



Jornades de Foment de la Investigació

LOS MARCADORES LINGÜÍSTICOS EN LAS DEFINICIONES DEL GRUPO CONCEPTUAL “PROCESOS DE FABRICACIÓN CERÁMICA”

Autor

Esperanza VALERO.

**Los marcadores lingüísticos en las definiciones del grupo conceptual
“procesos de fabricación cerámica”**

Esperanza Valero Doménech
Equipo TecnoLeTTra
Universitat Jaume I

INTRODUCCIÓN¹

Son numerosos los estudios enfocados a extraer información de las definiciones de diccionarios existentes. La mayor parte de ellos van dirigidos a extraer relaciones taxonómicas (Amsler, 1980; Calzolari, 1984). Sin embargo, son necesarios más trabajos encaminados a aprovechar el resto de información presente en la definición, como las características específicas y las relaciones conceptuales no jerárquicas (Alshawi, 1987; Calzolari, 1992; Barnbrook, 2002). Para extraer esta información el mayor problema al que se enfrentan los sistemas automáticos es la falta de homogeneidad y sistematicidad en las definiciones. Gran parte de los diccionarios presentan información inconsistente e incompleta y términos que deberían recibir un tratamiento homogéneo son definidos de manera muy desigual (Boguraev y Briscoe, 1989). Muchos autores destacan la necesidad de elaborar definiciones más estandarizadas, precisas y en un formato comprensible para el ordenador y los humanos, de manera que facilite la extracción y reutilización de la información (Barque y Polguère, 2009; Barnbrook, 2002).

Uno de los objetivos de nuestro proyecto es desarrollar un sistema que asista al terminógrafo/a en la elaboración de definiciones. Este sistema generará de manera semi-automática definiciones a partir un patrón de definiciones y una ontología de domino. El patrón contendrá los marcadores lingüísticos necesarios para introducir cada característica en la definición. Un ejemplo de este patrón de definiciones es el propuesto por Alcina (2009) para el grupo conceptual “baldosa cerámica”:

A: baldosa cerámica de forma x y de tamaño y que está decorada con z y tiene como función q .

Donde A corresponde al concepto objeto de definición y las variables X, Y, Z son variables que se sustituirían por los valores que tomarían los atributos correspondientes a forma, tamaño, decoración y función.

En este trabajo se han analizado manualmente 135 definiciones de procesos de fabricación cerámica extraídas de dos diccionarios especializados. El objetivo de este análisis es observar los tipos de característica que son relevantes para la definición de algunos grupos conceptuales de este ámbito y estudiar cómo se expresan lingüísticamente en la definición. En definitiva, hemos estudiado los marcadores lingüísticos utilizados en las definiciones para introducir cada tipo de característica.

¹ Esta investigación se enmarca en el proyecto “ONTODIC: Metodología y tecnologías para la elaboración de diccionarios onomasiológicos basados en ontologías. Recursos terminológicos para la e-traducción”, dirigido por Amparo Alcina, Universitat Jaume I, y financiado por el Ministerio de Educación y Ciencia de España (TSI2006-01911).

Este conjunto de características y marcadores serán de utilidad, por un lado, para la extracción de información de los diccionarios y, por otro lado, para la representación de información semántica en un diccionario, es decir para la generación de definiciones.

EXTRACCIÓN DE LA INFORMACIÓN SEMÁNTICA DE LAS DEFINICIONES DE DICCIONARIOS

La extracción de información semántica de contextos definitorios es una importante área de investigación de la lexicografía y la terminología. Sin embargo, los estudios existentes se centran sobre todo en el inglés y el francés, existiendo una carencia en español (Soler y Alcina, 2008; Sierra et al, 2008). Los autores coinciden en que una buena manera de extraer esta información e los textos es mediante la búsqueda de patrones recurrentes. De acuerdo con Marshman et al (2002:2), los *knowledge patterns* son “words, word combinations or paralinguistic features which frequently indicate conceptual relations. Las autoras distinguen entre tres tipos de patrones:

- Patrones léxicos, que son palabras y sus variantes morfológicas que indican una relación, por ejemplo, *is a* indica hiperonimia.
- Patrones gramaticales, son combinaciones de categorías gramaticales. El ejemplo propuesto por las autoras es para el patrón de función que puede ser *sustantivo + verbo* como se ve en la frase: *L'unité centrale effectue le traitement*.
- Patrones paralingüísticos, en estos se incluye la puntuación, paréntesis y estructura del texto, que también puede ser indicativa de una relación. Por ejemplo, los paréntesis a veces indican la relación de hiperonimia.

