

NOMBRES de los AUTORES:

Diego Ferreyra

INSTITUCIÓN:

Universidad de Buenos Aires. Facultad de Filosofía y Letras. Departamento de Bibliotecología y ciencias de la información

DIRECCIÓN:

Puán 470 3er piso - Of. 335

EMAIL:

diego@r020.com.ar

TÍTULO del trabajo:

Metodología para el desarrollo de vocabularios multilingües*

RESUMEN del trabajo: Se presenta una experiencia orientada al análisis de viabilidad técnica y procedimental de una metodológica para la generación de vocabularios multilingües utilizando un vocabulario controlado monolingüe como punto de partida y vocabulario base. Se describirán los pasos y procedimientos realizados para el desarrollo e implementación del vocabulario multilingüe de géneros de ficción utilizando como base la versión ya existente en español para desarrollar sus versiones en inglés, francés e italiano. Asimismo se ha explorado el uso de estrategias de comunicación entre aplicaciones basadas en servicios web para la formalización de mapeos terminológicos entre distintos vocabularios controlados. En la experiencia se han articulado diferentes tecnologías, unas orientadas a la portabilidad de metadatos y otras orientadas establecer esquemas de comunicación entre herramientas de gestión de vocabularios controlados en condiciones de consistencia e integridad.

Presentación

Este trabajo presenta una experiencia orientada al análisis de la viabilidad técnica y metodológica del marco de armonización de vocabularios para la generación de vocabularios multilingües tomando un vocabulario controlado en una lengua como vocabulario base.

Los procesos de armonización de vocabularios controlados pretenden establecer relaciones de correspondencia terminológica y conceptual entre distintos sistemas formales de clasificación. Estos sistemas de clasificación permiten caracterizar, representar, organizar y acceder a repositorios de recursos descriptos a través de metadatos. La exposición de los mencionados metadatos a través de protocolos de comunicación en el contexto de la WWW permite desarrollar servicios de explotación cooperativa de contenidos.

Se realizará una presentación de los sistemas formales de clasificación como contexto conceptual y de los metadatos como marco técnico. Luego se describirán las instancias metodológicas y

* El resultado del presente trabajo se encuentra publicado en el sitio <http://www.vocabularyserver.com/ficcion/> y forma parte de la memoria del trabajo final realizado en el marco del *Master II professionnel sur la «Gestion de l'Information et de la Connaissance* de la *Université Paul Valéry - Montpellier III*.

tecnológicas utilizadas en el proceso de definición y formalización de las correspondencias terminológicas y conceptuales y, finalmente, se describirá el proceso de elaboración e implementación realizado. Para el desarrollo de la experiencia se han articulado diferentes tecnologías, unas orientadas a la portabilidad de metadatos y otras a la de vocabularios controlados. Se desarrolló e implementó un vocabulario multilingüe de géneros de ficción utilizando como base la versión ya existente en español para desarrollar sus versiones en inglés, francés e italiano. El resultado del presente trabajo se encuentra publicado en el sitio <http://www.vocabularyserver.com/ficcion/> y forma parte de la memoria del trabajo final realizado en el marco del *Master II professionnel sur la «Gestion de l'Information et de la Connaissance de la Université Paul Valéry - Montpellier III*.

Los sistemas formales de clasificación

La clasificación, en su acepción más amplia, puede considerarse en tanto la segmentación espacio-temporal de la experiencia: un sistema de clasificación es un conjunto de cajas en donde dividir y poner el mundo o nuestra experiencia. Esta segmentación, analizada en tanto dispositivo social, es considerada por Foucault ([1966] 2008) como un elemento constitutivo de lo social como posibilidad, como modelo de orden y categorización de las prácticas. Bowker y Star (1999), realizan un estudio orientado a analizar la dinámica de operación de los sistemas de clasificación en diferentes contextos y escalas a través del estudio de distintos vocabularios controlados y taxonomías. Al respecto, los autores consideran que los sistemas formales de clasificación son *artefactos* en el sentido de que son *un aspecto del mundo material que puede ser creado y modificado socialmente y que tiene la particularidad de ser simultáneamente ideal (conceptual) y material¹*, podría decirse que son artefactos lingüísticos con claras consecuencias materiales y físicas.

En esta línea señalan a los sistemas formales de clasificación como un tipo particular de artefacto, concretamente, *artefactos mediadores* caracterizados por su capacidad para:

- a) dotar de comparabilidad semántica a través de distintos contextos;
- b) establecer condiciones de visibilidad presentando categorías (definir lo visible y lo invisible: lo no categorizable es invisible); y
- c) controlar los procesos de significación definiendo el alcance conceptual de términos, categorías y clases.

En tal sentido, los sistemas formales de clasificación son artefactos mediadores que permiten llevar adelante procesos de circulación discursiva a través de diferentes contextos, diferentes comunidades de prácticas o a través del tiempo.

