Los avances científicos y tecnológicos de las últimas décadas junto con la globalización de los mercados implican un aumento en la comunicación multilingüe. El uso de las tecnologías es imprescindible para el desarrollo de los recursos y diccionarios terminológicos. A su vez las tecnologías requieren el desarrollo de estos recursos con vistas a diversas aplicaciones, como la web semántica, la traducción automática o la extracción automática del conocimiento monolingüe y multilingüe.

Este libro presenta un panorama de las necesidades que surgen en torno al lenguaje y la terminología en relación con la nueva Sociedad del conocimiento, para después adentrarse en las técnicas de procesamiento automático del lenguaje y los métodos para extraer y representar el conocimiento.

Los distintos temas son abordados por expertos de reconocido prestigio en Lingüística Computacional, Procesamiento del lenguaje natural, Terminología, Traducción, Ingeniería del conocimiento y Ciencias de la computación. Los profesionales y universitarios de estos ámbitos encontrarán un panorama actual sobre el desarrollo de estas disciplinas así como ideas sugerentes sobre su futuro en relación con las tecnologías.

Amparo Alcina es Profesora Titular en la Universitat Jaume I. Dirige el Máster Oficial TECNOLOC: Tecnologías de la traducción y Localización. Dirige el grupo de investigación TecnoLeTTra: Tecnologías del lenguaje, la terminología y la traducción. Ha publicado numerosos artículos, algunos en revistas tan prestigiosas como Meta, Target y Terminology. Actualmente su investigación se centra en el desarrollo de ontologías y diccionarios onomasiológicos.

Esperanza Valero colabora en el grupo de investigación TecnoLeTTra y prepara su tesis doctoral sobre la definición terminológica y su elaboración semiautomática.

Elena Rambla colabora en el grupo de investigación TecnoLeTTra y prepara su tesis doctoral sobre la formalización e implementación de la relación conceptual hiperónimo-hipónimo.

ISBN 978-3-03911-593-8

9 783039 115938

Peter Lang

Terminología y Sociedad del conocimiento

Terminología y Sociedad del conocimiento

Bibliografische Information Der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über ‹http://dnb.d-nb.de› abrufbar.

Cover picture: Sergio Ballesteros
Cover design: Eva Rolli, Peter Lang Bern

ISBN 978-3-03911-593-8

© Peter Lang AG, Internationaler Verlag der Wissenschaften, Bern 2009
Hochfeldstrasse 32, CH-3012 Bern
info@peterlang.com, www.peterlang.com, www.peterlang.net

All rights reserved.
All parts of this publication are protected by copyright.
Any utilisation outside the strict limits of the copyright law, without the permission of the publisher, is forbidden and liable to prosecution.
This applies in particular to reproductions, translations, microfilming, and storage and processing in electronic retrieval systems.

Printed in Germany

Índice

Introducción: Perspectivas sobre Terminología y Sociedad del Conocimiento 7
AMPARO ALCINA

Necesidades y aplicaciones de la Terminología en la Sociedad del conocimiento

Retos en la interacción de la recuperación de información y el procesamiento del lenguaje natural 15
RICARDO BAEZA-YATES

Metodología y tecnologías para la elaboración de diccionarios terminológicos onomasiológicos 33
AMPARO ALCINA

Terminología y traducción automática 59
CELIA RICO PÉREZ

La formación terminológica y en las tecnologías del lenguaje: un valor añadido de la actividad profesional de traductores e intérpretes 77
PURIFICACIÓN FERNÁNDEZ NISTAL

Procesamiento automático del lenguaje y terminología: extracción de términos y anotación sintáctica y semántica de corpus

Panorama actual de la ingeniería lingüística 99
ANTONIO MORENO SANDOVAL

Los extractores de terminología: logros y escollos 117
ROSA ESTOPÀ BAGOT

Técnicas cuantitativas para la extracción de términos en un corpus 147
JORDI PORTA ZAMORANO

Anotación morfo-sintáctica y semántica de corpus. Adquisición y uso 179
HORACIO RODRÍGUEZ

Una visión interdisciplinar de la anotación semántica 219
GUADALUPE AGUADO DE CEA, INMACULADA ÁLVAREZ DE MON Y REGO Y ANTONIO PAREJA LORA

Ontologías en Informática 255
ISMAEL SANZ Y ERNESTO JIMÉNEZ-RUIZ

Técnicas y estrategias para la extracción y representación del conocimiento

Terminología multilingüe y ontologías 289
PAMELA FABER BENÍTEZ

FrameNet Español: un análisis cognitivo del léxico del español 309
CARLOS SUBIRATS RÜGGEBERG

Terminología y estructura del conocimiento: métodos y técnicas de visualización 321
ROSA CASTRO PRIETO

Extracción y representación de conocimiento a partir de corpus 341
CLARA INÉS LÓPEZ RODRÍGUEZ

Extracción de información conceptual basada en ontologías 375
JORDI VIVALDI PALATRESI

Notas sobre los autores 405

Introducción: Perspectivas sobre Terminología y Sociedad del Conocimiento[1]

AMPARO ALCINA

La idea de este libro surgió a partir de los cursos de verano *Terminología y Sociedad del conocimiento* y *Terminología e Ingeniería lingüística*, que organicé en la Universitat Jaume I de Castellón en 2004 y 2006. Los cursos pretendían dar a conocer los nuevos avances y perspectivas tecnológicas que podían tener un impacto en Terminología, con el objetivo de contribuir a la formación de los estudiantes que entonces comenzaban a colaborar y formarse en el grupo de investigación de Tecnologías del Lenguaje, la Terminología y la Traducción, Tecnolettra, de esta Universidad. A estos cursos asistieron también estudiantes procedentes de distintas Universidades que mostraron su interés y entusiasmo por los objetivos y las líneas que allí se expusieron, participando activamente en los debates y proponiendo nuevas ideas. Esto me hizo ver, en gran medida, que el interés científico y formativo en este ámbito interdisciplinar iba más allá de los propios cursos y de los proyectos que estábamos gestando en ese momento en nuestro grupo (proyectos TxtCeram y ONTODIC), y que había una necesidad de formación, por una parte, y un vacío editorial, por otra, que era necesario satisfacer.

El propósito del volumen es, pues, reunir algunas claves que nos lleven a vislumbrar mejor cuál es o cuál puede ser el papel de la Termino-

1 Esta publicación ha sido posible gracias a la financiación recibida del Ministerio de Educación y Ciencia y Generalitat Valenciana para los proyectos *ONTODIC: Metodología y tecnologías para la elaboración de diccionarios onomasiológicos basados en ontologías. Recursos terminológicos para la e-traducción* (TSI2006-01911) y *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* (GV05/260).

— (1994): "Natural Language Processing: she needs something old and something new (maybe something borrowed and something blue, too)", *Presidential address, Association for Computational Linguistics*, 1-9.

STRZALKOWSKI, T. (ed.) (1999): *Natural Language Information Retrieval*, Boston, Kluwer.

STRZALKOWSKI, T., F. LIN, J. WANG y J. PEREZ-CARBALLO (1999): *Evaluating NLP Techniques in IR, in Natural Language Information Retrieval*, Boston, Kluwer.