En este trabajo, en lugar de contextos ricos en conocimiento extraídos de un corpus, nos centramos en las definiciones. Según Calzolari (1984) “las definiciones utilizan un sublenguaje del lenguaje natural”. El objetivo de este estudio es encontrar los patrones que se utilizan en las definiciones para denotar las características y relaciones semánticas.

Las definiciones de los diccionarios poseen la estructura semántica del lenguaje más explícita y accesible para su análisis (Wilks et al., 1989: 196).

Conceptualmente la definición convencional tiene dos componentes: el *genus*, que refleja en la definición la organización conceptual jerárquica del campo de especialidad; y las *differentiae*, que sirven para distinguir el concepto definido de sus co-hipónimos y reflejan las relaciones que tienen lugar en el campo de especialidad.

Wilks et al. (1989:227) reconocen que para la semántica computacional extraer los términos genéricos de las definiciones resulta menos difícil que extraer las características específicas. Como afirma Barriere (1997) en la bibliografía no se puede encontrar una lista completa de la información semántica que se puede extraer de las definiciones. También Sager y L’Homme (1994: 366) consideran que esta parte de la definición todavía no ha sido susceptible de codificación siguiendo un conjunto restringido de tipos de características. De acuerdo con ellos, éste sería un paso importante para la sistematización de las definiciones terminográficas.

METODOLOGÍA DEL ESTUDIO EMPÍRICO

Selección de definiciones

Hemos analizado un total de 135 definiciones procedentes de dos diccionarios del ámbito de la cerámica: *Diccionario científico-práctico de la cerámica* y *Diccionario de cerámica*. Estos diccionarios fueron publicados en formato papel y han sido digitalizados para el proyecto ONTODIC (Alcina et al, 2007). Hemos extraído de los diccionarios las definiciones de conceptos englobados en la categoría de “procesos de fabricación cerámica”. El objetivo es identificar un conjunto de características relevantes para la descripción de un grupo conceptual completo.

Identificación de las características conceptuales en la definición

En este trabajo, hemos analizado las definiciones para obtener una serie de características para dos categorías de este dominio. Se ha seguido la propuesta Meyer et al (1997:102) que distingue el nombre de la característica y su valor. El nombre de la característica funciona como etiqueta indicadora del contenido de la misma y el valor ofrece información específica de un concepto. Para el concepto abrasión, causa sería la etiqueta y fricción su valor.

El análisis se ha realizado segmentando la información contenida en la definición y asignando una etiqueta o código que describe el tipo de información que cada fragmento representa, como se puede ver en la figura 1. De esta manera, hemos observado que cada grupo conceptual comparte una serie de características.

Figura 1. Identificación de las características conceptuales

Tamizado. Acto en que se lleva a cabo el paso del material a través del tamiz [PROCEDIMIENTO]. Aerografía. Método de aplicación de capas de esmalte o de engobe [OBJETIVO] mediante un aerógrafo [INSTRUMENTO].

Los marcadores lingüísticos en las definiciones del grupo conceptual “procesos de fabricación cerámica”

Para realizar este análisis hemos utilizado el programa de análisis cualitativo Atlas.ti. Este programa nos permite segmentar la información, asignar un código descriptivo a cada segmento, crear relaciones entre ellos o entre sus códigos, obtener representaciones gráficas de la estructura conceptual de los datos y realizar búsquedas textuales o por los códigos asignados como en una base de datos.

El resultado de esta fase es un listado de características esenciales para la descripción de cada grupo conceptual y un conjunto de valores para cada una de estas características.

