



Jornades de Foment de la Investigació

**PATRONES LIN-
GÜÍSTICOS PARA
LA BÚSQUEDA
DE INFORMACIÓN
CONCEPTUAL EN
EL CORPUS TEX-
TUAL ESPECIALI-
ZADO DE LA CERÁ-
MICA TXTCeram^a**

Autora
Victoria SOLER PUERTES

1. INTRODUCCIÓN

A lo largo de estas últimas décadas se ha producido un aumento de la necesidad del estudio de la lingüística de corpus con fines investigadores. Por eso, los corpus se han convertido en un elemento de referencia que permite estudiar de manera empírica el comportamiento de la lengua basándose en datos reales. La tecnología también ha tenido un papel importante en el desarrollo de esta disciplina ya que, gracias a ella, es posible tratar grandes volúmenes de información de manera objetiva y precisa. Estas ventajas electrónicas para el tratamiento de corpus afectan tanto a su constitución como a su procesamiento y explotación.

Este trabajo se enmarca en el proyecto de investigación *TXTCeram: Extracción semiautomática y análisis conceptual formal de términos de la cerámica a partir de un corpus electrónico. Su eficacia y utilidad en la mediación lingüística* cuya finalidad es la creación de un corpus electrónico del campo de la cerámica industrial para su posterior extracción terminológica.

Nuestro principal objetivo es encontrar patrones lingüísticos que ayuden a localizar las relaciones semánticas que expresen metonimia en el corpus de la cerámica TXTCeram. Gracias a dichos patrones, se podrá automatizar la búsqueda de relaciones conceptuales en el corpus para la creación de bases de conocimiento. Es decir, al introducir en un corpus textual electrónico una serie de estructuras lingüísticas, estas estructuras nos dan información sobre las relaciones semánticas (hiponimia e hiperonimia, metonimia, sinonimia, etc.) que existen entre los conceptos del corpus.

1.1. LOS CORPUS ELECTRÓNICOS: UNA HERRAMIENTA PARA LA INVESTIGACIÓN

Según Sinclair (1991: 171), un corpus es «a collection of naturally-occurring language text, chosen to characterize a state or variety of a language». Sin embargo, esta definición puede resultar hoy en día un tanto obsoleta porque un corpus implica algo más que simplemente una colección de textos auténticos. Por esto mismo, a esta definición cabría añadirle una serie de aspectos fundamentales que determinan la validez del corpus con el que se va a trabajar. En primer lugar, en la creación de un corpus hay que tener en cuenta la representatividad de los textos elegidos con respecto a la lengua que se quiere investigar y, en segundo lugar, aunque no menos importante, es muy relevante indicar si el corpus está disponible en formato electrónico.

Sánchez Gijón (2003: 187-193) en su tesis doctoral realiza una detallada clasificación de los distintos tipos de corpus según las diferentes características y finalidades para las que hayan sido creados. Así, adoptamos la siguiente tipología de Sánchez Gijón según:

- Su modalidad → oral o escrito.
- El etiquetado → no etiquetado o etiquetado.
- Aspectos cronológicos → sincrónico/diacrónico, periódico o cronológico.
- Aspectos temáticos → general y especializado.
- La cantidad de información y su distribución → textuales, de referencia, equilibrados, extensos o monitor.
- La información extralingüística → documentado o no documentado.
- Las lenguas implicadas → monolingüe, bilingüe o multilingüe.
- La relación entre las lenguas → comparable, paralelo/alineado o de traducciones libres.
- Su especialidad → con finalidad especial o *ad hoc*.

La principal razón por la que la lingüística de corpus se ha convertido en una herramienta indispensable es por su capacidad de aplicación en distintos campos de estudio como son la terminología, la lexicografía o la traducción automática, entre otros. Esto se debe a que no existe un único fin para el que se constituyen los corpus sino que, dependiendo de la información que se pretenda obtener se seguirá una metodología para su explotación u otra. Sin embargo, sea cual sea esta metodología, para su procesamiento es necesario utilizar un programa de análisis de corpus textuales, como por ejemplo WordSmith, con el que se pueden extraer listados de frecuencias de uso, concordancias y listados de palabras clave.

1.2. CONTEXTOS RICOS EN CONOCIMIENTO

La terminología, como hemos mencionado más arriba, también hace uso de la lingüística de corpus para una de sus actividades más novedosas: la extracción automática o semiautomática de términos y de información conceptual. Esta actividad consiste en la búsqueda de contextos que ofrezcan información rica en conocimiento lingüístico (*knowledge-rich contexts*) y que ayuden en la elaboración de la estructura conceptual de un área de especialización concreta. Por tanto, tal y como afirma Pérez Hernández (2002: 160), «el estudio de la estructura de conocimiento que subyace a un dominio de especialidad es de vital importancia en el trabajo terminológico».

Meyer (2001: 281) define los contextos ricos en conocimiento del siguiente modo:

By knowledge-rich context, we designate a context indicating at least one item of domain knowledge that could be useful for conceptual analysis. In other words, the context should indicate at least one conceptual characteristic, whether it be an attribute or a relation.

A partir de los contextos ricos en conocimiento es posible conseguir información que dé cuenta de las relaciones semánticas que existen entre varios conceptos, como por ejemplo de las relaciones de hiponimia o de metonimia. Además de estas relaciones, también es posible encontrar contextos ricos en conocimiento desde los que extraer información que ayude a la creación de las definiciones terminológicas del área de dominio.

