



TBX en SDL MultiTerm

Detlef Reineke
Universidad de Las Palmas de Gran Canaria
Comité Técnico 37 de la Organización Internacional
de Normalización (ISO TC 37)
detlef_reineke@yahoo.es

RESUMEN

El intercambio de datos terminológicos entre aplicaciones de gestión terminológica se realiza mediante una serie de interfaces como, por ejemplo, TermBase eXchange (TBX = 30042). El presente artículo analiza el grado de conformidad de la recién lanzada versión SDL MultiTerm 2014 para con el estándar TBX, y propone mecanismos de conversión alternativos en casos de no conformidad.

Palabras clave: TBX; ISO 30042; intercambio de datos terminológicos; SDL MultiTerm

RESUM (TBX en SDL MultiTerm)

L'intercanvi de dades terminològiques entre aplicacions de gestió terminològica es porta a terme mitjançant una sèrie de interfícies, com ara TermBase eXchange (TBX = 30042). El present article analitza el grau de conformitat de la recent versió SDL MultiTerm 2014 amb l'estàndard TBX, i proposa mecanismes de conversió alternatius en casos de no conformitat.

Paraules clau: TBX; ISO 30042; intercanvi de dades terminològiques; SDL MultiTerm

ABSTRACT (TBX in SDL MultiTerm)

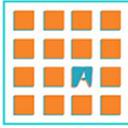
Exchanging terminological data between tools can be achieved by using exchange formats like TermBase eXchange (TBX = ISO 30042). The present article examines if the recently launched version SDL MultiTerm 2014 is in compliance with TBX and proposes alternative conversion routines in case of non-compliance.

Keywords: TBX; ISO 30042; terminological data exchange; SDL MultiTerm

1. Introducción

La terminología es un elemento fundamental, no sólo en uno de sus campos pioneros, la traducción, sino también en muchos otros ámbitos, directa o indirectamente relacionados con la producción mono y multilingüe de contenidos digitales, como son la documentación técnica en su acepción tradicional (manuales, contratos, etc.), la Web semántica, la gestión de datos de productos, el abastecimiento electrónico, etc.

En lo referente a la traducción, todas las versiones de MultiTerm han desempeñado un papel predominante entre las herramientas de gestión terminológica desde su aparición en 1990. Encuestas llevadas a cabo entre traductores, terminólogos y redactores técnicos confirman que entre el 60% y el 80% de los usuarios trabaja con MultiTerm (Reineke/Sánchez-León, 2005 y Senge, 2010). Como consecuencia, algunos fabricantes han incluido interfaces que permiten intercambiar datos terminológicos directamente a través del respectivo formato de MultiTerm sin tener que pasar por un formato intermedio. No obstante, esos escenarios representan sólo una parte en el conjunto del mundo del intercambio de



datos terminológicos (Schmitz/Straub, 2010). Por todo ello, la comunidad de la terminología lleva trabajando, más de tres décadas, en el desarrollo y la mejora de estándares de intercambio de bases de datos terminológicos universales, pasando por MATER, Micro-MATER, MARTIF y, finalmente, TBX (Reineke, 2010).

2. TBX

TBX está basado en XML y constituye un derivado del estándar TMF (= Terminological Markup Framework = ISO 16642). El corazón de TMF es un metamodelo (figura 1) que define la estructura genérica de una base de datos terminológica y aporta un marco genérico para la creación de lenguajes de marcas terminológicos expresados en XML, todo ello, independientemente de aplicaciones concretas y destinado a facilitar el intercambio de datos entre aplicaciones sin pérdida de información.