Las principales características identificadas en las definiciones se pueden ver en la tabla 1, donde se indica también la frecuencia de aparición en el corpus. Los ejemplos de los valores que pueden adoptar estas características están reflejados en el anexo 1.

Tabla 1. Características de los procesos de fabricación

Característica	frecuencia	Característica	frecuencia
PROCEDIMIENTO	69	INSTRUMENTOS	22
OBJETIVO	103	ESTADO DEL MATERIAL	15
PACIENTE	56	FASE POSTERIOR	8
FASE ANTERIOR	6		

Identificación de los marcadores lingüísticos

En esta fase hemos anotado la estructura lingüística o marcador que precede a cada característica. En algunas definiciones es posible identificar una serie de patrones léxicos que introducen un tipo de característica. Como se ve en la figura 2, la característica OBJETIVO va introducida en la primera definición por el marcador lingüístico *que sirve para* y en la segunda, *con el fin de*.

Figura 2. Ejemplos de marcadores lingüísticos en la definición

cocción de decoración: cocción que sirve para madurar los efectos decorativos aplicados previamente a las pieza [OBJETIVO]
compactación de polvos. Operación de condensar los materiales pulverulentos con el fin de obtener productos con la mayor densidad posible [OBJETIVO]
cocción lenta. cocción que se desarrolla a una velocidad más pequeña de la acostumbrada. [PROCEDIMIENTO]

Sin embargo, en muchos casos no ha sido posible identificar un marcador, sino estructuras sintácticas recurrentes en la expresión de las características. Como se puede observar en la figura 3, la característica PROCEDIMIENTO está expresada en ambas definiciones mediante una oración en la que el verbo principal está en gerundio.

Los marcadores lingüísticos en las definiciones del grupo conceptual “procesos de fabricación cerámica”

Figura 3. Estructuras sintácticas en las definiciones

colar. Formar una pieza vertiendo un líquido o una masa plástica en un molde [PROCEDIMIENTO], por fraguado o enfriamiento
 aclarar, diluir (un color). Rebajar la intensidad de un color añadiéndole componente blanco o agente blanqueante [PROCEDIMIENTO].

RESULTADOS

Los resultados obtenidos en este análisis consisten en una serie de marcadores lingüísticos y sintácticos que denotan los diferentes tipos de características en las definiciones de este grupo conceptual. Los marcadores lingüísticos preceden a la característica, un ejemplo de ello sería “con el fin de” para describir el OBJETIVO de un proceso. Por otro lado, los patrones sintácticos constituyen la estructura mediante la cual se expresa al característica, por ejemplo, la característica PROCEDIMIENTO aparece muchas veces expresada sin un marcador lingüístico que la introduzca, pero con un patrón sintáctico recurrente que es una oración cuyo verbo está en gerundio. La tabla 2 muestra los marcadores lingüísticos detectados para cada característica y la tabla 3, los patrones sintácticos recurrentes.

Tabla 2. Marcadores lingüísticos

Marcador lingüístico	Ejemplos del corpus
OBJETIVO	
<i>para +inf</i> (15)	para dotarle de las propiedades adecuadas.
<i>consistent* en</i> (2)	consistente en eliminar de las piezas moldeadas las rebabas y otras protuberancias
<i>que serv* para</i> (1)	que sirve para madurar los efectos decorativos
PROCEDIMIENTO	
<i>dirigid*a</i> (1)r	Acción dirigida a que un esmalte que da superficies brillantes se transforme...
<i>por+SN</i> (14)	por inmersión de la pieza en un baño de esmalte
<i>Mediante</i> (3)	Mediante su inmersión en un baño de esmalte
<i>que se logra</i> (1)	Que se logra pasando la mezcla a través del medio adecuado.
<i>se llev*a cabo</i> (1)	se lleva a cabo el paso del material a través del tamiz.
FASE PREVIA	
<i>se desarroll*</i>	que se desarrolla a una velocidad más pequeña de la acostumbrada.
<i>después de</i> (3)	después de haber sido esmaltada

