

CASO PRÁCTICO DE CREACIÓN DE UN GLOSARIO TERMINOLÓGICO TRILINGÜE (INGLÉS-ESPAÑOL-ALEMÁN) SOBRE MINERÍA DE DATOS

JOSÉ LUIS LLAVONA ARREGUI

*Departamento de Lingüística Aplicada a la Ciencia y a la Tecnología. Universidad Politécnica
de Madrid, España*

Esta comunicación revisa el proceso de creación de un glosario terminológico sobre minería de datos en inglés, español y alemán. Los términos fueron recogidos en fichas terminológicas normalizadas en las que se incluyeron aquellos campos que permitieran una referencia más rápida para los futuros usuarios, ya fueran técnicos o académicos. Los resultados finales muestran cómo en la mente del técnico prima la precisión del término, el hecho de que refiera más claramente al concepto que designa, por encima de consideraciones estéticas, lingüísticas o de cualquier otro tipo.

Palabras clave: glosario terminológico, minería de datos, normalización, especialista.

1 Introducción

En el presente trabajo se describe el proceso de creación de un glosario terminológico sobre minería de datos en inglés, español y alemán. La elección del tema se debe a la inexistencia constatada de glosarios terminológicos sobre la materia (ni en España, ni en Alemania), dado lo reciente de su aparición como especialidad informático-estadística de análisis de información oculta en grandes bases de datos. En cuanto a las lenguas de trabajo, el inglés es el origen de la totalidad de la literatura existente sobre la materia; el español es la lengua materna del autor y ofrece un número considerable de usuarios potenciales del glosario resultante, y, por último, el alemán permite contrastar los procesos de incorporación de nueva terminología a una lengua no-románica con tradición en la producción propia de términos técnicos.

Este estudio nos da la oportunidad de analizar la capacidad de absorción y producción propia de términos de ambas lenguas, y, en nuestro caso, de comprobar los procesos que siguen los especialistas al enfrentarse por primera

vez a la necesidad de incorporar a la docencia universitaria en sus respectivas lenguas maternas (español en el caso de la Dra. Menasalvas de la Universidad Politécnica de Madrid y Alemán en el del Dr. Lattner, de la Universidad J. W. Goethe de Frankfurt am Main) unos términos que han aprendido en su forma original inglesa. Por razones de índole práctica se optó por la elección de un equipo de trabajo formado por un terminólogo (el autor) y dos técnicos. El resultado fue posteriormente revisado por un terminólogo de reconocido prestigio, el Doctor Klaus-Dirk Schmitz, de la Universidad de Colonia (1).

2 Metodología

El proceso comenzó con la búsqueda de los glosarios monolingües en inglés ya existentes. La fuente de referencia obligada –una vez constatado que no existía un glosario semejante en el ámbito académico- eran las publicaciones de las principales empresas dedicadas a la minería de datos. De este modo se obtuvieron cuatro glosarios de partida, (por orden cronológico de publicación: IBM 1997, Clementine 2002, Two Crows Corporation 2006 y Kurt Thearling 2008) que recogen una amplia selección de lo que estas cuatro empresas consideran terminología imprescindible de minería de datos. El hecho de que la publicación de estos glosarios se extendiera en un plazo de diez años permitiría, además, realizar un pequeño estudio diacrónico de alguno de los términos.

Estos cuatro glosarios fueron cruzados para evitar repeticiones, y el listado de

(1) Modelo organizativo “D” (terminology centred) según establece la Norma ISO 15188.

candidatos a término resultante (2) fue sometido al análisis de los técnicos (los citados doctores Menasalvas y Lattner) de los países de las lenguas de destino, español y alemán. El listado incorporaba, además, una definición en inglés y una propuesta de traducción al español. Como referencia para los especialistas, se les facilitó la siguiente orientación metodológica sobre cómo proceder ante las propuestas de entradas al glosario:

- Sólo deberían considerar como “términos” aquellas expresiones de una o más palabras cuyo uso y comprensión resultara imprescindible para transmitir los conocimientos teóricos propios de la minería de datos.
- Denominarían “semi-términos” a aquellas palabras y expresiones de otros campos paralelos (fundamentalmente matemáticas e informática) cuyo uso fuera imprescindible para precisar conceptos relativos a procesos o técnicas propios de la minería de datos.
- Las demás propuestas del listado que no cumplieran con los dos requisitos mencionados anteriormente debían ser descartadas.