La tecnología también está presente en este momento del trabajo terminológico ya que ayuda a acceder a la información con mayor celeridad y precisión. Uno de los programas que se pueden utilizar para la búsqueda de contextos ricos en conocimiento es el programa de análisis textual WordSmith con el que se consiguen listados de concordancias, de frecuencia y de palabras clave. No obstante, WordSmith no es el único programa que puede desempeñar estas funciones ya que en la Universidad de Ottawa se ha desarrollado un programa llamado Text Analyzer (TA) que también permite la extracción de conocimiento para el trabajo terminológico. Como bien explica Davidson (1998: 3):

The “Conceptual Operations” function of the TA allows users to extract (from an electronic corpus treating a given subject field) sentences that express the semantic relations that hold between the concepts in that field.

Sin embargo, para poder recuperar esta información tan útil del corpus para el estudio terminológico es extremadamente necesario conocer algún mecanismo que permita la obtención inmediata de

la información que buscamos. Por ejemplo, se pueden usar los patrones lingüísticos para explorar el corpus en busca de contextos ricos en conocimiento.

1.3. HACIA EL ESTUDIO DE PATRONES LINGÜÍSTICOS

Al expresarnos utilizamos una serie de estructuras lingüísticas con las que organizamos nuestro discurso y que ayudan al receptor a asimilar la información conceptual que emitimos. Por lo tanto, si encontramos estas estructuras, que llamaremos patrones, y las introducimos en un programa de extracción de información como WordSmith, nos aportarán la información requerida.

El uso de patrones lingüísticos para la extracción de relaciones conceptuales es bastante reciente en terminología por lo que todavía no posee una denominación establecida. Por eso, cabe destacar las diferentes maneras con las que los diversos autores han designado lo que nosotros llamamos en este trabajo patrones lingüísticos, como por ejemplo la denominación *patterns* que utiliza el grupo COGNITERM dirigido por Meyer (2001). El siguiente listado de términos lo hemos tomado de Marshman y otros (2002: 2): *formulae* (Lyons, 1977), *diagnostic frames* o *test frames* (Cruse, 1986), *frames* (Winston y otros, 1987), *knowledge probes* (Ahmad and Fulford, 1992), *definitional metalanguage* y *defining expositives* (Pearson, 1996). En español, destacamos a Bach (2005) que utiliza la denominación de *marcadores de reformulación (MR)* y Faber (2001) que también utiliza el mismo término que en el presente artículo, *patrones*. A pesar de las diferentes denominaciones, todos los autores se refieren al mismo concepto que Condamines (2002: 145) llama *marqueurs* en francés o *patterns* en inglés y define como:

A discursive structure used as an indication of the possible transition from the discourse to a model, allowing the more or less direct construction of a model in the form of a semantic relation depending on its relation with the context.

De una manera más sencilla, Marshman y otros (2002: 2) los definen como «words, word combinations or paralinguistics features of texts which frequently indicate conceptual relations». A partir de esta definición, estas autoras señalan en este mismo artículo que existen tres tipos de patrones lingüísticos para la detección de contextos ricos en conocimientos:

- 1) Patrones léxicos. Son los más visibles y consisten en palabras o grupos de palabras que indican las relaciones conceptuales.
- 2) Patrones gramaticales. Implican combinaciones de estructuras gramaticales que ofrecen relaciones semánticas entre conceptos aunque de un modo más limitado que los patrones léxicos.
- 3) Patrones paralingüísticos. Se trata de los elementos textuales como por ejemplo comas, paréntesis, puntos, etc.

En resumen, gracias a estos patrones de conocimiento, el terminólogo es capaz de detectar, de una manera rápida y eficaz, las relaciones semánticas y las estructuras lingüísticas que le ayudarán en su trabajo.

1.4. SITUACIÓN ACTUAL DE LAS INVESTIGACIONES SOBRE PATRONES LINGÜÍSTICOS

El estudio de la búsqueda de los mencionados patrones es todavía una vía de investigación muy reciente y por eso, el número de trabajos que se han llevado a cabo sobre este tema es muy reducido. Además de esta escasa cantidad de estudios, también es importante subrayar que el número de lenguas de reconocimiento internacional con las que se ha trabajado aún es menor, reduciéndose únicamente al inglés, al francés y al español.

Los estudios de Ahmad y Fulford (1992), Pearson (1998) y Meyer (2001), que aparecen citados en la obra de Marshman y otros (2002: 5), son algunas de las investigaciones más desarrolladas sobre patrones en inglés de las que se tiene constancia. Junto a estos estudios en inglés, también podemos destacar los realizados para el francés por el grupo de trabajo COGNITERM de la Universidad de Ottawa, en la que se enmarcan los trabajos de investigación de Davidson (1998) y Morgan (2000).

Además del anterior grupo, igualmente cabe prestar atención a las investigaciones de Condamines y Rebeyrolle (2001) en la búsqueda de patrones lingüísticos que detecten relaciones semánticas.

En España, el grupo *ONCOTERM: Sistema bilingüe de información y recursos oncológicos*, con miembros de varias universidades españolas, trabaja asimismo en esta línea de investigación aunque su finalidad es la de crear una ontología de dominio en el área de la oncología. En Faber y Jiménez (2002), este grupo hace un repaso a los objetivos principales del proyecto y a los logros conseguidos.