El estudio de este tipo de artefactos de mediación permite analizar las condiciones socio-técnicas en las que fueron creados a través de la explicitación de los mecanismos de apropiación y relación entre contextos que pone en práctica, ya sean contextos ideológicos, institucionales, informales (comunidades de prácticas) o relativos a la apropiación del pasado. Se trata de una tarea que no resulta trivial, en tanto muchas veces el éxito de este tipo de artefactos, según Bowker y Star (1999), puede ser evaluado según su capacidad para recorrer el arco que va de la normalización a la naturalización, logrando condiciones de invisibilidad que permiten maximizar su eficacia.

1 Bowker, G. C., & Star, S. L. (1999). *Sorting things out Classification and its consequences*. Cambridge, Mass: MIT Press. p. 289

Con respecto a los modelos de análisis posibles para abordar los sistemas formales de clasificación en contextos sociales, Star y Ruhleder (1996) proponen un acercamiento que pretende aportar herramientas que permitan analizar las producciones simbólicas en tanto elementos materiales capaces de intervenir en transformaciones materiales concretas, al igual que un locomotora o un puente, y considerar, como se dijo antes, que las clasificaciones son a la vez simbólicas y materiales². Las autoras proponen utilizar la noción de *infraestructura* para analizar los sistemas formales de clasificación. Según su perspectiva, una infraestructura es susceptible de ser reconocida a partir de las siguientes propiedades:

- Relación embebida: se encuentra inserta dentro de otras estructuras sociales, tecnológicas, etc.
- Transparencia: la infraestructura resulta transparente para el uso, a su vez, el soporte que brinda a otras tareas o estructuras resulta invisible.
- Ámbito de aplicación: tiene un alcance espacial o temporal que va más allá de un sólo evento o de una práctica aislada
- Aprendidas en el proceso de inclusión: la adopción y uso de una determinada infraestructura es una condición sine qua non para la pertenencia a una determinada comunidad de práctica.
- Que guarda relaciones entre convenciones de la práctica: la infraestructura tiene una naturaleza dual, constituye y es constituido por las convenciones de una comunidad de prácticas.
- Articulación con normativas: la infraestructura se articula de manera transparente a través de estándares y normas con infraestructuras y herramientas.
- Que está consolidada sobre una base previa: las infraestructuras no son creadas desde cero o como algo nuevo en su totalidad, siempre se articulan sobre bases de infraestructuras ya instaladas.
- Que se hace visible cuándo falla: el accionar normalmente invisible de la infraestructura se torna visible cuándo falla (ej: un apagón o la caída de un sistema).
- Que cambia de manera gradual: debido que la infraestructura resulta un estructura articular y compleja, provista de distintos significados locales, sus transformaciones no son globales ni tampoco se presentan como una cambio único y total. Sus transformaciones requieren tiempo, negociación y proceso de ajustes con respecto a los demás actores y sistemas involucrados.

A partir de las características propuestas es posible considerar a los sistemas formales de clasificación en tanto una infraestructura omnipresente en todos los procesos que requieren de representaciones formales. Detrás y por delante de todo proceso formal existe y operan uno o más sistemas formales de clasificación.

El presente trabajo está centrado en un tipo específico de sistemas formales de clasificación denominados vocabularios controlados. Dichos sistemas se diferencian de otros esquemas de clasificación (ej: clasificación por color) por estar constituidos por términos lingüísticos y por la delimitación lógica y semántica de las relaciones posibles entre los mencionados términos lingüísticos considerados como entidades de representación conceptual.

2 Bowker, G. C., & Star, S. L. (1999). *Sorting things out Classification and its consequences*. Cambridge, Mass: MIT Press. p. 289

Los vocabularios controlados como infraestructura

A continuación se realizará una caracterización de los tipos de vocabularios controlados estableciendo un marco de diferenciación según las relaciones semánticas que es posible establecer y formalizar entre sus términos. Se utilizará el marco propuesto por la norma Z39.19³ para la diferenciación y denominación de los distintos tipos de vocabularios controlados.

La norma considera que la eficacia de un vocabulario controlado mantiene una relación directa con el cumplimiento de ciertas capacidades consideradas en tanto *principios de consistencia*. Dichas capacidades son las siguientes:

1. Control de ambigüedad: la capacidad de vocabulario para resolver, identificar y representar problemas de homografía y polisemia entre términos.
2. Control de sinonimia: la capacidad del vocabulario para establecer relaciones de equivalencia entre términos.
3. Relaciones semánticas: la capacidad del vocabulario para establecer relaciones de asociación recíprocas (ej: conceptos implicados) o asimétricas (ej: inclusión, jerárquicas) entre términos.
4. Condiciones de validación: la relación que establece un vocabulario con respecto a las prácticas lingüísticas en la literatura especializada; las prácticas lingüísticas de la comunidad de usuarios; y con las políticas lingüísticas de la institución o de los objetivos del vocabulario.