Los términos aceptados fueron recogidos en fichas terminológicas normalizadas (3) en las que, una vez incorporadas las indicaciones y sugerencias de los doctores Menasalvas y Lattner, se incluyeron los siguientes campos:

.Término en inglés, español y alemán.

.Definición en inglés y en español, con referencia a temas afines.

(2) No podemos denominar a las palabras de este listado previo términos stricto sensu, puesto que la percepción del concepto “término” difiere sensiblemente en función de que lo considere como tal un académico o un ejecutivo. La empresa precisa comunicarse con clientes no-especialistas, con lo cual su enfoque es notablemente más laxo y carece en ocasiones de rigor científico.

(3)Según Norma ISO 12620.

- .Sinónimos en las tres lenguas.
- .Área de conocimiento de la que procede el término.
- .Apartado de minería de datos a que pertenecen(4).
- .Ejemplo de uso del término en un texto del corpus.
- .Posición del concepto en el área de conocimiento.

Evidentemente, esta relación no comprende la totalidad de los apartados recogidos en la Norma ISO 12620. Como ya indicábamos anteriormente, el espíritu práctico que rige este trabajo nos obliga a centrarnos en aquellos apartados que consideramos más relevantes para los usuarios finales del glosario resultante: los técnicos en minería de datos. En otras palabras, se renuncia a un análisis lingüístico más profundo en aras a conseguir lo que en informática se llamaría un “entorno más amigable” para los destinatarios del producto final. Estimamos que la ficha resultante recoge la totalidad de la información que un técnico en minería de datos precisa a la hora de ubicar con precisión un término en su área de conocimiento.

El trabajo de revisión del listado por parte de los especialistas se desarrolló en dos fases, primero en Madrid en el Campus de Montegancedo con la Doctora Menasalvas y una segunda en Frankfurt am Main, Alemania, en el Campus de Bockenheimer con el Doctor Lattner. No podemos por menos que recordar aquí que ambos profesores están considerados entre las máximas autoridades en materia de minería de datos en el mundo académico de sus respectivos países

(4) Los especialistas distinguen hasta siete apartados distintos en el proceso de minería de datos: Datos, Proceso, Técnicas y Algoritmos, Tipos de problemas, Tipos de Resultados, Parámetros de Evaluación y Genérico.

y sin su colaboración y ayuda este trabajo no habría sido posible. En ambos casos ratificaron la conveniencia de contar con glosarios sobre la materia en sus lenguas maternas y la oportunidad que el proyecto suponía de cubrir un vacío evidente.

En la primera fase le fue presentada a la Dra. Menasalvas una relación inicial de fichas que incluían los términos con su denominación original en inglés, la correspondiente definición en este mismo idioma y una posible traducción al español del término y su definición. A lo largo de una serie de reuniones de trabajo fueron revisadas las fichas una por una, obteniendo los siguientes resultados:

- La propuesta de traducción (literal) del término fue modificada en la mayoría de los casos.
- La propuesta de traducción (también literal) de la definición del término fue alterada todos los casos bien parcial o totalmente, siendo añadidas referencias a otros sub-apartados y precisiones aclaratorias.
- la definición original en inglés fue mejorada.
- El campo correspondiente a sinónimos quedó desierto en aproximadamente el 75% de los casos.
- A los campos inicialmente propuestos se incorporó uno nuevo, el correspondiente a la ubicación del término en su correspondiente sub-apartado dentro del campo general de la minería de datos.
- Se completó el campo referente al origen del término.

-Se completó el campo correspondiente a la posición del concepto en el área de conocimiento de procedencia.

La doctora Menasalvas constató la ausencia de algunos términos imprescindibles en el listado propuesto, fundamentalmente los relativos al proyecto CRISP DM, razón por la cual y siguiendo sus orientaciones de documentación se procedió a incorporar el glosario Clementine.

El mismo método, basado en reuniones de trabajo periódicas en el despacho habitual del técnico, se siguió en Alemania con el Doctor Lattner a lo largo de cuatro meses. En esta fase del proyecto el volumen de información de cada ficha era mucho mayor que el presentado a la Dra. Menasalvas, lo que nos permitiría añadir, a parte de la información lingüística, una constatación técnica de los contenidos teóricos que presentaban los campos originales y los recién incorporados.

Los resultados de estas sesiones de trabajo con el Dr. Lattner son los siguientes:

- Se completó el campo correspondiente a equivalencia de cada término en alemán. En este caso no había sugerencia previa de traducción.
- Se revisaron y mejoraron las definiciones originales en inglés.
- A sugerencia del Dr. Lattner se incorporó en algunos casos una anotación a la definición, correspondiente a referencias imprescindibles a temas relacionados.