1.5. LA RELACIÓN DE METONIMIA

Iris y otros (1988: 261) aseguran que «the nature of the part-whole relation has been particularly mysterious and controversial». Sin embargo, a pesar de la ambigüedad de esta relación, la problemática de la parte y el todo no es la finalidad del presente trabajo. Para la situación que nos ocupa, nos acogeremos a la clasificación que Otman (1996: 83) ha realizado para la relación de metonimia que él denomina *partie-de*:

- a. La parte en el sentido estricto. Por ejemplo, un dedo es parte de una mano.
- b. El constituyente, es decir, los elementos constituyentes de una mezcla que se no pueden disociarse del elemento compuesto, como en el caso del aceite parte de la mayonesa.
- c. El atributo, la propiedad característica de un todo. Por ejemplo, el amarillo y el rojo forman parte de la bandera española.

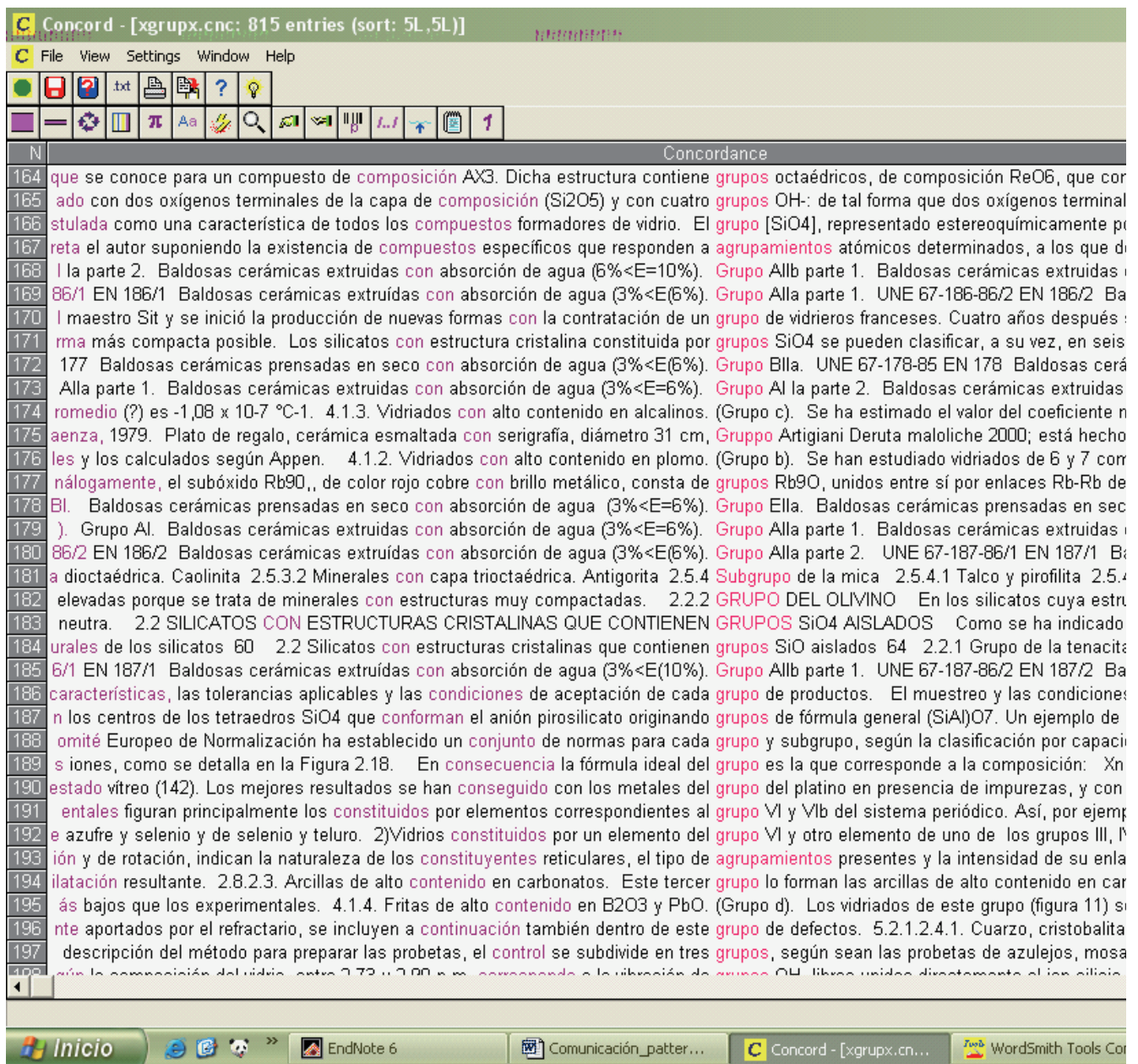
No obstante, Otman (1996: 84) afirma que esta clasificación no es del todo completa y que necesita ser revisada para que pueda considerarse totalmente válida.

2. FASES DE LA BÚSQUEDA DE PATRONES PARA LAS RELACIONES DE METONIMIA

Antes de comenzar a elaborar el estudio práctico para este trabajo, realizamos una perspectiva general de las metodologías empleadas por los diferentes autores mencionados más arriba. Tras esta revisión decidimos utilizar una metodología en la que se integrarán tanto el sistema arriba-abajo y como el

sistema abajo-arriba. Por lo tanto, los pasos seguidos para encontrar los patrones lingüísticos que ofrezcan información sobre la relación de metonimia son los siguientes:

- 1) A partir del corpus especializado en Cerámica, TXTCeram, que consta de 11 referencias, indicadas en el Anexo I al final del trabajo, y de un total de 909.523 palabras, se ha procedido a elaborar una lista de concordancias para la palabra de búsqueda *grup* con la aplicación *Concord* del programa de gestión de corpus WordSmith.