STRZALKOWSKI, T., G. STEIN, G.B. WISE y A. BAGGA (2000): "Towards the Next Generation Information Retrieval", RIAO 2000, París.

STRZALKOWSKI, T., J. PEREZ-CARBALLO, J. KARLGREN, A. HULTH, P. TAPANAINEN y T. LAHTINEN (1999): "Natural Language Information Retrieval: TREC-8 Report", *TREC Proceedings*.

SULLIVAN, D. (2001): *Document Warehousing and Text Mining*, Nueva York, Wiley Computer Publishing.

TAN, A. H. (1999): "Text Mining: the State of the Art and the Challenges", en *Proceedings of PAKDD'99. Workshop in Knowledge discovery from Advanced Databases*, Beijing, 71-76.

VAN HARMELEN, F. (2002): "How the Semantic Web will change KR: challenges and opportunities for a new research agenda", *The Knowledge Engineering Review*, 17: 1.

VOORHEES, E. (1999): "Natural Language Processing and Information Retrieval" en PAZIENZA M.T. (ed.) (1999): *Information Extraction: Towards Scalable, Adaptable Systems*, Lecture Notes in AI 1714, Berlín, Springer.

— (2003): "Evaluating the Evaluation: A Case Study using the TREC 2002 Question Answering Track", *Proceedings of HLT-NAAL 2003*, Edmonton, 181-188.

WHITE, R.W., J.M. JOSE y I. RUTHVEN (2002): "Using implicit contextual modelling to help users in information seeking", [poster] en *Proceedings of Building Bridges: Interdisciplinary context-sensitive computing*, Glasgow.

Metodología y tecnologías para la elaboración de diccionarios terminológicos onomasiológicos

AMPARO ALCINA

1. Introducción

La terminología constituye un pilar fundamental sobre el que se asienta el trabajo del traductor y mediador lingüístico. La industria de la traducción y la localización tiene un grave *cuello de botella* en la terminología como ha señalado el informe *Terminology Management Practices and Trends* publicado por LISA (<http://www.lisa.org>). La falta de diccionarios o su baja calidad obliga al traductor a ocuparse de la tarea terminológica con las consiguientes pérdidas de tiempo y eficacia en la tarea propiamente dicha de la traducción.

La multiplicación de ámbitos temáticos por el progreso de la ciencia y la tecnología, la rapidez con la que aparecen, cambian y evolucionan sus conceptos y la inmediatez con que se requiere que los nuevos términos y conceptos se divulguen y se traduzcan convierte a los diccionarios en instrumentos muy valiosos. Sin embargo, esta situación no ha motivado un aumento de diccionarios terminológicos en la misma proporción ni rapidez.

El problema de la baja calidad de muchos diccionarios terminológicos que se publican se debe a que la publicación de diccionarios no va precedida, en muchos casos, de un trabajo terminológico sistemático ni un auténtico tratamiento del concepto. En muchos casos, el proceso de elaboración se limita a reproducir las definiciones de otros diccionarios (con leves modificaciones), a traducirlas de diccionarios en otras lenguas o a adaptar puntualmente al formato de definición el contexto definitorio o explicativo encontrado en un manual. Ello produce pro-

blemas de ausencia de términos, circularidad o falta de información en las definiciones, cuando no definiciones desfasadas, erróneas o incoherentes.

Las causas de esta situación de escasez de diccionarios de calidad en un contexto de necesidad de diccionarios podemos intuir que están en la complejidad que conlleva el trabajo terminológico sistemático y quizás su coste económico. La necesaria participación de un equipo multidisciplinario formado por expertos del ámbito de especialidad concreto, terminólogos, expertos en las lenguas concretas del diccionario e informáticos, y la necesidad de recursos informáticos (ordenador, escáner, programas de gestión terminológica, conexión a Internet), libros u otros materiales textuales especializados y actuales, suponen un gasto quizás demasiado elevado en relación con el beneficio que se obtiene.

Esta situación se podría aliviar en gran medida si se contara con las herramientas informáticas adecuadas, así como con terminólogos expertos con una sólida base de formación en tecnologías lingüísticas.

En este artículo planteo cuáles son las limitaciones tecnológicas actuales y los retos que es necesario abordar para una nueva generación de herramientas en Terminótica. Seguidamente, explicaré cómo abordamos esta cuestión en el marco del proyecto ONTODIC para avanzar hacia el desarrollo de diccionarios onomasiológicos.

2. Hacia una nueva generación de herramientas en Terminótica

El panorama actual de la terminótica (Alcina, 2003; Alcina y otros, 2005) comprende herramientas y recursos informáticos que pueden realizar:

- La búsqueda y selección de textos en Internet.
- La extracción automática de términos simples y complejos (como TermExtract, KeyWords de WordSmith, TermoStat).

- La gestión de bases de datos terminológicas que facilitan la labor de almacenamiento y gestión de información de los términos: denominación y sus equivalencias en otras lenguas, contextos definitorios y lingüísticos, categoría gramatical, su definición (Multiterm de Trados).
- La edición y publicación de las fichas terminológicas (Multiterm de Trados).

Además, la red Internet permite el acceso a gran cantidad de textos que facilitan sin duda la extracción de terminología e información contextual. Por otra parte, estos programas informáticos facilitan la interacción con programas de traducción asistida y memorias de traducción, de forma que el traductor cuenta con acceso simultáneo a las memorias de traducción y bases de datos terminológicas. Algunos de estos programas cuentan con versiones que permiten el trabajo en red local e incluso a través de Internet, lo que permite que una misma base de datos sea alimentada por distintos usuarios simultáneamente, lo que facilita enormemente el trabajo en equipo.

Este panorama puede mejorarse notablemente gracias al espectacular despliegue de las tecnologías y a la cada vez mayor colaboración interdisciplinar. En concreto, creemos que el mayor escollo en Terminótica es el tratamiento del concepto. Si conseguimos desarrollar un auténtico tratamiento automatizado del concepto, podremos conseguir alcanzar objetivos estrechamente vinculados, como:

1. La automatización de la gestión de conceptos y elaboración de definiciones.
2. La ampliación de la gama de consultas del usuario: las consultas onomasiológicas.
3. El desarrollo de diccionarios *comprensibles* para el ordenador *(machine readable)*.

2.1 Automatizar la gestión de conceptos y elaboración de definiciones

La tarea de redacción de definiciones exige la lectura atenta de los contextos definitorios presentes en el corpus para comprender el concepto,

agrupar y clasificar el concepto junto con los afines, el establecimiento de relaciones entre los términos, la identificación de sus características, así como la selección de las características esenciales que, una vez ordenadas, deberán formar parte de la definición.

Una herramienta que permita gestionar automáticamente toda la información conceptual y facilite la elaboración de definiciones contribuiría a agilizar la tarea del terminólogo, hacerla más eficaz y evitar errores e inconsistencias.

Actualmente, también se está trabajando intensamente en la extracción de información conceptual a partir de corpus textuales, lo que facilitará sin duda la ardua tarea de la lectura atenta de los textos. Y resulta necesario también avanzar en la línea de diseñar herramientas que permitan analizar los conceptos y elaborar las definiciones.

Esta herramienta debería integrarse en un sistema eficaz y completo de *gestión de proyectos terminológicos*, un sistema que integre las herramientas necesarias para la automatización completa del trabajo terminológico (Alcina, 2001).