-La sinonimia en alemán subió por encima del 30%. En algunos casos se trataba del original en inglés, en otros era una expresión equivalente en alemán al término preferido, el original inglés, y en un número significativo de términos se planteaba una opción de término “en proceso de asimilación”, con una grafía alemana que representaba el original en inglés (sobre este punto se incide más extensamente en el apartado de conclusiones).

-Hubo una coincidencia próxima al 100% con los planteamientos de la Dra. Menasalvas en cuanto a ubicación de cada término en su campo correspondiente de minería de datos.

-Hubo una coincidencia próxima al 95% en cuanto al origen de cada término. En los casos en que hubo diferentes opiniones prevaleció la del técnico con más años de experiencia, es decir, la Dra. Menasalvas.

-Al contar con más datos (ejemplos en contexto, número de glosarios de partida en los que contaba cada término y frecuencia de aparición de cada término en los textos del corpus) se pudo realizar una profunda revisión técnica de las fichas en su conjunto que permitió realizar algunos descartes (véase apartado de conclusiones).

Simultáneamente al trabajo con los técnicos se realizó un análisis estadístico de frecuencias de uso de los términos en textos propios de la especialidad que avalara objetivamente su opinión. Si se trataba efectivamente de términos propios del campo, deberían aparecer frecuentemente y en posiciones preponderantes en dichos textos.

Con este fin, se seleccionó un corpus sobre minería de datos de una extensión aproximada de 200.000 palabras que comprendiera las diversas temáticas propias de la materia. Los textos que forman el corpus proceden tanto de fuentes académicas (documentación de referencia para alumnos universitarios) como de guías de empresas que proporcionan servicios de minería de datos a sus clientes.

La localización, recuento y consiguiente validación de los términos se realizó por medio del análisis con herramientas informáticas (5) de los textos del corpus. Estos programas fueron utilizados para conseguir los siguientes objetivos:

.Localizar el término buscado en su contexto de uso, para obtener ejemplos que ilustraran el campo correspondiente en cada ficha.

.Obtener los porcentajes aparición del término en la totalidad del corpus. A mayor porcentaje, mayor certeza de pertenencia al campo de la minería de datos.

La razón por la que se emplearon dos analizadores, uno de textos españoles y otro de textos ingleses está en que ambos se complementan para conseguir un análisis más fidedigno dadas las características de este tipo de programas.

El empleo de uno sólo mostró señales evidentes de error en los datos

(5) Analizadores de Textos Trobes Version 6.2.4 (build 0013) de Acetic (español) y Zoom Version 7.0.0 (build 0041) de Acetic (inglés)

porcentuales de frecuencias de aparición por término, dado que sólo reconoce nombres o grupos nominales; ello daba lugar a que términos como *sampling* no fueran localizados por la versión en inglés.

De este modo, podemos resumir el proceso diciendo que los datos obtenidos a partir de los glosarios de partida fueron filtrados en primer lugar por el criterio de los técnicos y que el producto resultante de esta etapa fue finalmente validado mediante el análisis estadístico de la frecuencia de aparición de cada término en el corpus de referencia. El glosario así obtenido fue nuevamente sometido a la consideración de los técnicos, quienes dieron su visto bueno al resultado final.

El proyecto fue sometido en su etapa final a la supervisión de un terminólogo de constatada experiencia, el Profesor Dr. Klaus-Dirk Schmitz, de la Universidad de Colonia. El Dr. Schmitz, además de su dilatada experiencia en la dirección de proyectos terminológicos, es Presidente del Consejo de Terminología Alemana (RaDT) y Presidente del Centro Internacional de Información para la Terminología (Infoterm).

Una vez presentado el proyecto al Dr. Schmitz se incorporaron las siguientes mejoras:

- En el apartado de definiciones de los términos sugirió que se hiciera una depuración y mejora por parte de los técnicos.

