

FELIU, J.; SOLÉ, E.; TEBÉ, C.; CABRÉ, M. T. (2002) «Las relaciones conceptuales: un elemento esencial en la estructuración del conocimiento especializado». En: *Actas del VIII Simposio Iberoamericano de Terminología*. Cartagena de Indias, 28 a 31 de octubre de 2002. ISBN: 958-33-4022-7.

**VIII Simposio Iberoamericano de Terminología
Cartagena de Indias, octubre 2002
Sección: Terminología y Nuevas Tecnologías**

**Las relaciones conceptuales: un elemento esencial en la
estructuración del conocimiento especializado***

Judit Feliu, Elisabet Solé, Carles Tebé, M. Teresa Cabré
Grupo IULATERM
Institut Universitari de Lingüística Aplicada
Universitat Pompeu Fabra (Barcelona)

0. Introducción

En los últimos tiempos, se ha producido un cambio en la concepción de la terminología y, en consecuencia, en el estudio de las relaciones conceptuales causado por la irrupción de aportaciones teóricas desde las ciencias cognitivas, la lingüística y la semántica léxica y, desde el punto de vista aplicado, gracias a las aproximaciones desde el campo de la inteligencia artificial y los sistemas de representación y recuperación automática de la información. Estos nuevos enfoques, que amplían la noción y la caracterización de las relaciones semánticas, permiten dar cuenta de la vehiculación del conocimiento en los textos especializados y establecer cuáles son los elementos necesarios para proceder a la estructuración, y posterior recuperación selectiva, de la información especializada contenida en los textos.

El objetivo de esta comunicación es, en primer lugar, contrastar la tipología de relaciones conceptuales establecida en Feliu (2000) sobre un texto especializado del ámbito del Genoma humano y avanzar en el trabajo presentado en la edición anterior del Simposio (Feliu, Solé, Tebé, en prensa). Para contrastar esta tipología, se usará una lista preestablecida de marcadores lingüísticos que nos permitirá detectar las relaciones que efectivamente se

* Este artículo se ha llevado a cabo en el marco del proyecto financiado TEXTERM: Textos especializados y terminología: selección y recuperación automática de la información (BFF2000-0841).

establecen entre las unidades de conocimiento especializado en el ámbito tratado. En segundo lugar, y a partir del análisis de cada tipo de relación detectada, se observará qué tipo de unidad de conocimiento especializado vincula y qué otro tipo de información se establece sobre la unidad.

Finalmente, y desde el punto de vista aplicado, esbozaremos las principales características de un prototipo de sistema de representación de la información que, a partir de las relaciones detectadas en el texto mediante marcadores lingüísticos explícitos, sea capaz de proponer una etiqueta del tipo o subtipo de relación conceptual que expresa. Además, el sistema propondrá la unidad de conocimiento que aparece a ambos lados de la relación. La propuesta del sistema será una representación semiautomática de los nudos y de las principales relaciones de un texto que, evidentemente, deberá ser validada o descartada por el usuario.

1. ¿Qué es una relación conceptual?

Después de un estudio en profundidad de las relaciones semánticas y conceptuales desde la filosofía, la terminología, la lingüística y la inteligencia artificial (desde la óptica de las ontologías), consideramos que una relación conceptual se puede definir como un vínculo semántico, esto es, una predicación entre dos o más unidades de conocimiento especializado¹. Esta definición puede formalizarse de la manera siguiente:

$$a R b, n$$

donde R representa la relación; a y b son las unidades de conocimiento especializado que pertenecen a una clase semántica en particular; y, considerando que algunas relaciones permiten la recursividad, n prevee los casos en que una relación une más de dos conceptos a y b .

¹ Para más información sobre este tema véase Feliu, J. (2000).

El siguiente ejemplo ilustra la noción de relación conceptual que acabamos de definir:

[El genoma nuclear]_a [está formado por]_R [22 pares de cromosomas autosómicos]_b y por [los cromosomas sexuales]_n designados X e Y.