**Los marcadores lingüísticos en las definiciones del grupo conceptual
“procesos de fabricación cerámica”**

FASE POSTERIOR	
previa/o (1)	previa homogeneización
<i>antes de</i> (3)	antes de ser esmaltada
INSTRUMENTOS	
mediante (5)	mediante un aerógrafo.
con +SN (2)	Con piedra o herramienta de acero
en +SN (2)	en un molde
por +SN (2)	Por prensas mecánicas
<i>empleando</i> (1)	empleando molinos de rodillos, bolas o guijarros
<i>haciendo uso de</i>	haciendo uso de un dispositivo (pistola)
<i>se utiliz*</i> (1)	se utilizan resistencias eléctricas
<i>con ayuda de</i> (1)	con ayuda de un calibre o plantilla
<i>mediante el uso de</i> (1)	mediante el uso de resinas sintéticas de intercambio iónico.
por medio de (1)	por medio de un pistón
se requiere/en (1)	Se requieren tres tipos de reactivo: espumantes, colectores y controladores
dentro de (1)	dentro del molde
PACIENTE	
de+ SN (11)	de unas materias primas
en+SN (6)	en una pasta o barbotina
a+SN (5)	al material
se aplica/n a (1)	se aplica en el esmalte
ESTADO DEL MATERIAL	
en estado (1)	en estado plástico
en forma de (1)	en forma de barbotina.
en +SN (1)	en polvo

Los marcadores lingüísticos en las definiciones del grupo conceptual “procesos de fabricación cerámica”

Tabla 3. Estructuras sintácticas recurrentes

Patrón sintáctico	Ejemplos del corpus
OBJETIVO	
<i>oración de infinitivo</i> (25)	aplicar un color sobre una superficie
<i>oración de infinitivo</i> (25)	aplicar un color sobre una superficie
PROCEDIMIENTO	
<i>oración simple</i> (23)	este se sumerge en la barbotina y se escurre
<i>infinitivo</i> (11)	someter un material a la acción del calor a temperaturas altas.
<i>gerundio</i> (10)	añadiéndole componente blanco o agente blanqueante.
<i>SN</i> (3)	colocación de piezas unas sobre otras
ESTADO DEL MATERIAL	
<i>adjetivo</i> (10)	bizcochado

DISCUSIÓN

Hemos analizado 135 definiciones pertenecientes al grupo conceptual “procesos de fabricación cerámica”. En estas definiciones hemos detectado siete características conceptuales: OBJETIVO, PROCEDIMIENTO, PACIENTE, ESTADO DEL MATERIAL, INSTRUMENTOS, FASE PREVIA y FASE POSTERIOR. Hemos identificado una serie de marcadores y patrones sintácticos que preceden a estas características en las definiciones. Los resultados muestran algunos problemas relativos a la diversidad en la expresión lingüística de estas características, la ausencia de un *genus* común para todos los conceptos de un mismo grupo y otros problemas identificados en estudios previos ((Marshman et al, 2002:2 y Barrière, 2002) como son la polisemia de los patrones, la variedad morfosintáctica y la no contigüidad de los elementos del patrón. A continuación explicamos los problemas detectados en la expresión de las características de los procesos de fabricación cerámica:

Variedad de patrones sintácticos y marcadores léxicos

Tanto a nivel conceptual como a nivel lingüístico se puede ver que estas definiciones son muy heterogéneas. Conceptualmente podemos ver que dos conceptos similares se describen por medio de características diferentes. En una definición de un proceso de fabricación se expresa el PROCEDIMIENTO, los INSTRUMENTOS y el OBJETIVO, mientras que otro concepto del mismo grupo se define únicamente mediante las características de PROCEDIMIENTO y FASE ANTERIOR.