- La denominación “semi-término” debía ser eliminada; cualquier expresión cuyo uso sea imprescindible para explicar algún apartado de minería de datos debe ser considerada término, aunque proceda de otro campo.
- En la incorporación de nuevos términos al alemán indicó que debería establecerse el género correspondiente de los sustantivos.
- En la sección correspondiente a origen del término, indicó la conveniencia de introducir la nota “subject field 1”, dado que es la forma estandarizada de referirse a este campo en las fichas terminológicas.
- La ubicación del término en el apartado correspondiente de minería de datos debería ser referida como “subject field 2”, por las mismas razones indicadas en el punto anterior.
- Estableció como suficiente la extensión del corpus de referencia, (200.000 palabras), dadas las características del campo a investigar. La calidad de los mismos debería primar sobre su número.
- No estimó necesario incluir la frecuencia de aparición por texto de cada término.
- Como punto final, dio su visto bueno a los procedimientos de documentación, análisis técnico y tratamiento informático de los términos que constituyen los pilares de este proyecto.

El glosario resultante quedaba así certificado.

Conclusiones

En este proyecto terminológico hemos seguido, según mencionábamos en el apartado anterior, el modelo que establece la Norma ISO 15188, optando por un equipo de trabajo formado por un terminólogo y dos técnicos y sometiendo los procedimientos y resultados a la evaluación de un terminólogo de reconocida experiencia. Hay datos referentes a los técnicos que pueden resultar relevantes a la hora de interpretar los resultados: los doctores Menasalvas y Lattner se enfrentaban por primera vez a un proyecto semejante; no había por su parte ideas preconcebidas en cuanto a proyectos terminológicos. Su visión pragmática sobre lo que una ficha terminológica debía aportar a quien consultara el glosario contribuyó a la mejora del mismo. Como consecuencia de sus comentarios y sugerencias se añadió a la ficha la entrada “subject field 2” (denominada en principio *posición del término en minería de datos*), en la que se ubica el término en su ámbito correspondiente dentro de la minería de datos. Del mismo modo se incluyó en los casos más relevantes la nota “Términos Relacionados” como referencia necesaria para una mejor comprensión del concepto.

Ambos profesores tienen un elevado dominio de la lengua inglesa, tanto oral como escrito; el uso del inglés es constante en su vida académica, como medio de acceso a las fuentes de documentación originales y por ser el la lengua de comunicación habitual con sus colegas de especialidad. El idioma de origen no supuso una barrera en absoluto; bien al contrario, es el idioma vehicular de su especialidad. Aun así, el recurso al préstamo se da sólo en un 11% de los

casos en español y en un 17% en alemán (el mayor porcentaje es comprensible en este caso al tratarse de dos lenguas germánicas).

El recurso al préstamo se produce únicamente en aquellos casos en que una solución alternativa (habitualmente la traducción préstamo) podría suponer un obstáculo en la comprensión del concepto; consideramos que en una fase tan temprana de la evolución de la terminología de esta área de conocimiento no resulta extraño que algunos conceptos clave conserven su término original.

Pero la función de los técnicos no se limitó a proporcionar el término final y ubicarlo en su área de conocimiento: todas las definiciones procedentes de los glosarios de partida fueron depuradas con su concurso, dotando de una perspectiva académica a lo que eran mayoritariamente definiciones para profanos.

Pero al igual que el terminólogo debe permanecer en su puesto de fedatario de aquellos términos que se emplean en el área de conocimiento objeto de estudio (de la que no es por principio especialista), el técnico no debe invadir el territorio del lingüista. La apreciación del técnico está limitada fundamentalmente porque su aproximación al concepto es funcional: conoce el objeto y lo ubica en un sistema en el que los elementos interactúan en una dinámica problema-resultado. De este modo, en varias ocasiones (y con ambos técnicos) al intentar este autor conseguir una precisión mayor al definir un término, o su inclusión en un sistema jerárquico, la respuesta fue la misma: un diagrama en el que se mostraba la relación funcional de los términos. Al insistir

sobre la conveniencia de establecer una taxonomía que facilitara la comprensión a los usuarios del glosario, se apreciaba una cierta reacción de perplejidad y el comentario era el mismo: nunca se lo habían planteado desde ese punto de vista. La definición por fórmulas y diagramas les resulta familiar y en ella se sienten cómodos. Trasladar esa información gráfica a texto escrito es una labor para la que el concurso del lingüista es a veces imprescindible.

El técnico se plantea la inclusión o no del término en el glosario a partir, fundamentalmente, de la relevancia del concepto en el campo de su especialidad y la frecuencia con que se emplea dicho término. Es la perspectiva que proporciona su propia experiencia. Mientras que para los términos clave esa opinión es más que autorizada, en aquellos casos en que el término es compartido por áreas colindantes (fundamentalmente estadística e informática en nuestro caso), es preciso contar con una información de la que el técnico carece: la que proporciona el tratamiento informático del corpus de referencia con el que sí cuenta el terminólogo.