La clasificación en ambientes digitales

El uso de tecnologías basadas en la digitalización de datos ha generado importantes transformaciones en las prácticas sociales en general y en particular en las metodologías y concepciones aplicables a la organización y representación de bienes culturales y simbólicos.

Ya a fines de las década del 70, Lyotard ([1979],1993) consideraba la emergencia de la digitalidad como un punto de inflexión en la construcción del *saber*: “ ... al normalizar, miniaturizar y comercializar los aparatos, se modifican ya hoy en día las operaciones de adquisición, clasificación, posibilidad de disposición y de explotación de los conocimientos.”⁴ Al menos desde entonces, resulta relevante analizar las transformaciones ocurridas en la naturaleza de la noción de saber y los procesos de circulación discursiva a partir de la emergencia de la digitalidad.

Asimismo, los medios electrónicos han posibilitado el desarrollo de formas y ritmos de producción de información que no resultan siquiera comparables a etapas anteriores y por lo tanto no pueden ser siquiera evaluados en términos de crecimiento con respecto a éstas. Tanto en la web como a nivel privado e institucional, el volumen de información ha cambiado en términos de orden de magnitud, tal como menciona John Gillis (1994) en relación con la voluntad de información y la multiplicación de los *actos del recordar*, “en la actualidad, cada ático es un archivo y cada living un museo. Nunca antes hubo tanto por registrar, coleccionar; y nunca antes el acto de recordar fue tan compulsivo”⁵. Esta nueva dimensión en términos de magnitud, se ve a su vez incrementada ya que,

3 : National Information Standards Organization (U.S.). (2005). Guidelines for the construction, format, and management of monolingual controlled vocabulary. National information standards series. Bethesda, Md: NISO Press

4 Lyotard, Jean François ([1979], 1993). La condición postmoderna: informe sobre el saber. Barcelona: Planeta-Agostini. p. 14-15

5 Bowker, G. C. (2006). Memory practices in th science. Cambridge, Mass: MIT Press. p 15

tal como señalan Weinberger (2007) y Lash (2005), las estrategias orientadas a generar marcos de diferenciación, descripción y organización de la información, requieren de la generación de más información.

En relación con las modalidades de circulación discursiva, la digitalidad ha incorporado nuevas capacidades y modelos para la organización de los bienes culturales. Estas transformaciones se explicarían, según Manovich (2003), básicamente a partir de ciertos atributos constitutivos de los medios electrónicos, a saber:

- Modularidad: los objetos tienen una estructura fractal. La misma estructura en diferentes escalas
- Automatización
- Variabilidad: los objetos disponen de múltiples versiones, varían en tamaño, resolución, forma y contenido. Personalización, actualización continua.
- Representación numérica: todo objeto debe poder ser descrito en términos numéricos y formarles.
- Transcodificación: los nuevos medios son continentes con dos capas: la cultural y la informática.

Las mencionadas características le permiten al autor analizar las bases de datos en tanto una nueva forma cultural que propone nuevas formas de representación del mundo: “As a cultural form, database represents the world as a list of items and it refuses to order this list. In contrast, a narrative creates a cause-and-effect trajectory of seemingly unordered items (events). Therefore, database and narrative are natural enemies. Competing for the same territory of human culture, each claims an exclusive right to make meaning out of the world.”⁶

Tal como se ve, es posible identificar la inquietud acerca de la emergencia de un nuevo régimen de orden, cuyas formas de legitimación y operación transformarían las prácticas sociales y, necesariamente, los procedimientos de clasificación y categorización, estableciendo preguntas acerca del futuro de los sistemas de clasificación en el medio digital. Al respecto, Weinberger (2007) ha desarrollado un análisis basado en la identificación de tres formas de orden y una periodización de las sucesivas tecnologías y prácticas asociadas a cada uno de estas formas, a saber: el orden físico, el orden de la escritura y el orden digital.

El orden físico estaría basado en la organización del espacio y los objetos; se correspondería con las los edificios, depósitos, almacenes, estanterías, las memorias espaciales. El orden de la escritura incorpora las capacidades que permite la separación entre el objeto y su representación, estaría basado en la organización a partir de la representación física del espacio y los objetos. Este segundo orden separaría el primer orden de los objetos mismos, se correspondería con los catálogos, las listas, los códigos, las clasificaciones, las memorias lineales. Finalmente, el tercer orden, el orden digital, separa el segundo orden de la linealidad del lenguaje. Incorpora las capacidades que permite una relación arbitraria entre las modalidades de creación, almacenamiento, representación y acceso a la información, se correspondería con las bases de datos, los objetos digitales, los ambientes virtuales, las clasificaciones personalizadas, las memorias no lineales.