2.2 Ampliar la gama de consultas del usuario: las consultas onomasiológicas

La Sociedad del conocimiento demanda mejor acceso a la información y al conocimiento, en un mundo en que las disciplinas científicas se multiplican al igual que la cantidad de terminología que utilizan, la globalización nos lleva a un permanente multilingüismo y las tecnologías están cada vez más presentes en cada momento de nuestra vida y nuestro trabajo.

Las ventajas del diccionario electrónico frente al impreso apenas han experimentado mejoras en los últimos años, y éstas siguen siendo la rapidez de respuesta en las consultas, la posibilidad de ordenación por distintos campos, la ordenación alfabética en distintos idiomas, la actualización inmediata, el intercambio de datos y las múltiples posibilidades de difusión (Alcina, 1997).

Es necesario *reinventar* el diccionario en formato electrónico de manera que se aprovechen las ventajas que puede ofrecer este formato para dar respuesta a las demandas actuales de la Sociedad del conocimiento, que exige la mejora en el acceso a la información. Un sistema que facilite al terminólogo la gestión automática de conceptos también permitiría abrir una vía hacia el desarrollo de diccionarios onomasiológicos, es decir, diccionarios que permitan realizar consultas no sólo a partir de las denominaciones, sino también a partir de los significados.

Las consultas semasiológicas, que son las que permiten los diccionarios actuales, sólo cubren una pequeña parte de las necesidades de información de los usuarios de diccionarios. Hasta ahora, al diccionario le podemos preguntar: ¿qué significa el término *x*? ¿qué categoría gramatical tiene el término *x*? o ¿cuál es el equivalente en inglés del término *x*? Para responder a esta pregunta, basta introducir el término *x* en la casilla de consulta y nos devolverá la ficha en donde se encuentra la información con la definición del término *x*, su categoría gramatical o su equivalencia al inglés.

Pero el usuario que consulta un diccionario, sea un experto en el ámbito de conocimiento específico o un mediador lingüístico, requiere otras informaciones que a su vez exigen formas de acceso diferentes a las que se plantean tradicionalmente desde el punto de vista semasiológico.

Es necesario dar un paso más allá y plantearnos cómo conseguir que el diccionario responda a preguntas onomasiológicas. Se trata de preguntas más complejas como ¿qué entidad tiene la propiedad *y*? o ¿qué términos comparten la propiedad *y*? o ¿qué tipos de *x* existen? o ¿qué proceso da lugar a *y*?, por poner algunos ejemplos. El usuario del diccionario terminológico ha aprendido que estas preguntas no se formulan directamente, sino que debe conocer bien la estructura del diccionario y emplear distintas estrategias de consulta para encontrar la respuesta adecuada. Probablemente, el usuario del diccionario tendrá que buscar uno o varios términos, leer una o varias definiciones y sacar sus conclusiones a partir de la comprensión de la información que halle.

2.3 Desarrollar diccionarios machine readable

Muchos sistemas informáticos que se diseñan para el acceso a la información (consulta de bases de datos en Internet o Web Semántica) o para integrarse en sistemas de procesamiento del lenguaje natural (traducción automática) requieren el desarrollo de diccionarios que contengan información lingüística y, cada vez más, también información semántica. La información conceptual que acumulan los diccionarios o bases de datos terminológicas sería de gran utilidad para estos sistemas si esta información se pudiera procesar informáticamente. Sin embargo, las definiciones redactadas de los diccionarios sirven para que el usuario humano las comprenda, pero un sistema de procesamiento del lenguaje natural no puede sacar provecho de ellas como quedó patente en el proyecto ACQUILEX (Calzolari y Briscoe, 1992), entre otros.

Sería muy útil que los diccionarios y bases de datos almacenaran no sólo el resultado del análisis del concepto, es decir, su descripción mediante una definición redactada, sino que también se pudiera tener acceso a la información conceptual pormenorizada realizada en la fase de análisis conceptual.

Hoy en día, la tecnología proporciona flexibilidad en las formas de acceso a la información, permite almacenar gran cantidad de información y ofrece rapidez en el acceso, por lo que las cuestiones técnicas no deberían ser un impedimento. De algún modo, desde la Terminología debemos conseguir que el ordenador sea capaz de *comprender* el significado que encierran los conceptos y para ello, debemos afanarnos en explicitar la información que contienen. Éste es el punto de partida del proyecto ONTODIC.

3. ONTODIC: hacia el diccionario onomasiológico basado en ontologías

En el proyecto ONTODIC señalamos la necesidad de avanzar en el desarrollo de una nueva generación de herramientas en Terminótica que permita realizar consultas onomasiológicas, que facilite la generación automática o semiautomática de definiciones y que sea comprensible por el ordenador. Lo que requiere este panorama es avanzar hacia el análisis formal y sistemático de los conceptos, de modo que se pueda implementar en un sistema informático que almacene correctamente la información y la recupere cuando se le pida. Para ello, la información debe estar completamente estructurada y formalizada, y deben establecerse reglas lógicas que permitan establecer deducciones sobre la información proporcionada.

Estos objetivos requieren abordar los siguientes aspectos, que también pueden ser considerados como etapas:

1. Desarrollar fundamentos y metodología para llevar a cabo el análisis sistemático y formal de conceptos, que incluya la identificación y formalización de sus elementos constituyentes como características, las relaciones conceptuales, la clasificación de conceptos y su estructuración.
2. Diseñar un modelo formal que incluya los elementos que intervienen en la comprensión del concepto y que explicite su interacción. La formalización debe conducir a establecer deducciones a partir de la información proporcionada.
3. Ensayar y describir tecnologías que permitan la implementación del modelo formal y el almacenamiento y recuperación de la información conceptual.

Para validar la metodología y las técnicas elaboraremos un diccionario onomasiológico de la cerámica industrial. La elección del ámbito de la cerámica industrial viene motivado por la facilidad de obtener de nuestro entorno socioeconómico más próximo, la provincia de

Castellón, la documentación necesaria (revistas, monografías, informes, diccionarios, etc.) así como el asesoramiento por parte de expertos. Esta facilidad de acceso ha motivado también la generación de distintos recursos en proyectos anteriores, como la base de datos terminológica Cerámica con más de 4.000 términos y el corpus textual digitalizado TXTCeram con más de 2 millones de palabras procedentes de textos de cerámica. No obstante, aunque la base de conocimiento se desarrolle en el ámbito de la cerámica industrial, los resultados de este proyecto pretenden ser, en gran medida, extensivos a otros ámbitos de especialidad.

A continuación explicaré cómo pretendemos abordar estas cuestiones en el proyecto ONTODIC. Para ello, me centro en las nociones de concepto y sus elementos constituyentes (las características, las relaciones conceptuales y las clases de conceptos), y la metodología que se puede seguir para su formalización. Por último, se señalan los sistemas informáticos que se han utilizado para implementar aspectos relacionados con el concepto, sus limitaciones y cómo debería ser el sistema informático capaz de gestionar los conceptos.

