt****= neg,curs,subr,otr.*
PP	Patrones pragmáticos	Dan información sobre el uso de los términos. Los tres patrones considerados en este rubro son: *Autoría* (Aut); *Patrones pragmáticos temporales* (Temp) y *Patrones instruccionales* (Inst).
Cf	Correferencia	Contiene las relaciones de referencia que se dan dentro del CD. En los atributos se marca si la *Cf* se da con el término (TERM) o con cualquier otro elemento del CD que opere como referente (ORef). Se especifica si la *Cf* es una *frase nominal* (fn), *frase nominal con demostrativo* (frdem), o tiene otra estructura (otr). A partir de números se marca el *índice* de la Cf (idcf) que permite ligarla con su referente (REF).
Anf	Anáfora	Marca las anáforas dentro del CD. En los atributos se marca si la *Anf* se da con el término (TERM) o con cualquier otro elemento del CD que opere como referente (ORef); se especifica también el tipo de

		anáfora o tipo de pronombre. Igual que en el caso anterior, el *índice* para ligar con su referente, es señalado con números.
REF	Referente	Contiene al referente (REF) o antecedente de las correferencias y a las anáforas presentes en el CD. En los atributos se señala como *índice* (indcf/indanf), si el término definido (TERM) es el referente o es cualquier otra entidad (ORef) del CD.

4.1 Herramientas para el análisis del CCDs

Después de establecer el sistema de etiquetas del CCD, se evaluó su eficacia para identificar tanto elementos textuales como otras unidades constitutivas de los CDs. Los primeros documentos fueron etiquetados de manera manual en editores de XML. Sin embargo, al tratarse de un proceso que resultaba lento, se optó por agilizarlo manipulando los textos a través de macros Microsoft Word. La ventaja de esto es la posibilidad de hacer semiautomático el proceso de etiquetado. La barra de acceso es la siguiente:

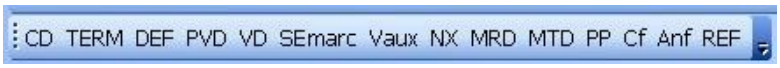

Figura 4: Barra de herramientas de macros "CONTEXTOS"

A continuación, mostramos una tabla con la salida de cada macro:

Tabla 3: Macros con salida de etiquetas “CONTEXTOS” de XML

BOTON	Macro	BOTON	Macro
1. CD	<CD num= “”> </CD>	**8.NX**	<NX tipnx= “”> </NX>
2. TERM	<TERM term= “” tipterm= “”> </TERM>	**9. MRD**	<MRD> </MRD>
3. DEF	<DEF tipdef= “”></DEF>	**10.MTD**	<MTD mdef= “” mt= “”> </MTD>
4. PVD	<PVD> </PVD>	**11. PP**	<PP tippp= “”> </PP>
5. VD	<VD lema= “” args= “” mdo= “”> </VD>	**12. Cf**	<Cf cf= "" stcf= "" indcf= ""> </Cf>
6. SEmarc	<SEmarc tipsemarc= “”></SEmarc>	**13. Anf**	<Anf anf= "" tipanf= "" indanf= ""> </Anf>
7. Vaux	<Vaux> </Vaux>	**14. REF**	<REF ref= "" indcf= "" inanf= ""> </REF>

Como se puede observar, los atributos en las etiquetas de apertura aparecen sin valor, éste debe ser asignado manualmente y en caso de no existir en el CD se deja en blanco.

Las etiquetas y las herramientas han sido modificadas durante el proceso de la investigación y son de gran utilidad para identificar los CDs a partir de los cuales se construirá el CCDs. Por último mostramos un ejemplo de un CD tal y como lo obtenemos del documento TXT y un ejemplo de las etiquetas que se le aplican:

- M. Godron y G. Merriam entre otros, quienes consideran a la Ecología del Paisaje como: "la ecología de los sistemas movibles y heterogéneos , estudiando entonces la influencia de la estructura del paisaje sobre los procesos ecológicos , tanto a escala local como regional "

- <CD num= ""> <PP tippp= "Aut"> <REF ref= "ORef" indcf= "" inanf= "1"> M. Godron y G. Merriam </REF> </PP> entre otros <Anf anf= "ORef" tipanf= "rel" indanf= "1"> quienes </Anf> <PVD><VD lema= "considerar" args= "3" mdo= "fin"> consideran</VD> <NX tipnx= "prep"> a </NX> la <TERM term= "L" tipterm= "fn"> Ecología del Paisaje</TERM> <NX tipnx= "adv"> como </NX> </PVD><MTD mdef= "dp" mt= ""> : </MTD> <DEF tipdef= "GD"><MTD mdef= "" mt= "otr">"</MTD> la ecología de los sistemas movibles y heterogéneos , estudiando entonces la influencia de la estructura del paisaje sobre los procesos ecológicos , tanto a escala local como regional <MTD mdef= "" mt= "otr">" </MTD> </DEF> </CD>

5. Trabajo a futuro

5.1 Extracción terminológica

Uno de los procesos que favorece la búsqueda y delimitación de CDs y sus componentes es la extracción automática de términos. Dicha extracción resulta de gran utilidad tanto para aplicaciones lexicográficas como para construir glosarios y diccionarios. En el caso del CCDs, el hecho de ubicar la relación que hay entre PVD y definición tiene una incidencia directa en la localización de términos asociados a definiciones en textos especializados.