6 Manovich, Lev (2003) *The language of new media*. Boston: MIT Press. p. 199

La conceptualización propuesta por el autor conlleva una pregunta acerca del estatus de la clasificación: ¿cuál sería la dinámica, el modelo y mecanismo de acción de los sistemas formales de clasificación en tanto dispositivo social en el contexto de la digitalidad?

Tal como plantea Manovich (2003), en parte las características constitutivas de la digitalidad estarían dadas a partir de la variabilidad y transcodificación de los signos. Estas capacidades permiten concebir nuevas formas culturales, como las bases de datos y sus nuevos regímenes de orden posibles que permiten la concurrencia y coexistencia de múltiples modelos de orden y clasificación. Esta sería la condición de posibilidad para el desarrollo de lo que Weinberger (2007) considera la emergencia del mencionado tercer orden, una forma de orden basada justamente en las posibilidades inauguradas por la relación arbitraria y no lineal entre representación y registro.

Sobre la base de estas capacidades, y considerando los sistemas formales de clasificación en tanto artefactos sociales de mediación, es que nos proponemos desarrollar una modelo de apropiación de recursos basado en un proceso de correspondencia terminológica entre sistemas formales de clasificación. Se trata de un proceso que apela a la problemática de la traducción en tanto procedimiento de apropiación. Como señala Rofes (2007), los medios digitales actualizan las problemáticas propias de la traducción a través del solapamiento de hábitat de significado locales como así también las tensión entre el hábitat global y la experiencia local.

Considerando que una de las funciones de los sistemas formales de clasificación, es intentar regularizar y moderar el movimiento de información de un contexto a otro, regulando el sentido y acceso a información a través de diferentes contextos⁷, y en tanto los vocabularios controlados son sistemas formales de clasificación, la armonización de vocabularios podría considerarse y utilizarse como un modelo de solución para lograr portabilidad de contenidos entre diferentes comunidades de interpretación.

Para el desarrollo de la experiencia se resolvió seguir los siguientes pasos:

- Selección de un vocabulario de base.
- Migración del vocabulario hacia una herramienta digital de gestión de vocabularios controlados.
- Creación de los vocabularios correspondientes en italiano, francés e inglés.
- Mapeo terminológico con otros vocabularios disponibles dentro del dominio conceptual.

A continuación se describen brevemente los pasos y marco metodológicos realizados.

Selección de un vocabulario de base

Se seleccionó el vocabulario controlado en español Tesauro de géneros de ficción elaborado por Mónica Gabriela Pené, atendiendo a las siguientes consideraciones:

- Se trata de un vocabulario controlado liberado a través de una licencia CreativeCommons, lo cual nos permite desarrollar obras derivadas.
- Siendo que la presente experiencia se encuentra orientada a explorar las posibilidades y problemáticas propias del marco de armonización y mapeo terminológico, este vocabulario

7 Bowker, G. C., & Star, S. L. (1999). *Sorting things out Classification and its consequences*. Cambridge, Mass: MIT Press. p. 290

controlado ofrece una escala de experimentación óptima debido al acotado número de términos y la especificidad de la temática

- La formalización de un modelo de representación formal en la definición de los géneros literarios constituye un campo diferenciado dentro de los procesos de clasificación asistida por medios informáticos ya que la posibilidad de categorización de un texto en función de su pertenencia a un género literario es un procedimiento fuertemente condicionado por factores contextuales cuya legibilidad sólo es posible a la luz de la literatura en tanto sistema de relaciones textuales.

Migración hacia una herramienta de gestión de vocabularios controlados

Para la gestión del vocabulario controlado se utilizó la herramienta TemaTres. Más allá de las posibles facilidades o capacidades de la herramienta, la misma fue seleccionada en virtud de las siguientes capacidades diferenciales consideradas críticas para la presente experiencia:

- Permite dinámicas de gestión distribuida y multiusuario del vocabulario: considerando que en la experiencia debían intervenir más de un editor, este constituía un requerimiento crítico.
- Permite el desarrollo de vocabularios multilingües: al tratarse del desarrollo de un vocabulario multilingüe, esta capacidad constituía un requerimiento crítico.
- Permite exportar y exponer los vocabularios controlados a través de una variedad de esquemas de metadatos (Skos-Core, TopicMaps, Zthes, Dublin Core, entre otros): este tipo de funcionalidad se valoró en atención a la posibilidad de maximizar las condiciones de interoperabilidad y utilidad del vocabulario controlado resultante.
- Dispone de servicios web: las capacidades de la interfaz de servicios web de TemaTres permite generar subproductos independientes y articulares con respecto al vocabulario controlado como tal.

Una vez seleccionada la herramienta se elaboró un archivo fuente de datos terminológicos de acuerdo a los formatos de importación soportados por TemaTres. Resuelto este paso, se importó la versión inicial en español del Tesauro de géneros de ficción logrando una versión digital del vocabulario controlado compuesto por 483 términos con 394 relaciones terminológicas y 18 notas de alcance.