Donde:

- a: el genoma nuclear
- R: está formado por (meronimia)
- b: 22 pares de cromosomas autosómicos
- n: los cromosomas sexuales

1.1 Tipología de relaciones conceptuales

La tipología tradicional de relaciones conceptuales ha sido reutilizada repetitivamente en terminología pero ha debido ser reconsiderada cuando se ha trabajado con corpus. Nuestra revisión teórica nos ha llevado a establecer la tipología de relaciones conceptuales que se indica en la tabla siguiente. Para cada tipo de relación conceptual explicitamos el nombre y símbolo de la relación, la definición de la misma, el subtipo de relación conceptual y la lista de patrones lingüísticos prototípicos para cada tipo de relación:

TIPO DE RELACIÓN CONCEPTUAL	DEFINICIÓN	SUBTIPO DE RELACIÓN CONCEPTUAL	SÍMBOLO	PATRÓN LINGÜÍSTICO
Similitud (SEMBL.)	Indica equivalencia u oposición entre dos o más elementos.	positiva: - <i>equivalencia total o sinonimia</i> - <i>equivalencia parcial o similitud</i>	++ +	<i>Ser semblant a (ser pareccido a)</i>
		negativa: - <i>total u oposición</i> - <i>parcial o contraste</i>	-- -	<i>Ser diferent de (ser diferente de)</i>
Inclusión (INCL.)	Indica la inclusión de algunas características de un elemento en otro.	de clase o hiponímica	CLASS.	<i>Ser (un tipus) de (ser un tipo de)</i>
Secuencialidad (SEQ.)	Indica la localización o la aparición en el tiempo o en el espacio de dos o más elementos.	espacial: - <i>localización</i>	ESP. loc.	<i>Ser en / ser davant / ser darrere/ anar de x a y (estar en / estar delante/ estar detrás / ir de x a y)</i>
		- <i>dirección</i>	ESP. dir.	
		temporal:	TMP.	<i>Ser simultani / anterior /</i>

		- <i>simultaneïdad</i> - <i>anterioridad-posterioridad</i>	<i>sim.</i> TMP. <i>ant-pos</i>	<i>posterior a (ser simultaneo / anterior / posterior a)</i>
Causalidad (CAUS.)	Indica la causa y el efecto entre los elementos implicados.	causal: - <i>causa-efecto</i> - <i>proceso-resultado: con afectaci3n</i> - <i>proceso resultado: sin afectaci3n</i>	CAUS. CAUS-PROC. > <	<i>Causar / ser la causa de / ser l'efecte de (causar / ser la causa de / ser el efecte de).</i> <i>Produir / fer que (producir / hacer que)</i>
Instrumental (INST.)	Indica la relaci3n que se mantiene entre una herramienta y su funci3n.	<i>instrumento-funci3n</i>		<i>Servir per a / fer-se amb (servir para / hacerse con)</i>
Meronymia (MER.)	Indica la relaci3n que se establece entre un elemento considerado el todo y sus partes.	<i>componente-objeto</i> <i>miembro-colecci3n</i> <i>porci3n-masa</i> <i>material-objeto</i> <i>etapa-proceso</i> <i>lugar-3rea</i>	P-T.	<i>Ser una part / element de; tenir + SN / estar format / fet per; incloure; constar de / pertanyer a (ser una parte / elemento de; tener + SN / estar format / hecho por; incluir; constar de; pertenecer a)</i>
Asociaci3n (ASS.)	Indica la proximidad sem3ntica entre dos o m3s elementos que est3n relacionados tem3ticamente.			<i>Correlacionar-se amb (correlacionarse con)</i>

En el siguiente apartado presentamos un cat3logo de los marcadores lingüísticos detectados para cada tipo de relaci3n y que han sido retomados en el an3lisis del texto sobre el genoma humano.

1.2 Hacia un cat3logo de marcadores lingüísticos

Para cada tipo de relaci3n conceptual hemos detectado y aislado los siguientes marcadores lingüísticos².

Similitud

POS. ++	POS. +	NEG. --	NEG. -
O (<u> </u>) és a dir	<i>assemblar-se a</i>	<i>al contrari que ser el contrari a</i>	<i>diferenciar-se de</i>

Inclusi3n

² La lista de marcadores lingüísticos (en catal3n) que detallamos en este apartado procede de un experimento realizado sobre un corpus de cardiopatías en el marco de un trabajo de investigaci3n de doctorado, Feliu (2000) y de los marcadores estudiados para las relaciones meronímicas en Sol3 (2002) y Feliu, Sol3, Teb3 (en prensa).