Concordance

164	que se conoce para un compuesto de composición AX ₃ . Dicha estructura contiene grupos octaédricos, de composición ReO ₆ , que cor
165	ado con dos oxígenos terminales de la capa de composición (Si ₂ O ₅) y con cuatro grupos OH-: de tal forma que dos oxígenos terminal
166	stulada como una característica de todos los compuestos formadores de vidrio. El grupo [SiO ₄], representado estereoquímicamente pr
167	eta el autor suponiendo la existencia de compuestos específicos que responden a agrupamientos atómicos determinados, a los que di
168	l la parte 2. Baldosas cerámicas extruidas con absorción de agua (6%<E=10%). Grupo Alb parte 1. Baldosas cerámicas extruidas
169	86/1 EN 186/1 Baldosas cerámicas extruidas con absorción de agua (3%<E=6%). Grupo Alla parte 1. UNE 67-186-86/2 EN 186/2 Ba
170	l maestro Sit y se inició la producción de nuevas formas con la contratación de un grupo de vidrieros franceses. Cuatro años después :
171	ma más compacta posible. Los silicatos con estructura cristalina constituida por grupos SiO ₄ se pueden clasificar, a su vez, en seis
172	177 Baldosas cerámicas prensadas en seco con absorción de agua (3%<E=6%). Grupo Blla. UNE 67-178-85 EN 178 Baldosas cerá
173	Alla parte 1. Baldosas cerámicas extruidas con absorción de agua (3%<E=6%). Grupo Al la parte 2. Baldosas cerámicas extruidas
174	romedio (?) es -1,08 x 10 ⁻⁷ °C ⁻¹ . 4.1.3. Vidriados con alto contenido en alcalinos. (Grupo c). Se ha estimado el valor del coeficiente n
175	aenza, 1979. Plato de regalo, cerámica esmaltada con serigrafía, diámetro 31 cm, Gruppo Artigiani Deruta malolliche 2000; está hecho
176	les y los calculados según Appen. 4.1.2. Vidriados con alto contenido en plomo. (Grupo b). Se han estudiado vidriados de 6 y 7 con
177	nálogamente, el subóxido Rb ₉ O ₄ , de color rojo cobre con brillo metálico, consta de grupos Rb ₉ O ₄ , unidos entre sí por enlaces Rb-Rb de
178	Bl. Baldosas cerámicas prensadas en seco con absorción de agua (3%<E=6%). Grupo Ella. Baldosas cerámicas prensadas en sec
179). Grupo Al. Baldosas cerámicas extruidas con absorción de agua (3%<E=6%). Grupo Alla parte 1. Baldosas cerámicas extruidas
180	86/2 EN 186/2 Baldosas cerámicas extruidas con absorción de agua (3%<E=6%). Grupo Alla parte 2. UNE 67-187-86/1 EN 187/1 B:
181	a dioctaédrica. Caolinita 2.5.3.2 Minerales con capa trioctaédrica. Antigorita 2.5.4 Subgrupo de la mica 2.5.4.1 Talco y pirofilita 2.5.4
182	elevadas porque se trata de minerales con estructuras muy compactadas. 2.2.2 GRUPO DEL OLIVINO En los silicatos cuya estru
183	neutra. 2.2 SILICATOS CON ESTRUCTURAS CRISTALINAS QUE CONTIENEN GRUPOS SiO ₄ AISLADOS Como se ha indicado
184	urales de los silicatos 60 2.2 Silicatos con estructuras cristalinas que contienen grupos SiO ₄ aislados 64 2.2.1 Grupo de la tenacit
185	6/1 EN 187/1 Baldosas cerámicas extruidas con absorción de agua (3%<E=10%). Grupo Alb parte 1. UNE 67-187-86/2 EN 187/2 Ba
186	características, las tolerancias aplicables y las condiciones de aceptación de cada grupo de productos. El muestreo y las condiciones
187	n los centros de los tetraedros SiO ₄ que conforman el anión pirosilicato originando grupos de fórmula general (SiAl) ₂ O ₇ . Un ejemplo de
188	omité Europeo de Normalización ha establecido un conjunto de normas para cada grupo y subgrupo, según la clasificación por capaci
189	s iones, como se detalla en la Figura 2.18. En consecuencia la fórmula ideal del grupo es la que corresponde a la composición: Xn
190	estado vítreo (142). Los mejores resultados se han conseguido con los metales del grupo del platino en presencia de impurezas, y con
191	entales figuran principalmente los constituidos por elementos correspondientes al grupo VI y Vlb del sistema periódico. Así, por ejem
192	e azufre y selenio y de selenio y telurio. 2) Vidrios constituidos por un elemento del grupo VI y otro elemento de uno de los grupos III, I'
193	ión y de rotación, indican la naturaleza de los constituyentes reticulares, el tipo de agrupamientos presentes y la intensidad de su enla
194	ilatación resultante. 2.8.2.3. Arcillas de alto contenido en carbonatos. Este tercer grupo lo forman las arcillas de alto contenido en car
195	ás bajos que los experimentales. 4.1.4. Fritas de alto contenido en B ₂ O ₃ y PbO. (Grupo d). Los vidriados de este grupo (figura 11) si
196	nte aportados por el refractario, se incluyen a continuación también dentro de este grupo de defectos. 5.2.1.2.4.1. Cuarzo, cristobalita
197	descripción del método para preparar las probetas, el control se subdivide en tres grupos , según sean las probetas de azulejos, mosa
198	ús la composición del vidrio, entre 2,73 y 2,99 g. corresponde a la liberación de grupos OH- libres unidos directamente al ion silicio

- 2) Una vez obtenidos los contextos, éstos se analizan en busca de patrones que impliquen relaciones de metonimia y se confecciona un listado de las estructuras extraídas con las diferentes combinaciones en las que aparecen. Por ejemplo, en la figura anterior encontramos la entrada 164 «Dicha

estructura contiene grupos octaédricos, de composición ReO₆, que [...]» en la que *contiene* sería la estructura con la que se detecta una relación de metonimia. En el Anexo II se muestra una tabla en la que describimos, con más detalle, todas estas estructuras mencionadas anteriormente junto con los posibles candidatos a patrones lingüísticos.