Los marcadores lingüísticos en las definiciones del grupo conceptual “procesos de fabricación cerámica”

A nivel lingüístico, podemos encontrar una amplia gama de marcadores y estructuras para expresar la misma característica en las diferentes definiciones. Por ejemplo, la característica PROCEDIMIENTO aparece 10 veces expresada por una oración cuyo verbo principal está en gerundio, pero en 23 casos esta característica constituye una oración principal de la que no se puede extraer ningún patrón común en todos los casos. Esto se puede ver en los siguientes ejemplos: 1. *este se sumerge en la barbotina y se escurre*; 2. *el esmalte se proyecta sobre la pieza cerámica*; 3. *el producto bizcochado se sumerge en una suspensión de los ingredientes del esmalte en agua*; 4. *las partículas que han de ser pulverizadas se les da una carga electrostática opuesta a la de la pieza a esmaltar; esta atrae a las partículas hacia la pieza*. En estos casos la única vía para encontrar regularidades sería la anotación semántica, por ejemplo, de los verbos utilizados en las características. Como se puede ver en los ejemplos, muchos de los verbos principales manifiestan una proximidad semántica y todos se podrían considerar verbos de acción. Por ejemplo, *sumergir, escurrir, proyectar o pulverizar*.

Ausencia de descriptores iniciales en las definiciones.

Otro factor que afecta a la expresión lingüística de las características es la estructura de la misma. El modelo estándar de definición que contiene el concepto superordenado y características específicas no siempre se respeta. En ocasiones, en la definición no aparece un descriptor inicial sino que se introduce en primera posición del *definiens* una característica específica. Por ejemplo, la característica OBJETIVO en la descripción de los procesos de fabricación ocupa el primer lugar en las definiciones, por lo tanto no es posible detectar ningún marcador lingüístico que la preceda.

- Allanado. Eliminación del material de apoyo de la pieza después del bizcochado o de la cocción de brillo.
- Amasar. Formar o hacer una masa o pasta, mezclando yeso, tierra o una materia pulverulenta con agua u otro líquido.

Variedad morfosintáctica de los marcadores en las definiciones,

En cuanto a la variedad morfosintáctica, cabe destacar el caso de los patrones verbales, ya que éstos pueden aparecer en cualquiera de sus formas conjugadas. Es el caso del patrón *utilizar*, que en el corpus aparece en diferentes formas:

- *Usualmente utilizada como*
- *Se utilizan...*
- *Utilizar*

Los resultados de este análisis muestran que las definiciones de los diccionarios carecen de uniformidad y que algunos de los marcadores lingüísticos identificados en la expresión de las características ofrecen algunos problemas de cara a la extracción automática de información de estas definiciones.

CONCLUSIÓN

En este artículo hemos descrito la expresión lingüística de las características incluidas en las definiciones del grupo conceptual “procesos de fabricación cerámica”. El objetivo de este estudio es obtener una serie de marcadores lingüísticos o patrones sintácticos. Sin embargo, los resultados muestran una enorme heterogeneidad en la realización lingüística de estas características.

Nuestro proyecto tiene como objetivo diseñar una herramienta de terminografía asistida que ayude al terminógrafo/a a elaborar las definiciones de conceptos especializados. Estas definiciones, basadas en patrones de definición, serán completas y coherentes con todo el sistema de definiciones y contendrán la información de manera estructurada. Uno de los elementos del patrón de definiciones consiste en un conjunto de marcadores lingüísticos introductorios de las diferentes características en la definición. Este análisis nos ha ayudado a observar los marcadores que ocurren de manera natural en el texto de las definiciones. En futuros trabajos nos planteamos seleccionar uno o varios de estos marcadores para cada característica e introducirlos en el patrón de este grupo conceptual.

REFERENCIAS

- ALCINA A. (2009): «Metodología y tecnologías para la elaboración de diccionarios terminológicos onomasiológicos», en ALCINA A., E. VALERO Y E. RAMBLA (eds.): *Terminología y sociedad del conocimiento*, Peter Lang, Berna.
- ALCINA, A. Y E. VALERO (2008): «Análisis de las definiciones del diccionario cerámico científico-práctico. Sugerencias para la elaboración de patrones de definición», *Debate Terminológico*, 4.
- ALCINA, A., V. SOLER Y J. GRANELL (2007): «Translation technology skills acquisition», *Perspectives, Studies in Translatology*, 15 (4), 2007.
- ALSHAWI, H (1989): «Analysing the dictionary definitions», en Boguraev, B. y T. Briscoe
- AMSLER, R.A. (1980): *The structure of the Merriam-Webster pocket dictionary*, The University of Texas
- BARQUE L. Y A. POLGUÈRE (2009): «Structuration et balisage sémantique des définitions du *Trésor de Langue Française informatisé (TLFi)*», *Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'09)*, Montreal.
- BARNBROOK, G. (2002): *Defining Language. A local grammar of definition sentences*, John Benjamins, Amsterdam.
- BARRIÈRE, C. (1997): *From a Children's First Dictionary to a Lexical Knowledge Base of Conceptual Graphs*, Ph.D. Thesis, Ottawa University.
- BARRIÈRE, C. (2002): «Investigating the Causal Relation in Informative Texts», *Terminology*, 7(2).