INCL.
<i>ser (+SN)</i> <i>∅ (format)</i> <i><_> considerar-se</i> <i>∅ com (x o y)</i> :

Secuencialidad

ESP. LOC	ESP. DIR.	TMP. ANT.-POS.
<i>-iniciar-se en</i> <i>-produir-se en</i> <i>-tenir lloc a nivell de</i> <i>-quedar encarat amb</i> <i>-realitzar-se sobre</i> <i>-situar sobre</i> <i>-registrar-se en / a / des de</i> <i>-evidenciar-se a</i> <i>-originar a</i> <i>-veure's en</i> <i>-ocórrer (en presència de)</i> <i>-aparèixer fins que (=quan)</i>	<i>-propagar-se a través de x / cap a y</i> <i>-continuar per / fins</i> <i>-arribar a / a través de</i> <i>-mesurar-se des de x fins a y</i> <i>-apropar-se a</i> <i>-allunyar-se de</i> <i>-produir-se cap a</i>	<i>-correspondre primer a x i posteriorment a y</i> <i>-localitzar (primer) i després</i> <i>-(veure's) abans i després de</i> <i>-transcórrer des de x fins a y</i> <i>-ser seguit de / seguir-se de</i>

Causalidad

CAUS.	CAUS.-PROC. >	CAUS.-PROC. <
<i>-dependre de</i> <i>-fer que</i> <i>-ser la causa (principal) de</i> <i>-deure's a</i> <i>-implicar</i> <i>-aparèixer</i> <i>-contribuir a</i> <i>-dependre de</i> <i>-emascarar</i> <i>-donar lloc a</i> <i>-<cond.> trobar :</i> <i>-renforçar</i> <i>-provocar</i>	<i>-augmentar</i> <i>-produir</i> <i>-transformar-se en</i> <i>-augmentar</i>	<i>-obtenir-se (connectant / desplaçant / unint x amb y / prolongant)</i> <i>-donar lloc a</i> <i>-permetre (construir)</i>

Instrumental

INST.
<i>-servir com a</i> <i>-realitzar-se amb</i>

Meronymia

MER.
<i>-definir-se tres grups:</i> <i>-constar de</i> - : <i>-tenir</i> <i>-mostrar</i> <i>-incloure</i> <i>-caracteritzar-se per</i> <i>-presentar</i> <i>-zona</i> <i>-part</i> <i>-fase</i> <i>-etapa</i> <i>-període</i> <i>-compondre</i> <i>-constituir</i> <i>-formar</i> <i>-integrar</i> <i>-agrupar</i> <i>-aplegar</i> <i>-englobar</i> <i>-reunir</i> <i>-entre</i>

-unànimement -unànime, -en conjunt -en pes -nombrós, -osa -nodrit, -ida -escàs, -assa

Asociación

ASS. -correlacionar-se amb -correspondre a / considerar-se com a corresponent a -representar-se amb -reflectir -registrar / registrar-se en -ser imprescindible per realitzar -determinar -confondre's amb -simular -intervenir (en) -suggerir -ser indicatiu de -veure's en -manifestar-se com -indicar / ser indicatiu de -presentar-se com a
--

2. Los marcadores lingüísticos de relaciones conceptuales en un corpus textual sobre genoma humano

Para validar la lista de marcadores lingüísticos propuestos en trabajos anteriores sobre muestras de corpus textuales especializados de distintos ámbitos (las cardiopatías en Feliu (2000), y manuales médicos, económicos y jurídicos en Feliu, Solé y Tebé (en prensa)), proponemos a continuación la detección de las relaciones conceptuales que aparecen en un texto del ámbito del genoma humano. Este trabajo ha sido elaborado en el marco del proyecto Texterm del grupo IULATERM. Hemos seleccionado para esta presentación los dos apartados iniciales del artículo «Composición del genoma humano» publicado en el libro *Genoma Humano*, y que reproducimos en el anexo 1.