3.1 Elementos para la formalización de conceptos en ONTODIC

Según la norma UNE 1-066-91 "los conceptos son construcciones mentales que sirven para clasificar los objetos individuales del mundo exterior o interior mediante una abstracción más o menos arbitraria". En ese proceso de abstracción, se seleccionan las características relevantes que definen una clase de objetos. La referencia al concepto como construcción mental o unidad de pensamiento es repetida por distintos autores (Sager, 1990; Arntz y Picht, 1989).

La descripción de un concepto es el conjunto de las características interrelacionadas que describen de forma esencial y complementaria una clase de objetos de la realidad (Cabré, 1992), y puede realizarse mediante dos procedimientos: comprensión y extensión. El proceso de comprehensión consiste en la enumeración ordenada de características, desde la más general a la más específica. Así, por ejemplo, la

comprehensión del concepto 'lobo' es el conjunto de las características siguientes:

lobo: mamífero -> carnívoro -> fisípedo -> cánido

La extensión consiste en enumerar todas las posibles realizaciones de un concepto, todas sus especies. Por ejemplo, la extensión del concepto 'halógeno' es el conjunto de las especies posibles de halógenos:

halógenos: flúor, cloro, bromo, yodo, ástato

El primer paso para conseguir procesar el concepto es analizarlo en busca de los elementos de los que se compone y el proceso de abstracción que conlleva la elaboración del concepto, las interacciones que generan o los resultados que produce.

3.1.1 Las características

El concepto es una construcción mental, una unidad de pensamiento compleja, formada por elementos de significado más simples que se denominan *características*. Así, el concepto de 'baldosa cerámica' se puede descomponer en las siguientes características:

- pieza cerámica impermeable;
- constituida por un soporte de naturaleza arcillosa con o sin recubrimiento de esmalte;
- fabricadas a partir de composiciones de arcillas y otras materias primas inorgánicas, que se someten a molienda y/o amasado, se moldean y seguidamente son secadas y cocidas a temperatura suficiente;
- utilizadas para revestimiento de suelos y paredes.

Generalmente, el concepto se describe en los textos mediante explicaciones más o menos largas. En los diccionarios, se representan mediante definiciones redactadas conforme a determinados criterios lexicográficos.

No todas las características tienen la misma importancia de cara a describir un concepto, y algunas características resultan esenciales para su descripción, mientras que otras son complementarias. De manera

que las primeras deben estar necesariamente presentes en una definición terminológica, mientras que las segundas no.

Además, las características pueden hacer referencia a diferentes aspectos del concepto. Por ejemplo, una característica como la 'forma' de un objeto está más próxima conceptualmente a la característica 'tamaño' porque ambas refieren a aspectos físicos de dicho objeto. Mientras que si hacemos referencia a la utilidad práctica o al uso que se le da estaremos haciendo referencia a aspectos funcionales. De ahí que las características se agrupan en diferentes tipos: intrínsecas y extrínsecas (Cabré, 1992), de estado, funcionales y relacionales (Arntz y Picht, 1989).

Meyer estableció unas técnicas de análisis formal de conceptos que consistían en identificar las características y establecer parejas de *atributo: valor*. Esta concepción del análisis se podía representar mediante tablas. En la primera columna de la tabla se enumeran los términos y en las siguientes columnas se especifican las características atributo que sirven para describir un conjunto de conceptos. En cada fila, se introduce en la primera columna el término, y en las restantes columnas el valor que toma cada característica respecto a ese término. Así, para cada grupo de conceptos se identifican las características que los describen y el valor que cada característica puede tomar respecto de un concepto, formando una pareja de *atributo: valor*. Por ejemplo, en el caso del grupo de conceptos 'soportes de la información' tendríamos el concepto 'disco óptico' con la siguiente información (ejemplos tomados de Meyer y otros, 1997):

term	physical form:	diameter:	degree of writability:	encoding method:	storage capacity:	content:	standard:
optical disc	*disk*	*3.5, 4.72, 5.25, 8, 12 or 14 inches*	*one of read-only, write-once, erasable*	*digital or analog*	*up to 10 Gb*	*one or both of audio, textual data*	—

Tabla 1.

mientras que otro concepto como 'disco compacto' se analizaría mediante las mismas características pero estas características tomarían valores distintos:

term	physical form:	diameter:	degree of writability:	encoding method:	storage capacity:	content:	standard:
optical disc	*disk*	*4.72 inches*	*one of read-only, write-once, erasable*	*digital*	*up to 600 Mb*	*one or both of audio, textual data*	*one of Red Book, Yellow Book, Green Book, Orange Book*

Tabla 2.

Esta técnica se centra en identificar y establecer características que afectan a los conceptos distinguiéndolos de los valores que pueden tomar. De este modo, se pueden establecer grupos de conceptos que tienen en común un conjunto de características, independientemente de que sus valores puedan ser diferentes. En el ejemplo podemos ver que los términos *optical disk* y *compact disk* comparten el conjunto de características: forma, diámetro, escritura, forma de codificación de datos, capacidad de almacenamiento, o forma estándar. Precisamente lo que hace que estos términos identifiquen conceptos distintos es que los valores de las características son diferentes. Por otra parte, el hecho de que necesiten utilizar las mismas características para describir su significado nos lleva a concluir que esos conceptos deben adscribirse a un mismo grupo de conceptos, por ejemplo, el grupo de conceptos 'soportes de la información'.

Por último, una vez identificado un concepto bajo un grupo de conceptos, ese concepto deberá ser descrito con las características que corresponden a ese grupo.

Una vez formalizada la información conceptual, ésta se puede utilizar también para generar definiciones de forma automática (Sierra y McNaught, 2000). En ONTODIC pretendemos formular patrones de definición para cada grupo de conceptos que comparten un mismo hiperónimo, que actúa como descriptor en la definición y conjunto de características esenciales. Por ejemplo, el siguiente patrón se podría utilizar para generar las definiciones del grupo de conceptos '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. Así, en el caso de que A = bordón, el generador sustituirá X = rectangular, Y = muy estrecho, Z = lisos y monocromos y Q = componer cenefas, dando lugar a la definición:

> bordón: baldosa cerámica de forma *rectangular* y de tamaño *muy estrecho* que está decorada con *colores lisos* y tiene como función *componer cenefas*.

Por último, como señaló Bowker, los conceptos que forman parte de un mismo grupo pueden clasificarse atendiendo a más de una característica, a más de una dimensión, dando lugar al fenómeno de la multidimensionalidad (Bowker, 1997). Así, por ejemplo, la representación de la clase 'vehículo' es unidimensional si lo clasificamos desde el punto de vista de la característica 'medio de transporte' en vehículos de aire, vehículos de tierra y vehículos de agua. Sin embargo, también podemos utilizar las características o dimensiones 'método de propulsión' en la que tendríamos 'vehículo a motor' o 'vehículo sin motor', e incluso la dimensión 'tipo de carga que transporta', según la cual las posibilidades serían 'vehículo de pasajeros', 'vehículo de carga' y 'vehículos de pasajero y carga'.

Por tanto, en ONTODIC la formalización de las características de un ámbito de especialidad requiere:

1. Enumerar con detalle las características, y en particular los atributos, que resultan relevantes para un determinado ámbito de especialidad (en nuestro caso, la cerámica industrial).
2. Distinguir el atributo y valor que subyace a cada característica, e identificar el tipo de valor o gama de valores posibles para un determinado atributo.
3. Asignar las características al grupo o grupos de conceptos a los que describen.
4. Clasificar las distintas características según los aspectos que cubren (intrínsecas o extrínsecas, de estado o función) y el ámbito de conocimiento al que pertenecen.