Hay dos ejemplos que ilustran perfectamente lo que decimos: las entradas *data warehouse* y *overtraining*. En el primero de los casos había una fuerte resistencia por parte de ambos técnicos a incluirlo en el glosario final; alegaban que era un error frecuente, y que la minería de datos era independiente del *data warehousing*. De este modo en principio fue descartado del glosario, pero el análisis del corpus reveló que era un término presente en un 41,6% de los textos de referencia... ¿cómo prescindir de él? Acabó por ser aceptado.

En cuanto a *overtraining*, el recorrido del término fue el inverso: pasó de estar admitido originalmente como término “de pleno derecho” a ser considerado un sinónimo del verdadero término: *overfitting*. La razón estriba en dos datos de los que los técnicos, en principio, carecían: sólo uno de los glosarios de referencia (el más antiguo) lo incluía, y estaba presente en apenas uno de los textos del corpus. El término que designaba el mismo concepto -*overfitting*-, estaba presente en dos de los glosarios y se citaba en un tercio de los textos del corpus. La conclusión a la que se llegó fue que *overtraining* era el término más antiguo, el usado por IBM, pero que había terminado por caer en desuso. Las frecuencias de aparición de los términos en el corpus determinaron la inclusión o no en el glosario final de los casos dudosos: los que presentaban unos porcentajes más bajos (casos de *prevalence*, *resubstitution error* o *quantile*) quedaron descartados.

Así pues, los cuatro glosarios de partida y el corpus de referencia supusieron la base objetiva del proyecto. Pero no debemos extraer conclusiones erróneas del hecho de contar con listados previos de vocabulario propio de la especialidad (cabría pensar que bastaba hacer un traslado de aquellas listas a la nuestra de destino): sólo hay ocho términos del glosario final que estén presentes en los cuatro listados de partida, y, si contamos la totalidad de los que propone cada listado, tendríamos aproximadamente doscientos cincuenta. Los 122 términos del glosario final son el resultado de un largo proceso de selección, depuración y análisis. Llegados a este punto conviene recordar que los términos no están sujetos a copyright, pero los listados y las definiciones sí lo están. En este proyecto no se han solicitado licencias de reproducción porque para el glosario

resultante se tomaron como referencia las definiciones de los listados de partida, pero las que han quedado finalmente son las aportadas por los técnicos, con la mejora considerable que ello supone. En cuanto a los textos del corpus, son todos de libre acceso. Su selección obedece a criterios de calidad (en su mayor parte son obras de autoridades reconocidas como Piatetsky-Sapyro, Kurt Thearling...etc) y a su directa relación con la materia objeto del corpus. Se optó por un corpus de 200.000 palabras al comprobar que aumentar su número podría ir en detrimento de la calidad del contenido. Por otra parte, dado el alto nivel de especialización de la materia y su reciente aparición, la literatura existente no es muy abundante.

Pero este corpus por sí mismo no nos habría servido de mucho sin las herramientas informáticas necesarias para extraer la información que en él se guarda. El analizador de textos supone una ayuda imprescindible en el trabajo del terminólogo. Estimamos que las tareas de localización de términos y cálculo de frecuencias de aparición empleando estos programas supusieron más de 300 horas de trabajo en este proyecto; es difícil calcular el tiempo que hubiera sido necesario para realizar las mismas tareas "a mano".

El analizador de textos empleado permite localizar términos de una o varias palabras, así como indexar una base de datos a partir de varios archivos de texto. De este modo fue posible establecer las frecuencias de aparición de cada término.

Los resultados que aquí se presentan no son extrapolables a otros campos colindantes, como la estadística y la informática. Se deben analizar siempre teniendo en cuenta que nos encontramos ante una especialidad extremadamente joven (6) y con un ámbito de difusión muy restringido.