Abordar la detección de las distintas relaciones conceptuales (RC) a partir de la identificación de marcadores lingüísticos explícitos en este texto tiene tres objetivos inmediatos y otro a medio plazo:

- a) Validar y completar la tipología de marcadores lingüísticos propuestos en trabajos anteriores, y destacar los más prototípicos.

- b) Mostrar la posibilidad de marcadores polisémicos, es decir, como un mismo marcador lingüístico puede servir para expresar y, posteriormente, detectar distintas relaciones conceptuales. Y como tratar estas posibles interferencias entre RC cuando se plantea la detección automática.
- c) Analizar las características de las unidades de conocimiento especializado que intervienen en la relación conceptual como elemento *a*, que precede al marcador, y como elemento *b*, que sigue al marcador.
- d) postular un sistema semiautomático orientado a detectar y etiquetar las relaciones conceptuales y las unidades terminológicas que aparecen precediendo y siguiendo a cada marcador lingüístico explícito.

2.1 Algunas cuestiones previas

Siguiendo el esquema inicial de definición de relación conceptual, hemos dejado de lado las relaciones que se establecen *dentro* de las unidades terminológicas complejas puesto que en este estudio nos interesa básicamente la relación *entre* las unidades terminológicas. Por ejemplo, desestimamos casos como *un orgánulo de la célula*, que nos llevaría a establecer que *organulo is_part_of / MER* de una *célula*.

En este trabajo, tan sólo consideramos relevante la relación conceptual que aparece indicada en el texto mediante un marcador lingüístico **explícito** entre *a* y *b*, considerando ambas unidades como “unidades de conocimiento especializado”. Así pues, hemos detectado todos los fragmentos que aparecen en el texto pero hemos desestimado casos como los siguientes para el análisis posterior:

- a) cuando el texto presenta relaciones no explícitas verbalmente pero sí con marcadores no verbales (puntuación, tipografía, etc.) del tipo “la palabra «genoma»”, donde [*genoma*]_a [es]_{R_INCL} [una palabra]_b.

- b) cuando *a* o *b* no són una unidad de conocimiento especializado (unidad terminológica o no), sino una información enciclopédica como en el ejemplo: [El término *gen*]_a [fue acuñado por]_{R_CAUS} [Johansen]_b en 1909, donde *b* es un antropónimo.
- c) cuando el marcador no tiene *a* y *b* a ambos lados, sino de forma continua como en el ejemplo siguiente: [se sobreentiende por]_{R_PRED} [*gen*]_a [una secuencia de DNA con todos sus elementos reguladores de la transcripción que da lugar a una proteína o a un RNA]_b

2.2 Validación de la tipología

A continuación presentamos los distintos marcadores lingüísticos explícitos detectados en el texto que permiten la identificación de las distintas relaciones conceptuales:

- Relación de inclusión

enumerarse

[Los cromosomas autosómicos]_a [se enumeran según su tamaño:]_{R_INCL} [el cromosoma más grande es el 1,]_b [el siguiente en tamaño decreciente es el 2]_n y así sucesivamente.

- Relación de secuencialidad (temporal y espacial):

avanzar en

[El concepto de *gen*]_a [<ha ido cambiando a medida que> se ha avanzado en]_{R_SEQ} [su conocimiento]_b

localizar en

[el complejo mayor de histocompatibilidad]_a [se localiza en]_{R_SEQ} [el cromosoma 6, brazo pequeño, banda xxx]_b
 [un solo tipo de molécula de DNA]_a [localizado en]_{R_SEQ} [las mitocondrias]_b

aparecer en

[éstas <mitocondrias>]_a [aparecieron en]_{R_SEQ} [los eucariotas]_b

presente en

[el material genético]_a [presente en]_{R_SEQ (MER)} [virus]_b

existente en

[uno de los dos cromosomas homólogos de cada autosoma]_a [existentes en]_{R_SEQ (MER)} [una célula diploide en metafase]_b