5.2 Anáforas y otras relaciones de correferencia en la expansión de CDs

Existen casos en los que se pierde información dentro de un CD debido a que las relaciones de correferencia que hay entre los términos y otras unidades tales como frases nominales, nombres propios, comunes y pronombres no se han explicado a profundidad. Por ello es importante hacer un estudio que permita identificar el tipo de relaciones de correferencia y anáforas entre términos y otras unidades lingüísticas. Actualmente se lleva a cabo un análisis de este fenómeno con miras a establecer patrones recurrentes de referencias que faciliten el proceso de búsquedas automáticas.

5.3 Desarrollo de un modelo de descripción formal para definiciones

Finalmente, uno de los problemas que han abordado tanto la lexicografía y la terminología es describir cómo se estructura en un plano formal una definición, y si dicha estructura es única o muestra diferentes variaciones, de acuerdo con las necesidades que un terminólogo o un lexicógrafo tenga para plasmar un concepto. Al respecto, otra de las líneas de trabajo a futuro, es el desarrollo de un modelo lógico-formal que describa de qué manera se constituye una definición en lenguaje natural, atendiendo precisamente a las diversas formas que ésta asume al ser introducida dentro de un CD. La base de esta propuesta es ubicar algún tipo de formalismo que dé cuenta del modelo canónico Género Próximo + Diferencia específica, y a partir de esto derivar variaciones asociables a este patrón mencionado.

Agradecimientos

La realización de este trabajo ha sido posible gracias al patrocinio del Consejo Nacional de Ciencia y Tecnología (Proyecto 46832), y al co-financiamiento del Macro-Proyecto "Tecnologías para la Universidad de la Información y la Computación", dentro del marco del Programa Transdisciplinario en Investigación y Desarrollo de la Secretaría de Desarrollo Institucional, UNAM.

Referencias

Aguilar C., Alarcón R., Rodríguez C. y Sierra G. (2004) "Reconocimiento y clasificación de patrones verbales definitorios en corpus especializados". En Cabre, T., Estopà, R. y Tebé, C. (edts.), *La terminología en el siglo XXI*, IULA-UPF, Barcelona: 259-269.

Alarcón R. y Sierra G. (2003) "El rol de las predicaciones verbales en la extracción automática de conceptos". *Estudios de Lingüística Aplicada* 38: 129-144.

Auger A. (1997) *Repérage des énonces d'intérêt définitoire dans les bases de données textuelles*, Thèse de Doctorat, Neuchâtel, Suisse, Université de Neuchâtel.

Cabré T., Vivaldi G. (Coords.) (1997) *Corpus Tècnic del IULA de la UPF* (CT-IULA). http://bwananet.iula.upf.edu/indexes.htm.

Cabré T., Estopà R., Vivaldi J. (2001) "Automatic term detection. A review of current systems", en Bourigault D., Jaquemin C., L'Homme M.C. (eds.), *Recent Advances in Computational Terminology*, Amsterdam, John Benjamin Publish, 53-87.

Lara L.F. (2004): *De la definición lexicográfica*, México, COLMEX.

Medina A., Sierra G., Garduño G., Méndez C., Saldaña R. (2004) "CLI: An Open Linguistic Corpus for Engineering", en De Ita G., Fuentes O., Galindo M, (eds.), *Proceedings of IX Ibero-American Workshop on Artificial Intelligence*. Puebla, México, Autonomous University of Puebla, 203-208.

Meyer I. (2001) "Extracting a knowledge-rich contexts for terminography: A conceptual and methodological framework", en Bourigault D. Jaquemin C. L'Homme M.C. (eds.), *Recent Advances in Computational Terminology*, Amsterdam, John Benjamin Publish: 279-302.

Pearson J. (1998) *Terms in Context*, Amsterdam, John Benjamin Publish.

Rebeyrolle J. (2000) 'Utilisation des contextes définitoires pour l'acquisition de connaissances à partir des textes' en *Actes des Journées Francophones d'Ingénierie des Connaissances, IC'2000*, Toulouse, France, 105-114.

Rodríguez C. (2004) *Metalinguistic Information Extraction from specialized texts to enrich computational lexicons*, Ph. D. Dissertation, Universitat Pompeu Fabra, Barcelona.

Sager J.C. y Ndi-Kimbi A. (1995): "The conceptual structure of terminological definition and their linguistic realisations ". *Terminology*, 2(1): 61-85.

Sierra G., Alarcón R., Medina A., Aguilar C. (2003) 'Definitional Contexts Extraction from Specialised Texts' en Lewandowska-Tomaszczyk, B. (ed.), *PALC 2003 Proceedings: Language, Corpora and E-Learning*. Frankfurt, Peter Lang, 21-31.

Vossen P. y Copestake A. (1993) "Defaults in Lexical Representation" en Briscoe T., Paiva V. & Copestake A. (eds.), *Inheritance, Defaults and the Lexicon*, Cambridge, CUP: 246-274.