- BOGURAEV B. Y T. BRISCOE (eds.) (1989): *Computational Lexicography for Natural Language Processing*, Longman, London.
- CALZOLARI, N. (1992): «Acquiring and representing semantic information in a Lexical Knowledge Base», *Proceedings of the ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation*, Springer-Verlag.
- CALZOLARI, N. (1984): «Detecting patterns in a Lexical Data Base», *Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics*, Stanford.
- CHITI, J. F. (1984): *Diccionario de cerámica*. Buenos Aires, Condorhuasi.
- GUILLEM, C. Y M. C. GUILLEM (1987) : *Diccionario cerámico científico-práctico (español-inglés-alemán-francés)*, Castellón, Sociedad española de cerámica y vidrio..
- MALAISÉ, V., P. ZWEIGENBAUM Y B. BACHIMONT (2005): «Mining defining contexts to help structuring differential ontologies», *Terminology*, 11 (1), 21–53..
- MARSHMAN, E., T. MORGAN E I. MEYER (2002): «French patterns for expressing concept relations», *Terminology*, 8 (1).
- MEYER, I., E. KAREN Y D. SKUCE (1997): «Systematic concept Analysis within a Knowledge-Based Approach to Terminology», en S. E. Wright y G. Budin (eds.). *Handbook of Terminology Management*, John Benjamins, Philadelphia, 98-118..
- SAGER J. Y M.C. L'HOMME (1994): «A model for the definition of concepts: rules for analytical definitions in terminological databases», *Terminology*, 1(2), 351- 373.
- SIERRA, G., R. ALARCÓN, C. AGUILAR Y C. BACH (2008): «Definitional verbal patterns for semantic relation extraction», *Terminology*, 14(1), 74-98
- SOLER, V. Y A. ALCINA (2008): «Patrones léxicos para la extracción de conceptos vinculados por la relación parte-todo en español», *Terminology*, 14(1).
- WILKS Y., D. FASS, C. GUO, J. McDONALD, T. PLATE Y B. SLATOR (1989): A tractable machine dictionary as a resource for computational semantics, en B. BOGURAEV Y T. BRISCOE (eds.): *Computational Lexicography for Natural Language Processing*, Longman, London.

Los marcadores lingüísticos en las definiciones del grupo conceptual “procesos de fabricación cerámica”

ANEXO 1.

Ejemplos de valores de características encontrados en el corpus

Tabla 4. Características de los procesos de fabricación

Característica específica	Ejemplos en las definiciones
<i>PROCEDIMIENTO</i>	proyección de una barbotina de pasta cerámica en forma de lluvia finísima por la sacudida de una explosión dentro del molde por inmersión de la pieza en un baño de esmalte.
<i>OBJETIVO</i>	para cubrirla de una capa adecuada rebajar la intensidad de un color quitarle la humedad a los objetos
<i>PRODUCTO RECEPTOR</i>	esmalte, producto cerámico, pieza
<i>ESTADO DEL PRODUCTO</i>	cruda, bizcochada, esmaltada
<i>INSTRUMENTOS</i>	empleando molinos de rodillos, bolas o guijarros haciendo use de un dispositivo (pistola) mediante un aerógrafo.
<i>MÉTODO</i>	por vía seca por vía húmeda aplicados mediante compresor
<i>FASE PREVIA</i>	después del bizcochado, previa homogeneización, que se efectúa después del prensado
<i>FASE POSTERIOR</i>	para proceder a su secado para su almacenamiento. antes de la aplicación del barniz o esmalte.