- Relación de causalidad:

producir

[«secuencia de información»]_a [que produce]_{R_CAUS} [un producto funcional]_b

dar lugar a

[una secuencia de DNA con todos sus elementos reguladores de la transcripción]_a [da lugar a]_{R_CAUS} [una proteína]_b o [un RNA]_n

adquirir

[los eucariotas resultantes]_a [habrían adquirido]_{R_CAUS} [la capacidad de fosforilación oxidativa que permitía un crecimiento rápido en una atmósfera rica en O₂]_b

permitir

[la capacidad de fosforilación oxidativa]_a [permitía]_{R_CAUS} [un crecimiento rápido en una atmósfera rica en O₂]_b

dividir

[que <centrómero>]_a [divide]_{R_CAUS} [al cromosoma en dos brazos]_b

- Relación instrumental:

visible con

<Cuando las células se dividen por mitosis,> [los cromosomas...]_a [haciéndose visibles con]_{R_INST} [el microscopio óptico]_b

- Relación de meronimia:

contener, contenido en

[La palabra «genoma»]_a [contiene]_{R_MER} [la raíz griega *gen* (...)]_b y [la extensión también griega *oma* (...)]_n

[una pequeña molécula de DNA circular]_a [contenida en]_{R_MER} [un órgano de la célula denominado mitocondria (genoma mitocondrial)]_b

estar compuesto por, estar formado por

[El Genoma Humano]_a [está compuesto por]_{R_MER} [23 pares de moléculas de DNA contenidas en estructuras llamadas cromosomas, que se localizan en el núcleo de la célula (genoma nuclear), y por una pequeña molécula de DNA circular contenida en un órgano de la célula denominado mitocondria (genoma mitocondrial)]_b

[El genoma nuclear]_a [está formado por]_{R_MER} [22 pares de cromosomas autosómicos]_b y [los cromosomas sexuales designados X e Y]_n

tener

[Cada cromosoma]_a [tiene]_{R_MER} [una constricción central denominada centrómero que divide al cromosoma en dos brazos, designados «p» (para el brazo pequeño) y «q» (para el brazo grande; «q» sigue a «p» en el alfabeto)]_b

constar de

[El genoma mitocondrial]_a [consta de]_{R_MER} [un solo tipo de molécula de DNA localizado en las mitocondrias, pero del que pueden existir varios miles de copias por célula.]_b

existente en

[uno de los dos cromosomas homólogos de cada autosoma]_a [existentes en]_{R_MER (SEQ)} [una célula diploide en metafase]_b

presente en

[el material genético]_a [presente en]_{R_MER (SEQ)} [virus]_b

rico –a en

[una atmósfera]_a [rica en]_{R_MER} [O2]_b

- Relación de asociación:

aplicarse a

[El término «gen»]_a [aplicado a]_{R_ASSOC} [la herencia]_b

asociarse a

[las unidades de herencia]_a [asociadas a]_{R_ASSOC} [un carácter transmisible específico]_b

relacionarse con

[algunos genes (37)]_a [relacionados con]_{R_ASSOC} [el metabolismo oxidativo]_b

ligarse a

[el origen del genoma mitocondrial]_a [va ligado a]_{R_ASSOC} [al propio origen de las mitocondrias]_b

característico de

[el contenido total de material genético]_a [característico de]_{R_ASSOC (MER)} [un organismo]_b

Hemos dejado en último lugar las relaciones de similitud porque sus marcadores lingüísticos se entremezlan con los de las relaciones predicativas y metalingüísticas. Consideramos que se establece una *relación predicativa* cuando el marcador pone en relación unidades terminológicas (como elemento *a*) con sus definiciones o paráfrasis explicativas (elemento *b*), o bien el elemento *b* aporta información atributiva a la unidad terminológica (*a*);

información que puede reconvertirse en atributos de la unidad terminológica con su correspondiente valor.

Sobresalen los marcadores del tipo *significar, designar, ser, definirse como, tratarse de*:

[la raíz griega *gen*]_a [significa]_{R_PRED} [origen]_b

[El término «gen»]_a [para designar]_{R_PRED} [a las unidades de herencia asociadas a un carácter transmisible específico]_b

Una de las definiciones actuales más amplia de [gen]_a [sería]_{R_PRED} [«secuencia de información que produce un producto funcional»]_b

[«Genoma»]_a [puede definirse como]_{R_PRED} [el contenido total de material genético característico de un organismo]_b

[El tamaño del genoma mitocondrial]_a [es]_{R_PRED} [extraordinariamente pequeño (xxx) comparado con el del genoma nuclear (xxx)]_b