Creación de versiones en inglés, francés e italiano

Una vez seleccionado el vocabulario controlado y habiéndose migrado el mismo hacia la herramienta de gestión de vocabularios controlados TemaTres, estaban ya dadas las condiciones para la creación de las versiones correspondientes en francés, inglés e italiano. En base a la misma instalación de TemaTres, se crearon tres vocabularios, en principio independientes, uno para cada idioma. En base a los términos iniciales disponibles en español, se crearon sus correspondencias terminológicas para cada uno de los vocabularios.

Asimismo, surgió la posibilidad de establecer no sólo relaciones de traducción entre lenguas, sino también mapeos terminológicos con respecto a otros vocabularios controlados existente en el mismo dominio conceptual. En tal sentido se seleccionó el vocabulario *Guidelines on Subject*

Access to Individual Works of Fiction, Drama, etc Comic books, strips, etc. elaborado por la American Library Association.

A continuación se refieren los marcos conceptuales utilizados en cada caso.

Mapeo entre los vocabularios

La formalización de las relaciones de mapeo entre los vocabularios comprendió los siguientes niveles de armonización:

- Referencias cruzadas: este tipo de armonización se basa en la formalización de relaciones de equivalencia entre vocabularios. Se utilizaron las siguientes formas de equivalencia semántica entre conceptos:
 1. Equivalencia exacta: cuando dos conceptos tienen el mismo alcance y significado.
 2. Equivalencia parcial: cuando el alcance de un concepto en un vocabulario se corresponde completamente con el de un concepto de otro vocabulario, pero esto no ocurre a la inversa.
 3. Equivalencia inexacta: cuando el alcance de dos conceptos se solapan entre sí, sin llegar a corresponderse por completo.

No se utilizó la relación de no-equivalencia, referida al caso de no existencia en el vocabulario de destino de un concepto que contenga el alcance del concepto del vocabulario fuente, ya que en términos funcionales, ese tipo de relación resultó expresada por la nulidad o no existencia de relación de relación entre dos conceptos. De esta manera, para el desarrollo de esta experiencia se consideró que las relaciones de equivalencia entre conceptos de diferentes vocabularios, debían ser expresados siempre de manera positiva, considerando la no existencia de relación como una no-equivalencia.

Ejemplos de equivalencia exacta:

Español: Historieta – Francés: Historiette

Español: Ficción – English: Fiction

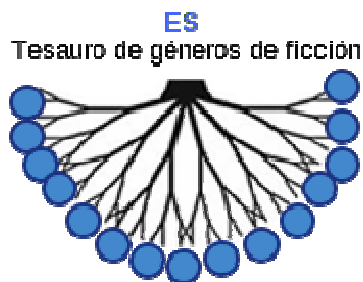
- Asociación: se trata de la articulación consistente entre un vocabulario especializado en el contexto de la estructura jerárquica de un vocabulario más general. Este tipo de armonización fue el predominante en el caso de las relaciones establecidas con vocabulario *Guidelines on Subject Access to Individual Works of Fiction, Drama, etc.* siendo este un vocabulario menos especializado en géneros de ficción literaria en particular. Por ejemplo:
 - Tesoro de géneros de ficción : Historieta = *Guidelines on Subject Access to Individual Works of Fiction, Drama, etc* : Comic books, strips, etc.

Detalle de las etapas del trabajo

Más allá de los marcos conceptuales utilizados, fue necesario establecer etapas, reglas y metodologías para la elaboración de la versión multilingüe del *Tesoro de géneros de ficción* y el mapeo con respecto al vocabulario controlado *Guidelines on Subject Access to Individual Works of Fiction, Drama, etc.* requirió de instancias sucesivas de consolidación y avance.

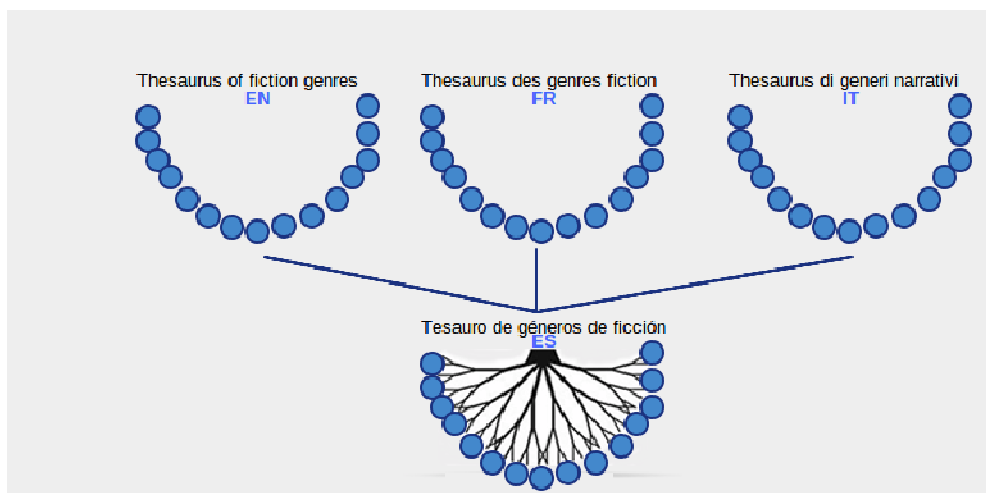
1) Incorporación del vocabulario en español a TemaTres.