- 3) Cada uno de los patrones obtenidos en la fase anterior se prueba en el corpus TXTCeram para confirmar si realmente nos indican la información conceptual buscada. Al mismo tiempo, se anota la cantidad de aciertos (contextos válidos) y de ruido (contextos no válidos) que produce cada uno de ellos. Por ejemplo, con el patrón *contien** encontramos 295 entradas de las que tan sólo 15 son ruido y el resto (280) son aciertos.
- 4) Finalmente, valoraremos la eficacia de cada uno de los patrones que indicaban metonimia mediante un porcentaje de éxito. Este porcentaje lo calcularemos al comparar el número de aciertos el número total de entradas. Es decir, si del patrón anterior, *contien**, obtenemos un total de 295 concordancias y la relación semántica buscada la expresan 280 de ellas, podemos asegurar que dicho patrón resulta muy eficaz porque logramos nuestro objetivo en un 94,9 % de los casos.

3. RESULTADOS: LISTADO DE PATRONES LINGÜÍSTICOS

Como veremos más adelante, los resultados de este estudio los hemos presentado en forma de tabla-resumen que hemos dividido en cinco columnas. En la primera presentamos el listado de patrones que hemos obtenido durante la investigación. Cada patrón contiene un asterisco para que el WordSmith entienda que ese comodín permite también extraer las formas que varían en género y número. Además, los números que aparecen entre corchetes representan el número de palabras que se le indican a este mismo programa y que pueden ocurrir entre una parte y otra del patrón. En la segunda columna indicamos el número de entradas que obtenemos al introducir la búsqueda de cada patrón en el corpus TXTCeram. En la tercera y la cuarta mostramos la cantidad de entradas que ofrecen concordancias válidas (aciertos) y no válidas (ruido). Finalmente, en la última columna tenemos el porcentaje de éxito de cada patrón, que mide la eficacia de los patrones para la búsqueda de relaciones de metonimia en el corpus estudiado.

En la siguiente tabla se muestran los datos de los resultados obtenidos en el trabajo:

Patrones lingüísticos	Entradas	Aciertos	Ruido	Porcentaje de éxito
COMPUEST* [5] POR	46	19	27	41,3 %
CONFORM*	104	11	93	10,6 %
CONSIST* [5] EN	234	39	195	16,7 %
CONSTA* [5] DE*	80	49	31	61,3 %
CONSTITU* [5] POR	260	237	23	91,1 %
[5] CONSTRUID* [10] CON	13	7	6	53,8 %
CONTEN*	832 (100)	80	20	80 % aprox.
CONTIEN*	295	280	15	94,9 %
DIVID* [5] EN	87	26	61	29,9 %
EN [10] EXIST*	284 (100)	6	94	6 % aprox.
FORMAD* [5] POR	203	149	54	73,4 %

FORMANDO	162	58	104	35,8 %
FORM* [5] PARTE DE*	56	44	12	78,6 %
[5] PERTENEC* [10] A*	107	36	71	33,6 %
PRESENTE* [5] EN	163	120	43	73,6 %

Los resultados que hemos alcanzado siguiendo la metodología descrita en el apartado anterior pueden considerarse de alta fiabilidad. Tal aseveración puede afirmarse al tratarse de un estudio realizado con un número bastante elevado de entradas para cada candidato a patrón lingüístico. Además el porcentaje de éxito lo obtenemos mediante una operación matemática que establece la eficacia de dicho patrón. Por esta razón, nos ha llamado la atención el trabajo de Davidson (1998) en el que determina la productividad de un patrón a partir de un número demasiado bajo de entradas para cada uno de ellos. Por ejemplo, Davidson señala que *composed of* posee una eficacia del 0 % porque la única entrada obtenida para esta estructura resulta ser ruido. Sin embargo, en nuestra opinión, esta estructura no debe descartarse como posible patrón lingüístico que indique metonimia aunque su porcentaje de éxito no resulte ser de los más elevados.