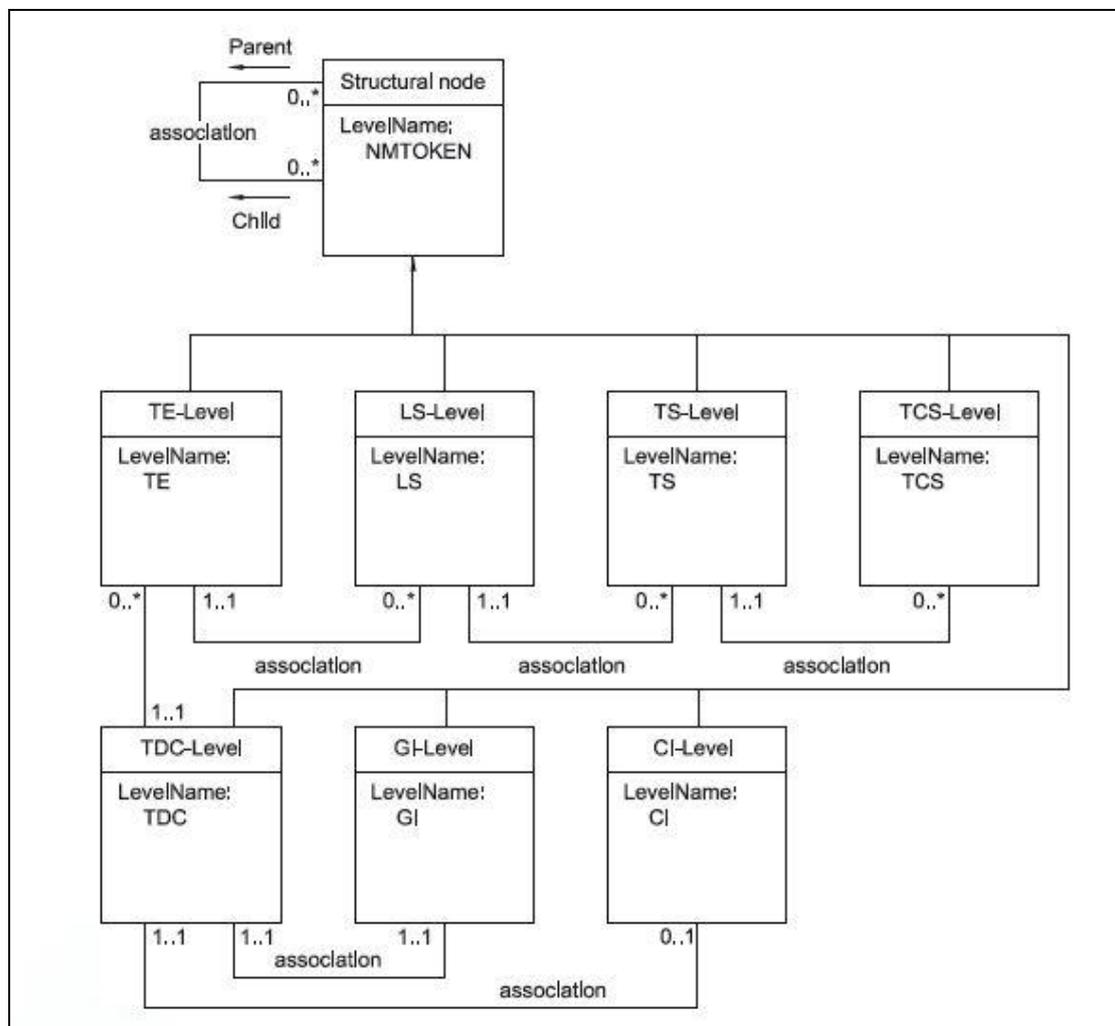
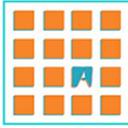


Figura 1. Metamodelo TMF diagrama en UML (ISO 16642:2003: 11).



Como ilustra el diagrama UML, el metamodelo consta de siete instancias (o contenedores). Según este modelo, cada base de datos terminológica (TDC = terminological data collection) debe contener, exactamente, un contenedor (1..1) para almacenar información global (GI = global information). Esa información concierne al conjunto de los elementos de una base de datos, como pueden ser el nombre de la base de datos, los datos de su creadores, el historial de modificaciones, los derechos de autor, etc. Los recursos compartidos (bibliografía completa, archivos multimedia, etc.), a los que se hace referencia desde las entradas terminológicas, deben depositarse en el contenedor CI (= complementary information) que puede obviarse o aparecer una vez (0..1). Cada base de datos puede comprender cualquier número (0..*) de entradas terminológicas (TE = terminological entry) que, a su vez, pueden encapsular un número indeterminado de contenedores de idiomas (LS = language section). A dichas entradas pueden asignarseles 0..* contenedores para la descripción de los términos (TS = term section). En caso de que se precise documentar los componentes pluriverbales de un término, pueden instanciarse los contenedores correspondientes (TCS = term component section, ocurrencia 0..*).