En esta instancia se realizó la migración de los términos y las relaciones terminológicas establecidas en la versión impresa del *Tesouro de géneros de ficción*.



2) Creación de las versiones en inglés, francés e italiano del vocabulario y generación de set de términos correspondientes en TemaTres

En esta instancia, a partir de la versión inicial en español, se realizó la traducción de todos los términos a los idiomas previstos.

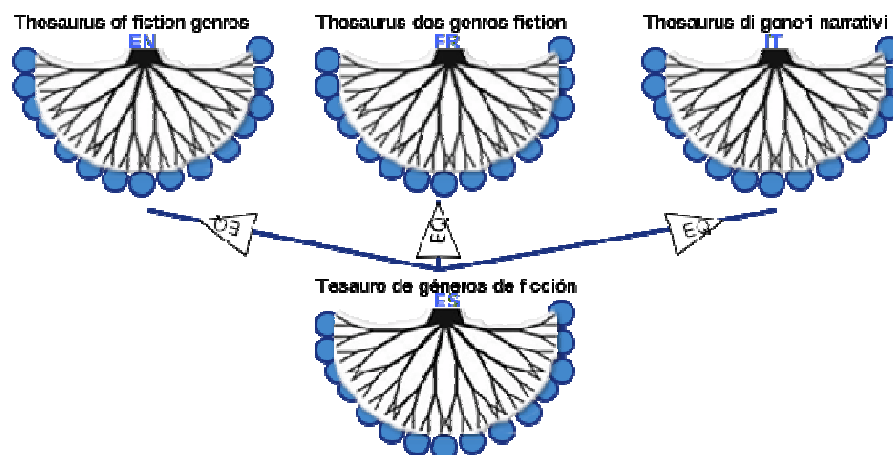


3) Generación de las relaciones terminológicas en inglés, francés e italiano.

En esta instancia, sobre la base de las relaciones terminológicas existentes en la versión en español, se estableció un algoritmo para replicar en cada una de las versiones idiomáticas del vocabulario controlados las relaciones terminológicas establecidas en la versión en español. El algoritmo de replica de relaciones terminológicas se diseñó según el siguiente esquema lógico:

Si el término A del Vocabulario 1 (A_{v1}) es igual a el término A de V2 (A_{v2}); entonces las relaciones del término A_{v2} deberían ser las mismas que las del término A_{v1}

Por ejemplo, si en la versión en español *Relato histórico* es un *término relacionado* (TR) de *Relato corto*; entonces en la versión en inglés *Historical relate* será un *término relacionado* (TR) de *Short story*.

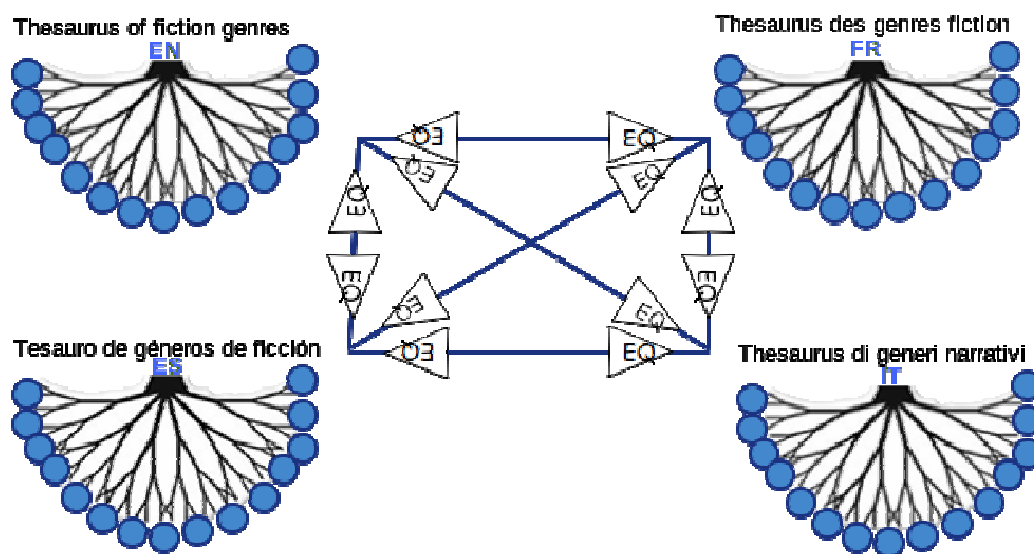


4) Homologación de relaciones de traducción entre todos los idiomas.