A continuación, realizaremos un análisis más detallado de cada uno de los patrones estudiados. Además, mostraremos un ejemplo de los resultados que se pueden obtener con cada patrón lingüístico, así como también valoraremos su utilidad.

a) Compuest* [5] por

Ejemplo: «*El pequeño rosetón central está compuesto por ocho gajos volumétricos iguales, y su anillo perimetral por piezas complementarias curvadas.*»

El ruido que aparece al aplicar este patrón se debe a que a esta estructura léxica no sólo le corresponde el sentido que buscamos, es decir, al introducir nuestra búsqueda también obtenemos contextos para el sustantivo «compuesto [químico]». Así, este patrón resulta eficaz siempre y cuando logremos distinguir un significado y otro.

b) Conform*

Ejemplo: «*Compacidad de probetas conformadas a partir de arcillas de cuarzos de distinto tamaño de partícula.*»

A pesar de haber eliminado desde un principio mucho ruido activando la opción *but excluding: conformaci**, la eficacia de este patrón es muy baja ya que el ruido que presenta es enorme. El lema *conform** también extrae muchas otras palabras que nada tienen que ver con el verbo conformar —en el sentido de “dar forma a algo”—, como por ejemplo, conforme, conformidad, conformado, etc.

c) Consist* [5] en

Ejemplo: «*La materia orgánica dispersa en las arcillas naturales consiste principalmente en ácidos húmicos y carbón.*»

Incluso tras haber evitado muchos contextos inservibles con el parámetro *but excluding: consistencia**, nos encontramos con una gran mayoría de entradas en las que el patrón tiene el sentido de “estribar, estar fundada una cosa en otra”. Por eso, los aciertos que produce el patrón son despreciables en comparación con el ruido que produce.

d) Consta* [5] de*

Ejemplo: «*El equipo consta básicamente de un tubo emisor de rayos X, una cámara circular donde se sitúa la muestra (goniómetro) y un detector que recoge los rayos difractados.*»

En contradicción con el porcentaje de éxito de este patrón, creemos que los contextos que puede llegar a ofrecer son bastante ricos en metonimia. Después de haber incluso aplicado la opción *but excluding: constante**, la cantidad de ruido, es decir, el número total de contextos que no extraen ninguna de las formas «consta de» o «constan de» es muy elevado. Por eso, aconsejamos el uso de este patrón pero siempre intentando evitar cualquier otra estructura que no sea las mencionadas anteriormente.

e) Constitu* [5] por

Ejemplo: «*El compresor centrífugo está constituido por una serie de ruedas impulsoras montadas sobre un eje de acero y contenidas dentro de una carcasa de hierro.*»

En un principio optamos por probar este patrón con la forma *constitu**, para que abarcase tanto las estructuras encontradas con o sin preposición. Sin embargo, al echar un primer vistazo a los contextos que nos ofrecía, pronto nos dimos cuenta de que la mayoría de las veces en las que aparecía esta forma sin la preposición *por*, los resultados eran ruido. Por eso, decidimos realizar la búsqueda con el patrón «*constitu* por*» y descubrimos que este patrón es uno de los más eficaces para encontrar relaciones de metonimia entre conceptos.

f) [5] Construid* [10] con

Ejemplo: «*Las paredes construidas con ladrillos texturados no requieren ninguno de estos tratamientos.*»

Con este patrón hemos extraído muy pocos contextos como para poder decidir con más exactitud su grado de utilidad. Aún así, cabe destacar que, en proporción al número de entradas, los aciertos que ofrece son satisfactorios.

g) Conten*

Ejemplo: «*La primera evidencia que se tuvo del efecto de álcali mixto fue la de la depresión secular del cero observada en los termómetros construidos a partir de vidrios que contenían sodio y potasio.*»

En este caso se han comprobado las 100 primeras entradas ya que el volumen de resultados para este patrón es tan grande que se ha analizado únicamente una muestra representativa del total. Aún así el número de entradas se redujo en un principio al incluir también en la definición de la búsqueda el parámetro «*but excluding: contenedor**». Por tanto, llegamos a la conclusión de que el patrón produce bastantes aciertos aunque una gran cantidad de entradas.

h) Contien*

Ejemplo: «*Los cementos-cola contienen, además del aglomerante hidráulico, una fuerte proporción de caseína, cuyo poder autoalisante dificulta algo la colocación, y son ligeramente solubles en agua.*»

A la vista del porcentaje de éxito que conseguimos con este patrón, podemos afirmar que éste es, con diferencia, el patrón más eficaz de los estudiados en la presente investigación.

i) Divid* [5] en

Ejemplo: «*La estructura interna de la tierra se divide en corteza (continental u oceánica), manto (superior e inferior) y núcleo (externo e interno).*»

La ambigüedad de este patrón dificulta la extracción de contextos ricos en relación de metonimia ya que, a partir de «dividir en», también se pueden detectar resultados para la relación de hiponimia, como por ejemplo «Los equipos de corte de materiales cerámicos se pueden dividir en dos categorías: -cortadores manuales, -cortadores eléctricos de disco de diamante.». Esto ralentiza enormemente la

extracción automática de relaciones de metonimia para este patrón porque se tendría que analizar con más detenimiento los resultados lanzados por el WordSmith.

j) En [10] exist*

Ejemplo: «*Otro pirosilicato importante es la hemimorfita, de composición $[Zn_4(OH)_2](Si_2O_7)$, en el que existen tetraedros de composiciones ZnO_3OH y SiO_4 que comparten 3 vértices...*»

Con este patrón también se han analizado únicamente las 100 primeras entradas porque rápidamente nos dimos cuenta de que su efectividad era tan pequeña que no valía la pena continuar con el análisis de todas las concordancias.

k) Formad* [5] por

Ejemplo: «*Forjados unidireccionales de hormigón. Descripción y ámbito de aplicación. Son forjados formados por nervios, viguetas o semiviguetas de hormigón armado o pretensado dispuestos paralelamente.*»

Este patrón puede adoptar dos sentidos diferentes: 1) la enumeración de los componentes y 2) la causa de la formación. En el caso de que ofreciese información de la primera opción, los contextos contendrían relaciones de metonimia, en cambio, si se trata de la segunda opción, las entradas no serían válidas para la relación que nos ocupa. Luego, siempre que distingamos entre las dos acepciones, podemos utilizar el patrón en la búsqueda de metonimia.