5. Agrupar los conceptos del ámbito de especialidad según los atributos de características que tengan en común y establecer grupos de conceptos basados en las características (atributos) que comparten.
6. Establecer, para cada grupo de conceptos, la relevancia de las características especificadas; según su importancia (esenciales y complementarias). Esta clasificación tiene consecuencias importantes para la elaboración de la definición.
7. Identificar los valores que para cada concepto corresponden a los distintos atributos, lo que nos permitirá distinguir conceptos o identificar sinónimos.

Así, por ejemplo, cada concepto perteneciente al grupo de conceptos o rama del árbol 'productos cerámicos' (del que forman parte los términos *ángulo*, *azulejo*, *baldosa*, *bocel*, *bordón*, etc.) sería definido por el mismo grupo de características esenciales (por ejemplo: características de fabricación, forma, tamaño, decoración, lugar de colocación y función).

3.1.2 Relaciones conceptuales

Las relaciones conceptuales son, probablemente, una de las cuestiones que más ha atraído la atención de investigadores y terminólogos, de manera que, donde antes se establecían únicamente dos tipos de relación: genérica y partitiva, más un conjunto de *otras relaciones*, actualmente el listado es bastante extenso. Sager ofrece la siguiente lista de relaciones conceptuales: causa-efecto, material-producto, material-propiedad, material-estado, proceso-producto, proceso-instrumento, proceso-método, proceso-paciente, fenómeno-medida, objeto-contraagente, objeto-contenedor, objeto-material, objeto-cualidad, objeto-operación, objeto-característica, objeto-forma, actividad-lugar (Sager, 1990).

A las relaciones conceptuales establecidas con carácter general, otros estudios han podido añadir subclasificaciones y especificaciones. Es el caso de la relación parte-todo, en la que se han distinguido hasta seis subtipos: componente-objeto integral, miembro-colección, porción-masa, materia-objeto, característica-actividad, lugar-área (Winston y otros, 1987).

Otras clasificaciones de las relaciones, como aquella que distingue las relaciones jerárquicas de las no jerárquicas parecen hacer alusión más bien a una propiedad de algunas relaciones, que a su vez conlleva propiedades como la transitividad. Por ejemplo, la relación género-especie es jerárquica, y ello implica en este caso que el elemento subordinado, el hipónimo, la especie, hereda del hiperónimo sus características que posee por el hecho de serlo.

Algunas relaciones pueden tener propiedades específicas de ese tipo de relación. Por ejemplo, los distintos subtipos de la relación partitiva que hemos indicado anteriormente pueden distinguirse por las propiedades funcional, homeomérica o separable. Así, por ejemplo, la relación partitiva entre los conceptos 'asa' y 'taza' tiene las propiedades funcional (cada parte tiene una función respecto al todo), no homeomérica (las partes no son similares entre sí), y separable (las partes pueden separarse una de la otra); mientras que la relación entre 'grano' y 'sal' es no funcional (cada parte no tiene asignadas funciones), homeomérica (las partes son similares) y separable (las partes se pueden separar). Estas propiedades descritas en las relaciones pueden contribuir a formalizar el significado de los conceptos y a obtener deducciones e inferencias. También hay que prestar especial atención a las combinaciones de varios tipos de relaciones (Otman, 1996).

Otro aspecto que contribuye a formalizar las relaciones conceptuales es la relación que pueden mantener con otras relaciones. Por ejemplo, la relación género-especie es inversa de la relación especie-género. Por tanto, al establecer que 'mamífero' es hiperónimo de 'lobo', se deduce automáticamente que 'lobo' es hipónimo de 'mamífero'. Esto añade poder de deducción a la base de conocimiento y ahorra esfuerzos al introducir los datos.

Otman presenta un modelo de representación semántica de relaciones conceptuales basado en redes semánticas de la Inteligencia artificial. En este modelo, una relación permite establecer un enlace conceptual entre otros conceptos. Otman lo representa formalmente como:

R (A, B)

Donde *R* representa el nombre de la relación y *A* y *B* los dos conceptos relacionados. Así, 'tipo-de (lobo, mamífero)' significa que *perro* es un tipo de *mamífero*.

La representación de Otman recoge la idea, en general compartida por otros autores, de que "una relación establece un enlace binario entre dos conceptos" (Otman, 1996: 55). No obstante, la práctica de la terminografía parece indicar más bien que las relaciones conceptuales no son binarias, sino que en uno o ambos lados del enlace podría haber más de un elemento. Así, Soler (2006) ha sugerido una representación del tipo:

$a\ R\ \{b_1, b_2, \ldots, b_n\}$

Donde *R* representa el nombre de la relación, *a* el primer término de la relación, *b* el segundo término de la relación, y los subíndices de *1* a *n* para *b* indican que puede haber un número variable de términos en la segunda parte de la relación. Por otra parte, se puede sacar la falsa idea de que un término sólo mantiene un tipo de relación R con otros conceptos. Por ello, Oster prefiere la representación mediante esquemas, que permiten reflejar distintos tipos de relación para un mismo concepto (Oster, 2006). En un modelo formal de representación de las relaciones, habrá que tener en cuenta estas cuestiones.

Por todo ello, en ONTODIC el desarrollo de un modelo formal de las relaciones conceptuales de un ámbito de especialidad requiere:

1. Elaboración de un catálogo de relaciones conceptuales generales que afectan a cualquier ámbito de especialidad y distinguir, en su caso, relaciones conceptuales especializadas que afecten a ámbitos de especialidad concretos.
2. Estudio de las propiedades lógicas y matemáticas que afectan a las relaciones conceptuales y sus combinaciones.
3. Análisis exhaustivo formal de cada relación conceptual. Este análisis debe incluir:
   a) La especificación de las clases conceptuales a las que pueda afectar;

b) Las propiedades de tipo lógico o matemático y las condiciones para que se produzcan;
c) La especificación de relaciones conceptuales que interactúan con ella (por ejemplo, si existe una relación conceptual inversa);
d) Las implicaciones de tipo lógico o matemático a que pueda dar lugar su aplicación entre conceptos unidos por esa relación y que puedan ser determinantes para establecer deducciones entre ellos.

3.1.3 Clases de conceptos

Sager (1990) distingue las siguientes clases de conceptos:

- entidades
- actividades o procesos
- cualidades o propiedades
- relaciones

Más adelante, Sager y Kageura (1995) establecen una clasificación de conceptos y estructuras de conceptos más compleja. No repetiremos completamente esta clasificación por su exhaustividad, pero para dar idea de su diversificación, veamos que, en un primer nivel, los conceptos de la clase 'entidad' se clasificarían a su vez en:

- entidades materiales
- entidades clasificatorias
- entidades abstractas
- entidades representativas

Cada uno de estas clases se divide a su vez en subclases. Por ejemplo, la clase 'entidades materiales' se subdivide a su vez en: animados (a su vez, personas u organizaciones) e inanimados (que pueden ser objetos físicos o lugares).