Entendemos, en cambio, como *relación metalingüística* los casos en los que marcadores parecidos a los anteriores —*designado (como), llamado, denominado, denominarse, expresarse, ser, etc.*—, ponen en relación la definición o paráfrasis explicativa de una unidad terminológica (*a*) con dicha unidad (*b*). Por ejemplo:

[estructuras]_a [llamadas]_{R_META} [cromosomas]_b

[un órgano de la célula]_a [denominado]_{R_META} [mitocondria]_b

[El conjunto de cromosomas metafásicos de una célula]_a [se denomina]_{R_META} [cariotipo]_b

[una constricción central]_a [denominada]_{R_META} [centrómero]_b

[el cromosoma más grande]_a [es]_{R_META} [el 1]_b

Finalmente hablamos de *relaciones de similitud* cuando tenemos dos unidades terminológicas *a* y *b* relacionadas con un marcador lingüístico que expresa equivalencia total o parcial. Por ejemplo:

[los genes]_a [habían sido designados como]_{R_SEMBL} [«elementos»]_b

[los cromosomas sexuales]_a [designados]_{R_SEMBL} [X e Y]_b

[dos brazos]_a [designados]_{R_SEMBL} [«p» (para el brazo pequeño)]_b y [«q» (para el brazo grande)]_n

2.3 Conclusiones acerca de los marcadores

Las dos conclusiones principales que se pueden extraer acerca de los marcadores lingüísticos explícitos son:

- a) Las relaciones pueden expresarse mediante elementos léxicos autónomos, mayoritariamente verbos aunque no únicamente, o bien a través de verbos o adjetivos que requieren una o más preposiciones.
- b) Algunos de los marcadores transmiten un valor semántico en particular, es decir, el significado de la unidad léxica transfiere el sentido de la relación (por ejemplo: *producir*, *compuesto por*, *relacionarse con*). En otros casos, algunos marcadores aparecen como unidades polisémicas y su consideración vehiculando una determinada relación depende, en la mayoría de casos, del tipo de unidades que vinculan. Por ejemplo, el marcador lingüístico *es un* se considera directamente como marcador de inclusión cuando el concepto *a* es hipónimo del concepto *b* (que es su hiperónimo) como ocurre en el ejemplo *el ADN es un ácido*. Ahora bien, la relación de inclusión no es tan clara en el caso de *el ADN es una doble hélice*, donde aparece la noción de similitud.

2.4 Interferencias entre relaciones conceptuales: marcadores polisémicos

De acuerdo con los datos analizados y pasando a cuestiones más específicas, quisiéramos destacar, tal como habíamos avanzado anteriormente, que hemos detectado algunos casos en los que el marcador o expresión lingüística explícita puede vehicular más de una relación conceptual. Por ejemplo: *característico de* (asociativa y meronímica), *presente en*, *existente en* (secuencialidad localización y meronímica), *mediante* (causalidad y instrumental), etc.

El análisis de algunos de estos casos y similares se está llevando a cabo en el marco de una tesis doctoral dentro del grupo IULATERM (Feliu, en curso) donde se trabaja en el tratamiento de parejas como las siguientes:

- Similitud-inclusión
- Secuencialidad localización-meronimia (lugar-área)
- Causalidad-secuencialidad temporal

2.5 Características de las unidades de conocimiento especializado

A la vista de los datos reportados en el apartado anterior, consideramos que una relación conceptual se establece, efectivamente, entre dos conceptos expresados por una unidad terminológica más o menos larga. Como se ha podido observar, en esta primera exploración del texto hemos recogido, no obstante, fragmentos de longitud superior para ver cómo puede ser tratada esta información (por ejemplo, en los casos en que este tipo de información especializada es necesaria para que la relación se manifieste).