En TBX, el metamodelo resulta en la siguiente estructura:

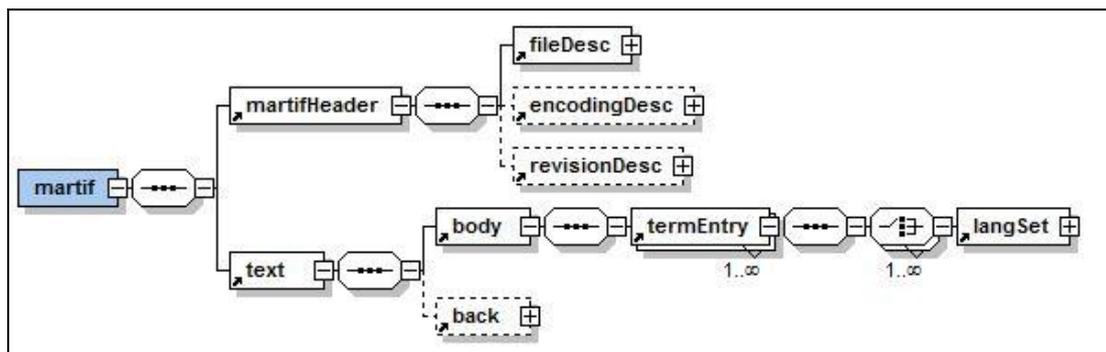


Figura 2. Elementos superiores de la estructura TBX (ISO 30042:2008: 10).

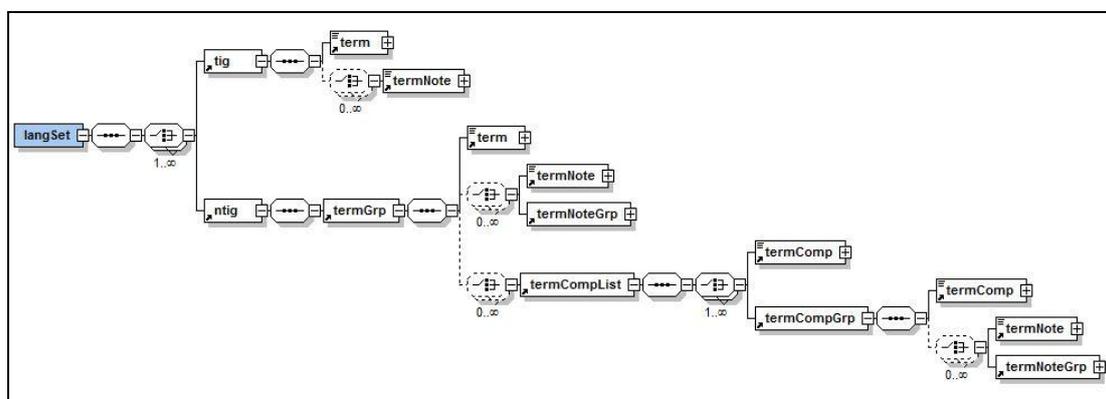
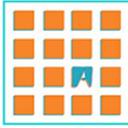


Figura 3. Elementos inferiores de la estructura inferior de TBX (ISO 30042:2008: 10).



Los nodos principales de la estructura TBX coinciden con las siete instancias del metamodelo (TDC = martif, GI = martifHeader, CI = back, LS = langSet, TE = termEntry, TS = tig/ntig, TCS = termCompList), pero distan en cuanto a las ocurrencias. En TBX, los elementos “termEntry”, “langSet” y “ntig”/“tig”^[1] deben ser instanciados, por lo menos, una vez (1..∞). Si se instancia el elemento “termCompList”, también debe haber, por lo menos, un elemento “termComp” o “termCompGrp”.

Todos esos nodos encapsulan las categorías de datos TBX que están definidas en ISOcat (www.isocat.org), el registro de categorías de datos del ISO TC 37. Por regla general, las categorías de datos no se instancian como marcas en el árbol XML, sino que van asignadas a metacategorías de datos (“descrip”, “admin”, “xref”, etc.) mediante un atributo “type” (por ejemplo, `<admin type="source">http://www.grafico.com</admin>`). Ese formalismo garantiza, por un lado, suficiente flexibilidad para dar cabida a un máximo número de modelos de datos y, por otro lado, conservar un cierto grado de rigidez para lograr un alto nivel de automatización en el intercambio de datos. Un número reducido de categorías de datos es instanciado directamente como atributo (por ejemplo, “xml:lang” en `<langSet xml:lang="es">`) o como elemento (“term”, “date”, “note”, etc.).