Esta clasificación de conceptos serviría como punto de partida, como un árbol de campo original que podría utilizarse al inicio de cualquier trabajo terminológico para ir adscribiendo los distintos conceptos específicos de un ámbito de especialidad. Además de tener una utilidad al

inicio de cualquier trabajo terminológico de una especialidad, también podría servir para establecer puntos en común o comparaciones entre los conceptos de distintas especialidades, al establecer elementos comunes por el hecho de estar adscritos a una misma clase de conceptos.

La clasificación de los conceptos en distintos tipos tiene, en ONTODIC, dos finalidades. La primera, como ya hemos visto, la de permitir adscribir los distintos conceptos de un ámbito temático en grupos de términos afines, con los que compartirá una serie de características. Las clases de conceptos contribuyen a tener una primera etapa de clasificación cubierta y funcionan como una metaclasificación.[1]

La segunda finalidad consiste en que las clases de conceptos servirán también para restringir el tipo de conceptos que pueden quedar adscritos a una propiedad o relación. Así, por ejemplo, 'resistencia a la heladicidad' es un concepto que señala una propiedad del ámbito de la cerámica industrial que puede adscribirse a ciertos tipos de baldosa cerámica, pero no a un horno cerámico o al proceso de esmaltado. Los tipos de conceptos pueden actuar, por tanto, como elementos que restringen el ámbito de actuación de un concepto.

3.1.4 Estructuración de conceptos

La estructuración de un ámbito de conocimiento en Terminología se ha representado mediante árboles de campo, utilizados principalmente en terminografía para delimitar el ámbito de un trabajo terminológico. Así, por ejemplo, el árbol de campo para el diccionario terminológico sobre Cerámica que proponemos consta de las siguientes ramas:

1. Producto
   1.1 Producto acabado
      1.1.1 Características del producto acabado
      1.1.2 Defectos del producto acabado
      1.1.3 Tipos de producto acabado
      1.1.4 Partes del producto acabado

1 En cierto modo, su utilidad sería similar a la de las metaontologías de la Ingeniería del conocimiento.

1.2 Producto intermedio
1.2.1 Características del producto intermedio
1.2.2 Defectos del producto intermedio
1.2.3 Tipos de producto
1.3 Otros productos

2. Fabricación
2.1 Procesos
2.2 Maquinaria y accesorios

3. Materias primas y aditivos
3.1 Materias primas
3.2 Aditivos

4. Colocación de baldosas
4.1 Procesos de colocación
4.2 Maquinaria y accesorios de colocación
4.3 Materiales para colocación
4.3.1 Materiales de base
4.3.2 Materiales de adherencia

5. Calidad
5.1 Procesos de control de calidad
5.2 Maquinaria y accesorios de control de calidad

6. Unidades de medida

7. Organismos, Instituciones

En ONTODIC, deberá quedar reflejado respecto a cada concepto, el ámbito temático al que se adscribe y la rama del árbol de campo que corresponda. En el desarrollo del diccionario actual sobre Cerámica, OntoCeram, cada concepto deberá quedar adscrito a una de estas ramas.

3.2 Terminología, ontologías y sistemas de gestión de conceptos

En la década de los 90 se inicia, en Inteligencia artificial y luego Ingeniería del conocimiento, la investigación sobre bases de conocimiento u ontologías. Gruber las define del modo siguiente:

> An ontology is an explicit specification of a conceptualization. (Gruber, 1993)

En un artículo anterior, se explicitan mejor los componentes de estas ontologías:

> An ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary. (Neches y otros, 1991)

En esta definición podemos apreciar la similitud entre los componentes de una ontología (términos, relaciones, vocabulario de un ámbito de especialidad, normas para la inclusión de nuevos términos) y los componentes de los conceptos terminológicos. Por ello no debe sorprendernos que hacia finales de la década de los 90, muchos autores señalen las ventajas que para unos y otros, ingenieros y terminólogos, puede tener el trabajo conjunto (Gamper y otros, 1999).

Ejemplo de ello es la numerosa bibliografía y reuniones científicas que han tenido lugar en estas líneas de investigación: el número especial dedicado en la revista Terminologies nouvelles al tema "Terminología e inteligencia artificial" en 1999, los congresos Terminología e Inteligencia Artificial (TIA), Terminology and Knowledge Engineering (TKE), OntoLex, Semantic Web y revistas como *Ontological Applied*, que dedica una sección a Terminología, confirman el cruce de caminos en los hallazgos desde la Terminología y la Ingeniería del conocimiento.

Esta intensa relación ha dado lugar al desarrollo de ontologías y de sistemas de gestión de conceptos en el ámbito de la Terminología. Así, a principios de los años 90, en la Universidad de Otawa se diseña la base de conocimiento CODE4 (que luego dio lugar a Ikarus y más tarde a COGNITERM) que contiene información altamente estructurada sobre atributos y relaciones de conceptos (Meyer y otros, 1997). Esta base de conocimiento dispone de interfaces de usuario que proporcionan distintas formas de mostrar la información que contienen. Dispone de un navegador gráfico que muestra los conceptos como nodos y las relaciones como flechas que enlazan nodos y que se etiquetan con el nombre de la relación. También dispone de un navegador textual que

presenta los distintos conceptos en una lista jerarquizada, en forma de viñetas indentadas. A partir de los datos introducidos, la base de conocimiento puede deducir automáticamente otras informaciones con ayuda de mecanismos como la herencia. Así, las características asignadas a un determinado concepto genérico serán heredadas por todos sus conceptos específicos salvo que se especifique otra cosa. Este tipo de mecanismos también sirven para detectar inconsistencias en una estructura de conocimiento.

Más tarde, el proyecto OncoTerm de la Universidad de Granada dirigido por Faber desarrolló una ontología de la Oncología donde se organizan sistemáticamente los términos según las relaciones existentes entre ellos. La ontología desarrollada se basa en la metaontología Mikrokosmos. En este mismo proyecto, se desarrolló una metodología sobre la adquisición de conocimiento y creación de la ontología, y también un sistema de gestión de conceptos propio denominado OntoTerm (Moreno Ortiz, 2002). Este programa de gestión de conceptos está conectado a una base de datos terminológica que garantiza la vertiente lingüística del estudio terminológico y garantiza la correspondencia entre conceptos y denominaciones (Faber, 2002; Faber y Jiménez, 2002). También desde el grupo IULA-Term, en el Instituto de Lingüística Aplicada de la Universitat Pompeu Fabra, dirigido por Teresa Cabré, se pone en marcha un proyecto sobre la terminología del Genoma humano y se utiliza el programa OntoTerm para el desarrollo de la taxonomía, al tiempo que se publica un informe sobre ontologías (Feliu y otros, 2002).

Actualmente se están desarrollando otros sistemas de gestión de conceptos como CAOS 2, en la Universidad de Copenhague, en el que se puede trabajar a través de Internet (véase <http://www.id.cbs.dk/caos>).

En ONTODIC nos hemos planteado que dada la necesidad de abordar la creación de diccionarios onomasiológicos con aplicaciones en sistemas avanzados de procesamiento del lenguaje natural, lo más adecuado es combinar los conocimientos lingüísticos de que disponemos con las actuales técnicas de las ontologías desarrolladas por la Ingeniería del conocimiento, o la reciente Ingeniería ontológica. El aprovechamiento de estas tecnologías tiene la ventaja de que facilitará la

interoperabilidad de los contenidos terminológicos creados con otros sistemas y por tanto su reutilización.