l) Formando

Ejemplo: «*Según Plumet, el sulfuro sódico resultante reacciona con la sílice formando un sulfosilicato de estequiometría variable, de naturaleza vítrea y de color pardo oscuro.*»

La utilidad de esta palabra para la obtención de contextos ricos en relaciones de metonimia es de baja calidad. La razón es básicamente la misma que la del patrón anterior, *formad* por*.

m) Form* [5] parte de*

Ejemplo: «*Durante la combinación de ZrO_2 y SiO_2 , el ión vanadio entra a formar parte de la estructura cristalina para dar color.*»

Era de esperar que el grado de éxito del patrón fuera bastante elevado ya que ésta es la estructura modelo con la que indicar, de manera ideal, la relación de metonimia entre conceptos. Por consiguiente, aconsejamos el uso de este patrón lingüístico en la búsqueda de esta relación entre conceptos.

n) [5] Pertenece* [10] a*

Ejemplo: «*A temperaturas elevadas, la distribución Al-Si de la estructura se hace menos ordenada, denominándose saniclina, la cual, al igual que la ortosa, pertenece al sistema monoclinico.*»

Del mismo modo que ocurría más arriba con el patrón «*divid* [5] en*», la ambigüedad de este patrón con la relación de hiponimia reduce su efectividad al mismo tiempo que aumenta el riesgo de equivocación en el análisis de las relaciones. Por lo tanto, aconsejamos su uso de una manera más cuidadosa.

o) Present* [5] en

Ejemplo: «*La mullita se halla presente en todos los refractarios aluminosos y en muchos electrofundidos.*»

Los resultados positivos que se extraen con esta estructura son bastante elevados por lo que la efectividad de este patrón queda demostrada.

4. CONCLUSIONES

El objetivo principal de este trabajo era comprobar que los patrones lingüísticos son muy útiles en la búsqueda de relaciones semánticas y, en este caso, nos hemos centrado en la metonimia. Por eso, tras un análisis del corpus TXTCeram y con todos los datos derivados del trabajo, podemos concluir afirmando, por tanto, que los patrones lingüísticos permiten extraer contextos con información conceptual de términos, o lo que es lo mismo, contextos ricos en conocimiento.

Además, hemos elaborado un listado de patrones para la obtención automática de relaciones de metonimia entre conceptos entre los que señalamos los más eficaces: *contien**; *constitu** [5] *por*; *form** [5] *parte de**; *present** [5] *en*; y *formad** [5] *por*.

Sin embargo, esta investigación no concluye con este trabajo ya que el potencial que poseen los patrones en el trabajo terminológico todavía no se ha explorado lo suficiente. Por eso creemos interesante ampliar este campo de trabajo en futuras investigaciones en las que se partirá de una serie de términos relacionados semánticamente para obtener los patrones lingüísticos que los unen.

5. BIBLIOGRAFÍA

- BACH MARTORELL, CARMÉ (2005): «Los marcadores de reformulación como localizadores de zonas discursivas relevantes en el discurso especializado» en *Debate Terminológico*, n. 1. [En línea] http://www.riterm.net/revista/n_1/bach.pdf [Consulta: 31 de mayo de 2005]
- CHAFFIN, ROGER y DOUGLAS J. HERRMANN (1988): «The nature of semantic relations: a comparison of two approaches» en MARTHA EVENS (ed.) (1988): *Relational models of the lexicon. Representing knowledge in semantic networks*, Cambridge: Cambridge University Press.
- CONDAMINES, ANNE (2002): «Corpus analysis and conceptual relation patterns», *Terminology*, 8/1, 141-162.
- CONDAMINES, ANNE y JOSSETTE REBEYROLLE (2001): «Searching for and identifying conceptual relationships via corpus-based approach to a Terminological Knowledge Base (CTKB). Methods and results» en DIDIER BORIGAUULT, CHRISTIAN JACQUEMIN y MARIE-CLAUDE L'HOMME (eds.) (2001): *Recent advances in computational terminology*, Amsterdam: John Benjamins.
- DAVIDSON, LAURA (1998): *Knowledge Extraction Technology for Terminology*, Ottawa: University of Ottawa. [Tesis doctoral]
- FABER, PAMELA y CATALINA JIMÉNEZ (eds.) (2002): *Investigar en terminología*, Granada: Comares.
- FABER, PAMELA, CLARA INÉS LÓPEZ RODRÍGUEZ Y MARÍA ISABEL TERCEDOR SÁNCHEZ (2001): «Utilización de técnicas de corpus en la representación del conocimiento médico» en *Terminology*, 7/2, 167-198.
- FELIU CORTÉS, JUDIT (2004): *Relacions conceptuals i terminologia: Anàlisi i proposta de detecció semiautomàtica*, Barcelona: Universitat Pompeu Fabra. [Tesis doctoral]