De entre los datos que nos muestran las tablas de resultados obtenidas cabría destacar notablemente el hecho de que la práctica totalidad de los términos en español y alemán son préstamos o traducciones préstamo. Concretamente en español hay un 13% de préstamos, y el 87% restante lo constituyen traducciones préstamo (con la única excepción de “datos atípicos” por *outliers*, lo cual no deja de ser anecdótico: el dato atípico del glosario es el que designa este concepto en minería de datos). En alemán el porcentaje de préstamos sube al 21,5%, pero el porcentaje de traducciones préstamo, sin dejar de ser muy elevado –un 72%–, es algo menor, puesto que en un 6,5% de los casos crea un término propio (*Beispiel, Spalte, Anwendungsphase, Abweichungserkennung, Merkmal, Artikel, Muster, Zielattribut*) para designar el concepto correspondiente. En la mayoría de los casos esto se debe a que el alemán carece de la palabra de origen latino que dio lugar al término en inglés. En ambas lenguas y, según suele ser habitual en otros campos, los términos expresados por medio de siglas y acrónimos permanecen invariables (CART, CHAID y CRISP DM).

También hemos podido constatar el hecho de que los porcentajes de sinonimia

(6) La fecha comúnmente aceptada de la aparición del término Data Mining es 1996, con la publicación en el número de otoño de Artificial Intelligence Magazine del artículo From Data Mining to Knowledge Discovery in Databases, de Usama Fayyad, Gregory Piatetsky-Shapiro y Padhraic Smyth.

son relativamente bajos. Sólo un 29% de los términos en español tienen un sinónimo (el propio término en inglés en dos casos), y este porcentaje sube apenas al 35% en el caso del alemán (el término en inglés en 18 casos). Al tratarse de una ciencia exacta este dato no es infrecuente.

Lo que sí resulta especialmente llamativo es el proceso de asimilación al alemán de algunos términos en inglés. Para centrar esta cuestión debemos comenzar por comentar que en los últimos años la lengua alemana ha adoptado como propias múltiples palabras de origen inglés. De este modo, al libros de ejercicios se le llama “das Trainer”, el teléfono móvil es “das Handy”, el ordenador “der Computer”, la camiseta “das T-shirt”, y un bebé “ein baby”... por citar algunos ejemplos de la vida cotidiana. Como podemos observar, en el proceso se da la correspondiente asignación de género y escritura con mayúscula de los sustantivos, indispensable en alemán; además, la pronunciación de la palabra se mantiene como en el original inglés. Según datos obtenidos por medio de una encuesta realizada entre varios profesores de la *Volkshochschule* y la Universidad J. W. Goethe de Frankfurt, la permeabilidad del alemán ante los anglicismos es elevada y el hablante común acepta con agrado su uso.

En nuestro estudio este proceso se ve reflejado en que en 26 casos de sustantivos el término original inglés se ha incorporado al alemán manteniendo su pronunciación inicial pero con grafía en mayúsculas y género propio. Esta atribución de género no es aleatoria: los masculinos y femeninos se asignan en

función de su equivalente en alemán, y el género neutro se utiliza en aquellos términos que designan proceso.

Es también especialmente llamativo el hecho de que los términos originales hayan variado tan poco al incorporarse a la lengua de destino: un caso en español y ocho en alemán. Esto viene a corroborar la teoría que hemos sostenido a lo largo de este trabajo: en la mente del técnico prima la precisión del término, el hecho de que refiera más claramente al concepto que designa, por encima de consideraciones estéticas, lingüísticas o de cualquier otro tipo.

Bibliografía

Arntz, R., Picht, H. (1995) *Introducción a la terminología*. Madrid. Fundación Germán Sánchez Ruipérez . Pirámide. (Traducción de 1989 *Einführung in die Terminologearbeit*, Hildesheim, Georg Olms).

Cabena, Hadjinian, Stadler, Verhees y Zanasi (1997) *Discovering Data Mining From Concept to Implementation*. Upper Saddle River, NJ. Prentice Hall.

Cabré, M.T. (1993) *La Terminología. Teoría, Metodología, Aplicaciones*. Barcelona. Ed.Antártida/Empuries.

Felber, H. y Pitch, H. (1984) *Métodos de Terminografía y Principios de Investigación Terminológica*. Madrid. Instituto "Miguel de Cervantes". CSIC

Rirdance, S. y Vasiljevs, A. (2006) *Towards Consolidation of European Terminology Resources. Experiences and Recommendations From EuroTermBank Project*. Riga. Tilde.

Ryszad S. Michalsky, Ivan Bratko, Miroslav Kubat (Eds) (1998) *Machine Learning and Data Mining: Methods and Applications*. Wiley

Sonneveld, Helmi B. y Loening, Kurt L. (1993) *Terminology; Applications in Inter-disciplinary Communication*. Amsterdam/Philadelphia. John Benjamins Publishing Co.

TSS 2006. Terminology Summer School . Escuela de Traducción. Universidad de Viena.