Así, el esquema de trabajo de base es UT_R_UT, donde UT significa unidad terminológica y R, relación conceptual. Sin embargo, hemos detallado, así mismo, los casos en los que aparecía una paráfrasis (P) en alguno de los dos lados de la relación siguiendo el esquema P_R_UT o bien UT_R_P. En estos casos, creemos que se establece una cierta equivalencia entre las dos unidades (una con valor especializado más elevado, la UT, y la otra actuando generalmente como explicación o definición, la P). O bien, donde una unidad terminológica se relaciona conceptualmente con una unidad de conocimiento especializado más genérica, que no es prototípica del ámbito analizado. Aunque no se trata de unidades terminológicas, hemos recogido estas unidades por cuanto representan una información especializada muy valiosa para definir y atribuir propiedades a una determinada UT (siguiendo el enfoque comunicativo de Cabré, 1999); valiosa puesto que puede convertir la información que en algunos casos aparece como predicativa en propiedades o atributos de la unidad terminológica con su correspondiente valor.

3. A modo de conclusión: hacia un sistema de detección semiautomática de relaciones conceptuales

Sobre la base del catálogo y del análisis que hemos presentado someramente en el apartado anterior, presentamos seguidamente las propiedades principales que debería tener un sistema semiautomático orientado a detectar las relaciones conceptuales y la terminología que aparece cerca de cada marcador de relación conceptual. A partir de los marcadores lingüísticos descritos anteriormente, el sistema podría detectar una unidad concreta y en un texto especializado *T*. Una vez se ha aislado la unidad en su forma lematizada, el sistema debería compararla con la lista de marcadores del catálogo. En un primer momento, el sistema trabaja analizando oración por oración y, cuando encuentra una correspondencia entre la unidad y algún marcador lingüístico, lo marca con una etiqueta de relación conceptual. Esta etiqueta debería ser validada manualmente, sobre todo en los casos en que el sistema proponga dos o más posibilidades para los casos de polisemia que ya hemos visto en los apartados precedentes.

Seguidamente, con la ayuda de un extractor de terminología que funciona en el marco del grupo de investigación (Estopà, 1999 y Vivaldi, 2001), el sistema propondría las unidades de conocimiento especializado que aparecen en cada fragmento textual.

Consideremos el ejemplo siguiente para el catalán:

*D'origen cardíac, [un dolor semblant al de l'angina de pit]_a **provoca** [alteracions a la vàlvula aòrtica]_b*

El sistema sería capaz de aislar el verbo *provocar* y compararlo con el catálogo de marcadores lingüísticos. En este caso, sería coincidente con uno de los indicadores de causalidad. El verbo estaría etiquetado como CAUS. Al mismo

tiempo, el extractor de candidatos a término propondría *dolor* y *alteració* (alteración) como candidatos a términos. La suma de la relación de causalidad y el posible valor de término de estas unidades incrementaría las posibilidades de este fragmento de ser considerado una unidad de conocimiento especializado que sería recuperada para una posterior representación de la información especializada del texto analizado.

Referencias bibliográficas

CABRÉ, M. T. (dir) (1999) *La terminología: Representación y comunicación. Una propuesta de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. (Sèrie Monografies, 3)

CABRÉ, M. T. (2001) «Sumario de principios que configuran la nueva propuesta teórica» En: CABRÉ, M. T.; FELIU, J. (ed.) (2001) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica (DGES PB96-0293)*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 17-26.

CABRÉ, M. T. (2001) «Consecuencias metodológicas de la propuesta teórica (I)» En: CABRÉ, M. T.; FELIU, J. (ed.) (2001) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica (DGES PB96-0293)*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 27-36.

CABRÉ, M. T. (en prensa) «Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción». En: *Actas de las VI Jornadas de Traducción: La traducción científico-técnica y la terminología en la sociedad de la información*. Universitat Jaume I de Castelló. 24-26 de octubre de 2001.

ESTOPÀ, R. (1999). *Extracció de Terminologia: Elements per a la Construcció d'un SEACUSE (Sistema d'Extracció Automàtica de Candidats a Unitats de*

Significació Especialitzada). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. [tesis doctoral]

FELIU, J. (2000) *Relacions conceptuals i variació funcional: elements per a un sistema de detecció automàtica*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. [trabajo de investigación]

FELIU, J. (2001) «Propuesta de clases conceptuales y de relaciones conceptuales: recopilación y análisis». En: CABRÉ, M. T.; FELIU, J. (ed.) (2001) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica (DGES PB96-0293)*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 143-154.