Una vía para conseguir esto es utilizar los editores de ontologías ya desarrollados en ingeniería del conocimiento y adaptarlos a las necesidades de la terminología. Muchos de estos programas disponen de una interfaz que permite su manejo por cualquier usuario, y en los que la complejidad reside no en el manejo de los programas, sino en que exista una verdadera estructuración de los contenidos que se desea introducir. Además de la interfaz de introducción de términos, relaciones y propiedades, pueden disponer de funciones como consultas, editores gráficos, etc. Cabe mencionar entre los editores de ontologías: Protégé, desarrollado en la Universidad de Standford y WebODE, desarrollado en la Universidad Politécnica de Madrid. El primero dispone de gran número de funciones, la posibilidad de introducir *plug-ins* que permiten adaptar o personalizar el entorno para las funciones requeridas por el usuario, y trabaja en un ordenador local. WebODE es mucho más robusto y es accesible a través de la Web, pero no dispone de tantas funciones y sobre todo no tiene la posibilidad de añadir otros programas informáticos como Protégé.

De ahí que en el proyecto ONTODIC hemos decidido ensayar la implementación de nuestro modelo formal en el editor de ontologías estándar utilizado en ingeniería del conocimiento Protégé (en este mismo volumen, véase Sanz y Jiménez-Ruiz, 2009).

3.3 *Metodología para el análisis conceptual*

La metodología y los fundamentos que deben guiar la investigación terminológica y la presentación de trabajos terminológicos está rigurosamente definida en la bibliografía (Sager, 1990; Cabré, 1992; Arntz y Picht, 1989; Alpízar Castillo, 1997), así como en distintas normas europeas e internacionales (normas ISO).

El grupo de Temmerman ha anunciado el desarrollo de una nueva metodología, la *termontografía (Termontography)*, que consiste en el desarrollo de una metodología terminológica que incluya el desarrollo

de ontologías y que se está utilizando en el marco del proyecto europeo FFPOIROT (<http://www.ffpoirot.org>) para la creación de terminología multilingüe en el ámbito de la inspección fiscal (Kerremans y otros, 2005).

Si queremos avanzar en el desarrollo de diccionarios onomasiológicos, es necesario especificar con mayor grado de detalle cómo se trabajan los distintos elementos que hemos ido mencionando y que afectan al análisis formal de conceptos. Para ello, en ONTODIC hemos tenido en cuenta los principios metodológicos desarrollados desde la ingeniería del conocimiento.

La ingeniería del conocimiento ha desarrollado distintas metodologías para la creación de ontologías, entre las que cabe destacar la metodología desarrollada en la Universidad Politécnica de Madrid METHONTOLOGY (Gómez Pérez, 2004: 107-197). En Methontology, la creación de ontologías se compone de tres tipos de actividades: actividades de gestión, actividades orientadas al desarrollo y actividades de apoyo.

1. Actividades de gestión. Comprende la planificación y calendario del proyecto y las actividades de control y garantía de calidad.
2. Actividades orientadas al desarrollo. Comprende las siguientes etapas:
   a) Especificación. Consiste en detallar el propósito de la ontología, cuáles son los usuarios finales.
   b) Conceptualización. Se estructura el ámbito de conocimiento como modelos de significado en el nivel del conocimiento.
   c) Formalización. Se transforma el modelo conceptual en un modelo formal o semi-computable.
   d) Implementación. Se construye el modelo computable en un lenguaje ontológico.

   Estas actividades de desarrollo van precedidas por una fase de pre-desarrollo en la que se estudia qué plataforma se utilizará para el desarrollo de la ontología, las aplicaciones en las que se integrará, etc. Y continúan con una fase de post-desarrollo en la que se llevan a cabo actividades de mantenimiento en la que se actualiza la ontología o se modifica si fuera necesario.

3. Actividades de apoyo. Se trata de actividades que se realizan al mismo tiempo que las actividades de desarrollo e incluyen la adquisición de conocimiento, evaluación, integración, mezcla, alineación, documentación y configuración.

4. Conclusiones

Consideramos prioritario, en una sociedad volcada por el conocimiento, avanzar hacia la elaboración de diccionarios terminológicos que permitan la consulta onomasiológica, y contar con herramientas que faciliten la elaboración de definiciones terminológicas e incluso su generación semiautomática y el desarrollo de bases de conocimiento útiles para otros sistemas de procesamiento del lenguaje natural.

El desarrollo de ontologías viene siendo un campo en auge debido a las múltiples aplicaciones en la sociedad del conocimiento, ya que permitirá, entre otras cosas, el desarrollo de la web semántica, favorecerá la recuperación de información en diversas aplicaciones mono y multimodales, así como las relacionadas con el procesamiento del lenguaje natural.

Por este motivo, en ONTODIC combinamos los fundamentos de la Terminología y de la Ingeniería ontológica. Por una parte, tomamos de la Terminología los fundamentos teóricos y aspectos metodológicos que sirven para la extracción de términos a partir de corpus textuales y los aspectos orientados al análisis formal de conceptos. De la Ingeniería ontológica tomamos los aspectos metodológicos y técnicas desarrollados para la creación de ontologías.

En primer lugar, se han diseñado estas etapas y los objetivos a alcanzar en cada una de ellas. Estos objetivos comprenden:

1. Profundizar en aspectos teóricos sobre el concepto, sus clases y sus elementos constituyentes como características o relaciones. Asimismo, es necesario indagar sobre las propiedades lógicas y matemáticas que caracterizan estos elementos.

2. Diseñar un modelo formal que incluya los elementos que forman parte del concepto y en el que se establezcan sus interacciones.
3. Ensayar la implementación del modelo formal en el editor de ontologías Protégé.

Los resultados alcanzados en las distintas etapas se validarán empíricamente mediante su aplicación a la elaboración de un diccionario onomasiológico de la cerámica industrial.

En relación con la implementación, nuestra hipótesis es que un editor de ontologías estándar como Protégé puede ser suficiente para implementar el modelo formal de ontología terminológica. Si el resultado es positivo, esto permitirá un impulso a las investigaciones en Terminología y al desarrollo de diccionarios. No obstante, si los resultados fueran negativos, esperamos haber desarrollado *feedback* para diseñar un editor de ontología terminológica eficaz.

En estos momentos, la investigación en ONTODIC ya ha dado lugar a avances en relación con la formalización de relaciones conceptuales (Maroto, 2007), la formalización de características (Estellés, 2006) y la extracción automática de conceptos (Soler, 2006).

Agradecimientos

Esta investigación ha sido posible gracias a la financiación recibida del Ministerio de Educación y Ciencia y Generalitat Valenciana para los proyectos *ONTODIC: Metodología y tecnologías para la elaboración de diccionarios onomasiológicos basados en ontologías. Recursos terminológicos para la e-traducción* (TSI2006-01911) y *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* (GV05/260).

Bibliografía

ALCINA, A. (1997): "Soportes de almacenamiento y formas de difusión de datos terminológicos. Las fuentes de información en terminología", *Revista Española de Lingüística Aplicada*, 12, 221-233.