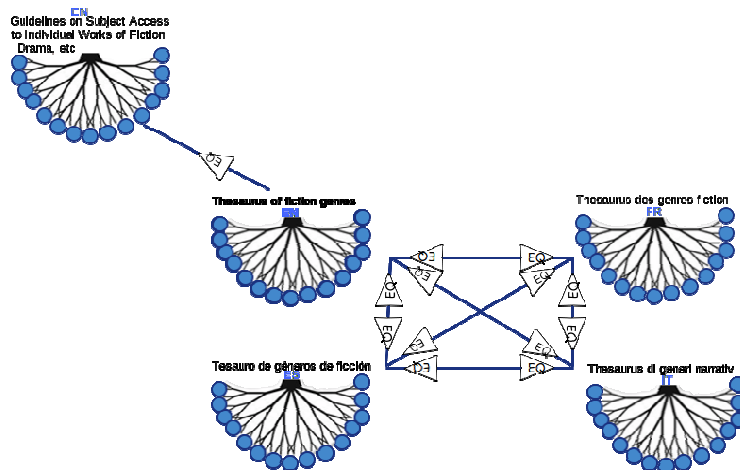
En esta instancia se establecieron las relaciones de traducción (equivalencia entre idiomas) entre todos y cada uno de los vocabularios. Hasta este momento, todos los vocabularios se referenciaban con respecto al vocabulario controlado inicial en español, conformando un esquema donde el vocabulario en español constituía el eje de las relaciones entre las distintas versiones idiomáticas, a partir de esta instancia se establecieron referencias entre todos los idiomas. El algoritmo para homologar las relaciones de traducción se diseñó según el siguiente esquema lógico:

Si el término A del Vocabulario 1 (A_{v1}) es igual a el término A de V2 (A_{v2}); y el término A del Vocabulario 1 (A_{v1}) es igual a el término A de V3 (A_{v3}); entonces el término A_{v2} es igual al término A_{v3} .

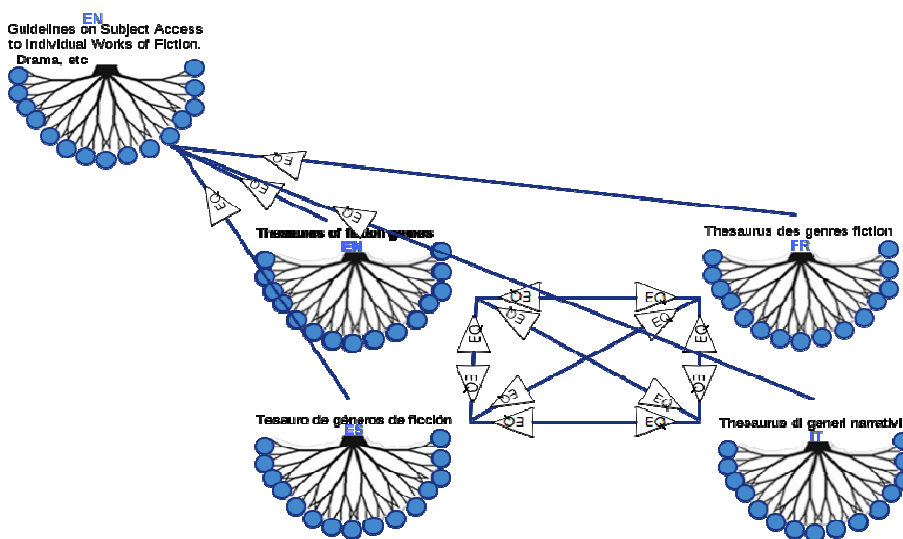
Por ejemplo, la traducción de *Relato histórico* al inglés es *Historical relate*; y la traducción de *Relato histórico* al francés es *Chronique*; entonces la traducción de *Historical relate* al francés será *Chronique*.



Con el fin de evaluar las condiciones de reuso y réplica de la metodología, esta última instancia del procedimiento fue luego repetida pero para establecer relaciones de mapeo terminológico entre el vocabulario controlado *Guidelines on Subject Access to Individual Works of Fiction, Drama, etc.* y la versión en inglés del *Tesauro de géneros de ficción*. Las relaciones de mapeo entre los vocabularios se establecieron en base al análisis conceptual de los términos.



Luego, una vez realizados los mapeos entre los dos vocabularios controlados, se volvió a aplicar el algoritmo de homologación de traducción para generalizar las relaciones establecidas entre los dos vocabularios en idioma inglés hacia las versiones en español, italiano y francés del *Tesauro de géneros de ficción*.



Conclusiones

La armonización resultó ser un proceso apropiado para la resolución de problemáticas generales de la normalización y la elaboración de herramientas multilingües de representación formal del conocimiento.