- IRIS, MADELYN ANNE, BONNIE E. LITOWITZ y MARTHA EVENS (1988): «Problems of the part-whole relation» en MARTHA EVENS (ed.) (1988): *Relational models of the lexicon. Representing knowledge in semantic networks*, Cambridge: Cambridge University Press.
- MARSHMAN, ELIZABETH, TRICIA MORGAN e INGRID MEYER (2002): «French patterns for expressing concept relations», *Terminology*, 8/1, 1-29.
- MARTÍ ANTONÍN, MARÍA ANTONIA (coord.) (2003): *Tecnologías del lenguaje*, Barcelona: Editorial UOC.
- MEYER, INGRID (2001): «Extracting knowledge-rich contexts for terminography. A conceptual and methodological framework» en BORIGAUULT, DIDIER, CHRISTIAN JACQUEMIN y MARIE-CLAUDE L'HOMME (eds.) (2001): *Recent advances in computational terminology*, Amsterdam: John Benjamins.
- MORGAN, TRICIA (2000): *A Comparative Study of Hypernymic Patterns for Knowledge Extraction*, Ottawa: University of Ottawa. [Tesis doctoral]
- OTMAN, GABRIEL (1996): *Les représentations sémantiques en terminologie*, París : Masson.
- PEARSON, JENNIFER (1998): *Terms in Context*, Amsterdam: John Benjamins
- PÉREZ HERNÁNDEZ, M. CHANTAL (2002): «Terminografía basada en corpus: principios teóricos y metodológicos» en FABER, PAMELA y CATALINA JIMÉNEZ (eds.) (2002): *Investigar en terminología*, Granada: Comares.
- SÁNCHEZ GIJÓN, PILAR (2003): *Els documents digitals especialitzats: utilització de la lingüística de corpus com a front de recursos per a la traducció*, Barcelona: Universitat Autònoma de Barcelona. [Tesis doctoral]
- SINCLAIR, JOHN (1991): *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

ANEXO I

Fuentes utilizadas para el corpus

- ALARCÓN, JAVIER, JUAN CARDA y JESÚS MARÍA RINCÓN (1992): *Nuevos productos y tecnologías de esmaltes y pigmentos cerámicos. Su fabricación y utilización*, Castellón: Faenza Editrice Ibérica, S.L.
- AMORÓS ALBERO, J. L., V. BARBA y V. BELTRÁN PORCAR (1994): *Estructuras cristalinas de los silicatos y óxidos de las materias primas cerámicas*, Castellón: Instituto de Tecnología Cerámica-Asociación de Investigación de las Industrias Cerámicas.
- AMORÓS ALBERO, J. L., V. BELTRÁN PORCAR, A. BLASCO FUENTES, J. E. ENRIQUE NAVARRO, AGUSTÍN ESCARDINO BENLLOCH e I. NEGRE MEDALL (1991): *Defectos de fabricación de pavimentos y revestimientos*, Castellón: Asociación de Investigación de las Industrias Cerámicas.
- BARBA, ANTONIO, VICENTE BELTRÁN, CARLOS FELIU, JAVIER GARCÍA, FERNANDO GINÉS, ENRIQUE SÁNCHEZ y VICENTE SANZ (1997): *Materias primas para la fabricación de soportes de baldosas cerámicas*, Castellón: Instituto de Tecnología Cerámica.
- ENRIQUE NAVARRO, J. E., F. NEGRE, V. BELTRÁN PORCAR y J. GUILLEM (1980): *Mayólica, gres y porcelana artística*, Castellón: Asociación de Investigación de las Industrias Cerámicas.
- ESCARDINO BENLLOCH, AGUSTÍN y MANUEL GONZÁLEZ CUDILLEIRO (1991): *Azulejos y pavimentos cerámicos españoles*, Ministerio de Industria, Comercio y Turismo.
- FERNÁNDEZ NAVARRO, JOSE MARÍA (1991): *El vidrio*, Madrid: Consejo Superior de Investigaciones Científicas.
- GARCÍA VERDUCH, ANTONIO (1993): *Colocación de pavimentos y revestimientos cerámicos*, Castellón: Instituto de Tecnología Cerámica.
- GIOVANNINI, R. (1989): *La serigrafía en la cerámica*, Barcelona: Omega, S.A.
- PORCAR, JOSE LUIS, ALFONSO PLEGUEZUELO, FERRAN RENAU y JOSE MARÍA GOMIS (1987): *Manual-guía técnica de los revestimientos y pavimentos cerámicos*, Castellón: Instituto de Tecnología Cerámica.
- TORRELLA ALCARAZ, ENRIQUE (1996): *La producción de frío*, Valencia: Universidad Politécnica de Valencia, Servicio de Publicaciones.

ANEXO II

Listado de estructuras que implican relación de metonimia y posibles candidatos a patrones lingüísticos

contienen, contienen a, contienen, contiene, contenidas en, contenga, contener, contengan	CONTEN* CONTIEN*
presente en, presentes en	PRESENTE*... EN
constituido por, constituidas por, constituidos por, constituida por, constituyen	CONSTITU*... POR
formada por, formando, formados por, formado por	FORMANDO FORMAD*... POR
construida con	CONSTRUID*... CON
consisten en	CONSIST*... EN
forman parte de	FORM*... PARTE DE*
consta de	CONSTA*... DE*
conforman	CONFORM*
compuestas... por	COMPUEST*... POR
dividida en	DIVID*... EN
pertenecientes al, pertenecientes a, pertenecen a	PERTENEC*... A*
en... existen	EN... EXIST*

ANOTACIONES

a Este trabajo ha sido realizado gracias a la financiación de la Generalitat Valenciana (GV05/260) y la Unidad de Apoyo Educativo de la Universitat Jaume I, en el marco de los proyectos TXTCeram y CREC, dirigidos por la Dra. Amparo Alcina. El corpus ha sido elaborado gracias a la colaboración de los estudiantes de Terminología e Informática aplicada a la Traducción de la licenciatura en Traducción e Interpretación.