FELIU, J. (en curso) *Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica*. Barcelona: Institut Universitari de Lingüística Aplicada. [tesis doctoral]

FELIU, J.; SOLÉ, E.; TEBÉ, C. (en prensa) «Las relaciones meronímicas en terminología: análisis semántico-textual y aplicaciones». Ponencia presentada en el *VII Simposio Internacional de Terminología*. Lisboa, noviembre 2000.

FELIU, J.; CABRÉ, M. T. (2002) «Conceptual relations in specialized texts: new typology and an extraction system proposal». *TKE2002. Terminology and Knowledge Engineering Proceedings. 6th International Conference. 28th-30th August 2002*. Nancy, 45-49.

OLIVA VIRGILI, R. (1996) «Composición del genoma humano». En: *Genoma Humano*. Barcelona: Editorial Masson, SA.

SOLÉ, E. (2001) «El concepto de colectivo representado nominalmente y su incidencia en textos de especialidad». En: CABRÉ, M. T.; FELIU, J. (ed.) (2001) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica (DGES PB96-0293)*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 155-164.

SOLÉ, E. (2002) *Els noms col·lectius en català: descripció i reconeixement*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. [tesis doctoral]

TEBÉ, C. (2001) «Elementos para la representación de los conceptos especializados en bases y bancos de datos». En: CABRÉ, M. T.; FELIU, J. (ed.) (2001) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica (DGES PB96-0293)*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 203-208.

VIVALDI, J. (2001) *Extracción de candidatos a términos mediante la combinación de estrategias heterogéneas*. Barcelona: Departament de Llenguatges i Sistemes Informàtics de la Universitat Politècnica de Catalunya, Universitat Politècnica de Catalunya. [tesis doctoral]

Anexo 1

Texto de análisis

COMPOSICIÓN DEL GENOMA HUMANO

DEFINICIÓN DEL GENOMA HUMANO

La palabra «genoma» contiene la raíz griega *gen*, que significa origen, y la extensión también griega *oma* empleada en sustantivos del vocabulario biológico y médico. El término «gen» aplicado a la herencia fue acuñado por Johansen en 1909 para designar a las unidades de herencia asociadas a un carácter transmisible específico. Previamente, los genes habían sido designados como «elementos» por Mendel (1866), *gemmales* por Darwin (1868), «pangenes» por De Vries (1889), «unidades fisiológicas» por Spencer (1864), o simplemente «carácter unidad», «factor unidad» o «factor». El concepto de gen ha ido cambiando a medida que se ha avanzado en su conocimiento. Una de las definiciones actuales más amplia de gen sería: «secuencia de información que produce un producto funcional». Normalmente se sobrentiende por gen una secuencia de DNA con todos sus elementos reguladores de la transcripción que da lugar a una proteína o a un RNA. En el capítulo 2 se analizará con mayor detalle el concepto de gen. «Genoma» puede definirse como el contenido total de material genético característico de un organismo. También se utiliza el término para indicar el material genético presente en virus.

GENOMA NUCLEAR Y GENOMA MITOCONDRIAL

NOMENCLATURA DE LOS CROMOSOMAS

El Genoma Humano está compuesto por 23 pares de moléculas de DNA contenidas en estructuras llamadas cromosomas, que se localizan en el núcleo de la célula (genoma nuclear), y por una pequeña molécula de DNA circular contenida en un órgano de la célula denominado mitocondria (genoma mitocondrial). Cuando las células se dividen por mitosis, los cromosomas se condensan, haciéndose visibles con el microscopio óptico y apareciendo como unidades discretas. El conjunto de cromosomas metafásicos de una célula se denomina cariotipo (fig., 1-1).

El genoma nuclear está formado por 22 pares de cromosomas autosómicos y por los cromosomas sexuales designados X e Y. Los cromosomas autosómicos se enumeran según su tamaño: el cromosoma más grande es el 1, el siguiente en tamaño decreciente es el 2 y así sucesivamente. Una excepción es la del cromosoma más pequeño, que es el 21 en lugar del 22. Esta excepción se debe al hecho de que inicialmente se pensó que el cromosoma 22 era el más pequeño, y aunque actualmente se sabe que esto no es así, la nomenclatura se ha mantenido.