— (2001): "Automatización de tareas en la elaboración de un diccionario terminológico", en CABRÉ, M. T., LL. CODINA y R. ESTOPÀ (eds.) (2001): *Terminologia i Documentació*, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 51-59.

— (2003): "La programación de objetivos didácticos en Terminótica atendiendo a las nuevas herramientas y recursos", en GALLARDO, N. (ed.) (2003): *Terminología y traducción: un bosquejo de su evolución*, Granada, Editorial Atrio, 79-89.

ALCINA, A., V. SOLER y A. ESTELLÉS (2005): "Internet como instrumento para la documentación en terminología y traducción. Hacia las plataformas de recursos electrónicos para el traductor especializado", en SALES, D. (ed.) (2005): *La Biblioteca de Babel: Documentarse para traducir*, Granada, Comares, 221-241.

ALPÍZAR CASTILLO, R. (1997): *¿Cómo hacer un diccionario científico-técnico?*, Buenos Aires, Memphis.

ARNTZ, R. y H. PICHT (1989): *Introducción a la terminología*, Madrid, Fundación Germán Sánchez Ruipérez.

BOWKER, L. (1997): "Multidimensional Classification of Concepts and Terms", en WRIGHT, S. E. y G. BUDIN (eds.) (1997): *Handbook of Terminology Management*, Philadelphia, John Benjamins, 133-143.

CABRÉ, M. T. (1992): *La terminologia. La teoria, els mètodes, les aplicacions*, Les Naus d'Empúries, Barcelona, Empúries.

CALZOLARI, N. y T. BRISCOE (1992): "Acquisition of Lexical Knowledge from Machine-readable Dictionaries and Text Corpora", *Computational Linguistics, COLING92*, Pisa.

ESTELLÉS, A. (2006): *Clasificación de características en el análisis de conceptos de la terminología cerámica*, Trabajo de investigación, Traducción y Comunicación, Universitat Jaume I.

FABER, P. (2002): "Oncoterm: Sistema bilingüe de información y recursos oncológicos", en ALCINA, A. y S. GAMERO (eds.) (2002): *La traducción científico-técnica y la terminología en la sociedad de la información*, Castellón, Publicaciones de la Universitat Jaume I, 177-188.

FABER, P. y C. JIMÉNEZ (eds.) (2002): *Investigar en Terminología*, Granada, Comares.

FELIU, J., J. VIVALDI y M. T. CABRÉ (2002): "Ontologies. A Review" [en línea], Serie Informes, 34, Barcelona, Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada. Disponible en: <ftp://ftp.iula.upf.es/pub/publicacions/02inf034.pdf> [Consulta: 26 nov. 2007].

GAMPER, J., W. NEJDL y M. WOLPERS (1999): "Combining Ontologies and Terminologies in Information Systems", en *TKE'99 Terminology and Knowledge Engineering*, Innsbruck.

GÓMEZ PÉREZ, A. (2004): *Ontological Engineering*, Nueva York, Springer.

GRUBER, T. R. (1993): "A translation approach to portable ontology specifications", *Knowledge Acquisition*, 5, 199-220.

KERREMANS, K., I. DESMEYTERE, R. TEMMERMAN y P. WILLE (2005): "Application-oriented terminography in financial forensics", *Terminology*, 11: 1, 83-106.

MAROTO, N. (2007): *Las relaciones conceptuales en la terminología de los productos cerámicos y su formalización mediante un editor de ontologías*, Tesis doctoral, Traducción y Comunicación, Universitat Jaume I.

MEYER, I., K. ECK y D. SKUCE (1997): "Systematic Concept Analysis within a Knowledge-Based Approach to Terminology", en WRIGHT, S. E. y G. BUDIN (eds.) (1997): *Handbook of Terminology Management*, Amsterdam, John Benjamins, 98-118.

MORENO ORTIZ, A. (2002): "Representación de la información terminológica en OntoTerm: un sistema gestor de bases de datos terminológicas basado en el conocimiento", en FABER, P. y C. JIMÉNEZ (eds.): *Investigar en Terminología*, Granada, Comares, 25-70.

NECHES, R., R.E. FIKES, T. FININ, T. R. GRUBER, T. SENATOR y W. R. SWARTOUT (1991): "Enabling technology for knowledge sharing", *AI Magazine*, 12: 3, 36-56.

OSTER, U. (2006): "Classifying domain-specific intraterm relations: A schema-based approach", *Terminology*, 12: 1, 1-17.

OTMAN, G. (1996): *Les représentations sémantiques en Terminologie*, París, Masson.

SAGER, J. C. (1990): *A practical course in terminology processing*, Amsterdam, John Benjamins.

SAGER, J. C. y K. KAGEURA (1995): "Concept Classes and Conceptual Structures: their Role and Necessity in Terminology", en *Symposium on French Language and linguistics, ALFA*, 191-216.

SANZ, I. y E. JIMÉNEZ-RUIZ (2009): "Ontologías en Informática", en ALCINA, A., E. VALERO y E. RAMBLA (eds.) (2009): *Terminología y Sociedad del conocimiento*, Berna, Peter Lang, 257-288.

SIERRA, G. y J. MCNAUGHT (2000): "Design of an onomasiological search system: A concept-oriented tool for terminology", *Terminology*, 6: 1, 1-34.

SOLER, V. (2006): *Búsqueda y clasificación de patrones léxicos para la extracción de información conceptual. El caso de la relación partetodo en los términos de la cerámica*, Trabajo de investigación, Departamento de Traducción y Comunicación, Universitat Jaume I.

WINSTON, M. E., R. CHAFFIN y D. HERRMANN (1987): "A Taxonomy of Part-Whole Relations", *Cognitive Science*, 11, 417-444.

Terminología y traducción automática

CELIA RICO PÉREZ

1. Introducción

Si bien es cierto que en la bibliografía especializada son escasas las referencias que tanto la traducción automática (TA) como la terminología se hacen mutuamente, no significa esto necesariamente que ambas disciplinas se ignoren. De hecho, uno de los elementos clave en todo sistema de TA, independientemente de su arquitectura o de la sofisticación de su uso, es, sin duda, el diccionario. Y no lo es tanto por su tamaño, esto es, por el número de entradas que contenga, como por el tipo de términos que incluya. Por ejemplo, un sistema de TA cuyo diccionario contenga un amplio léxico de uso general no resultará útil para traducir, pongamos por caso, las instrucciones de uso de una impresora a menos que recoja la terminología necesaria.

En general, cualquier sistema de TA de uso comercial[1] necesita una importante inversión de tiempo o dinero (o preferiblemente de los dos) para mantener actualizados los diccionarios y conseguir así la máxima productividad y calidad de los textos traducidos. De hecho, tal como señalan Bennet y Gerber (2003: 183-187), la gestión del léxico es uno de los aspectos clave en el mantenimiento a largo plazo de un producto comercial de TA. Esta tarea incluye no sólo la adaptación del diccionario a las necesidades del usuario (es decir, la terminología propia de la empresa y del tipo de textos que maneja) sino también su actualización conforme evolucionan los términos a lo largo de la vida de la empresa,

1 Durante años la TA ha sido un área reservada únicamente a la investigación y muy pocos han sido los sistemas desarrollados con fines comerciales. En la última década, se ha empezado a mirar al mercado y a los usuarios finales de la TA.