A través de metodologías replicables y aplicables a diversos escenarios y en diferentes escalas, fue posible generar un vocabulario multilingüe a partir de un vocabulario inicial en un único idioma. Asimismo fue posible generalizar relaciones de mapeo conceptual desde un vocabulario hacia otros.

Fue posible detectar ciertos problemas derivados de la aplicación del algoritmo de homologación de las relaciones de traducción en el caso de términos que sólo podían ser conceptualizados en el contexto de la lengua de origen y cuyas traducciones no pueden ser contextualizadas en el esquema de significado conformado por la estructura de relaciones terminológicas de la lengua de destino. El ejemplo más notable de este tipo de caso podría ser *Novela gaucha* o *Novela de la tierra*, términos cuya transposición a otros sistemas literarios o culturales no resulta trivial o transparente en tanto remite a construcciones sociales e históricas muy definidas y de difícil transposición.

Desde la realización de la presente experiencia la misma metodología ha podido ser replicada en escenarios de prueba y también en escenarios reales de uso y aplicación. Un ejemplo de replica para optimizar y sistematizar la metodología fue realizado con el Tesoro de la UNESCO, los resultados pueden ser consultados en <http://www.vocabularyserver.com/unesco/>. Asimismo, la misma metodología fue aplicada con éxito para la implementación de un nuevo vocabulario multilingüe en el caso del Tesoro del proyecto Central European Flood Risk Assessment and Management in CENTROPE (CEframe), los resultados pueden ser consultados en <http://www.ceframe.eu/thesaurus/>

Bibliografía

- Bowker, G. C. (2006). Memory practices in th science. Cambridge, Mass: MIT Press.
- Bowker, G. C., & Star, S. L. (1999). Sorting things out Classification and its consequences. Cambridge, Mass: MIT Press.
- Durkheim, Émile ([1901-1902], 1986) . Clasificaciones primitivas y otros ensayos de antropología positiva. Barcelona: Ariel.
- Eco, Umberto (1999). Kant y el ornitorrinco Barcelona: Lumen.
- Eco, Umberto (2005) La búsqueda de la lengua perfecta en la cultura europea. Barcelona: Editorial Crítica.
- European Committee for standardization (2005). CWA 15453:2005. Harmonisation of vocabularies for eLearning. Recuperado octubre 20, 2009, de : <http://fire.eun.org/cwa15453-00-2005-Nov.pdf>. Fecha de consulta: 10/10/2008
- Foucault, Michel ([1970], 1996). El orden del discurso Madrid: Las ediciones de la piqueta.
- Foucault, Michel ([1966] 2008). Las palabras y las cosas: una arqueología de las ciencias humanas Buenos Aires: Siglo XXI Editores, 2da ed.
- Heery, Rachel; Patel, Manjula. (2000). Application profiles: Mixing and matching metadata schemas, En: Ariadne, (25). Recuperado Agosto 26, 2007, de : <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>.
- International Federation of Library Associations and Institutions. (2009). Guidelines for multilingual thesauri. The Hague: International Federation of Library Associations and Institutions. Recuperado octubre 20, 2009, de : <http://www.ifap.ru/library/book411.pdf>
- Lagoze, C., Lynch, C. A., and Daniel, R. (1996). The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata. Cornell Computer Science Technical Report TR96-1593.

- Lyotard, Jean François ([1979],1993). La condición postmoderna: informe sobre el saber. Barcelona: Planeta-Agostini.
- Manovich, Lev (2003) The language of new media. Boston: MIT Press.
- National Information Standards Organization (U.S.). (2005). Guidelines for the construction, format, and management of monolingual controlled vocabulary. National information standards series. Bethesda, Md: NISO Press. Recuperado octubre 20, 2009, de : http://www.niso.org/kst/reports/standards/kfile_download?&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVa86hjP37r_hONsJghRDv2N-zj4TZCh8Dp01rZbmK3O-8vcVjh4hezP
- Pérez Agüera, José Ramón *Automatización de tesauros y su utilización en la web semántica*. En: BiD : textos universitarios de biblioteconomía i documentació (13), 2004. Disponible en <http://eprints.rclis.org/archive/00004176/>. Fecha de consulta: 10/10/2008
- Rofes, Octavi (2007) *Sobre la traducción: teorías, prácticas y actitudes* En: Agra, R, comp. Zapra, G, comp. (2007) *Muntadas Con/Textos II*. Buenos Aires: Nueva librería.
- Safari, M. (2004). *Metadata and the Web*. En: Webology, 1(2). Recuperado Agosto 26, 2007, de : <http://www.webology.ir/2004/v1n2/a7.html>.
- Veron, E. (2004). Fragmentos de un tejido. Barcelona: Gedisa.
- Weinberger, David (2007) Everything Is Miscellaneous: The Power of the New Digital Disorder. New York : Times Books.
- Wittgenstein, Ludwing ([1914-1916] 1986) Diario filosófico (1914-1916) Barcelona: Planeta-Agostini