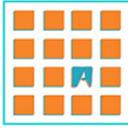
Junto con el estándar TBX se facilitan una serie de recursos estándares que permiten validar una instancia de documento TBX, en cuanto a su estructura y sus categorías de datos. La estructura de dichas instancias puede ser validada contra una DTD (TBXcoreStructV02.dtd) y una especificación de categorías de datos (TBXXCSV02.xcs y tbxxcsdtd.dtd) o contra un único modelo de contenido, un RNG Schema (TBX_RNGV02.rng). No obstante, la validación de instancias de documento TBX presenta ciertas limitaciones. Los editores XML convencionales no son capaces de procesar la DTD y la especificación de categorías en archivos separados. Por otra parte, sólo unos pocos editores XML procesan el lenguaje RNG Schema (por ejemplo, `</oXygen>`® XML Editor de Synchro Soft). Para salir del impasse, el grupo de investigación “Translation Research Group” de la Brigham Young University desarrolló el TBX Checker, una aplicación libre que valida instancias contra ambos modelos de contenido.

Cabe recordar, también, que la norma ISO 30042 es un macro estándar que cuenta con más de 200 categorías de datos. Está concebido para que, en la práctica, se instancien subconjuntos de TBX (dialectos), ya que apenas se necesitan toda la estructura y todas las categorías de datos TBX para modelar bases de datos terminológicas concretas. Esos dialectos, evidentemente, no están normalizados, pero si validan correctamente contra los recursos estándares, son considerados conforme a TBX.

3. Muestra de TBX

Para el análisis del grado de conformidad de MultiTerm con el estándar TBX, se ha creado una pequeña muestra de una instancia de documento TBX (dos conceptos – opacidad y transparencia – y dos idiomas) que incluye aspectos característicos en cuanto a la estructura y las categorías de datos. El ejemplo ilustrado en la figura 4 refleja un escenario, en el que la base de datos ha sido diseñada para su uso en un entorno colaborativo de una gran empresa de habla castellana. Los usuarios podrían ser desarrolladores de productos, autores de contenidos, traductores, distribuidores y publicistas. El idioma de trabajo es el castellano y la definición se ha asignado al nivel del concepto lo que garantiza una mayor

¹ El elemento “ntig” (= nested term information group) permite una estructura XML con mayor profundidad con el fin de poder describir los componentes de términos pluriverbales. En caso de que no sea necesario describir esos componentes, se suele utilizar el elemento “tig” (= term information group) que solamente cuenta con una profundidad de nodo igual a 1.

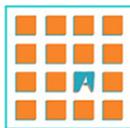


consistencia del mismo. La asignación de la definición o de las definiciones, al nivel del idioma, hubiera sido igualmente válida (y frecuente, sobre todo en entornos donde las labores son esencialmente de traducción). Pero, modelar los datos de esa manera tiene el inconveniente de una posible dispersión del concepto debido a la coexistencia de varias definiciones que, en pocos casos, coinciden de idioma a idioma.

A título de ejemplo, y en aras a ser breve, el siguiente código representa, solamente, aquella parte de la entrada terminológica relativa al concepto opacidad, así como la información que concierne al término en castellano. La sintaxis y el vocabulario XML son idénticos, tanto para el término inglés como para el concepto transparencia, variando , solamente, los contenidos de los elementos y el valor del atributo "xml:lang".

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE martif SYSTEM ".\TBXcoreStructV02.dtd">
<martif type="TBX" xml:lang="es">
  <martifHeader>
    <fileDesc>
      <sourceDesc>
        <p>TBX - Ejemplo Tradumática</p>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <p type="XCSUR!">.\TBXXCSV02.XCS</p>
    </encodingDesc>
  </martifHeader>
  <text>
    <body>
      <termEntry id="tid_1">
        <descrip type="subjectField">propiedades de los materiales</descrip>
        <descripGrp>
          <descrip type="definition">capacidad de un material de impedir el paso de la luz</descrip>
          <admin type="source">http://www.glosariografico.com</admin>
        </descripGrp>
        <xref type="xGraphic" target="opacidad.png"/>
        <xref type="xSource" target="http://docs.autodesk.com"/>
        <langSet xml:lang="es">
          <tig>
            <term>opacidad</term>
            <termNote type="partOfSpeech">sustantivo</termNote>
            <termNote type="grammaticalGender">f.</termNote>
            <termNote type="grammaticalNumber">sg.</termNote>
            <admin type="source">http://www.glosariografico.com</admin>
          </tig>
        </langSet>
        <langSet xml:lang="en">...</langSet>
      </termEntry>
      <termEntry id="tid_2">...</termEntry>
    </body>
  </text>
</martif>
```

Figura 4. Muestra de instancia de documento TBX.



4. MultiTerm

Con la versión iX se incorporaron cambios sustanciales a MultiTerm, entre otras cosas, el soporte XML, una estructura de entrada terminológica basada en TBX, así como una interfaz de exportación a TBX basada en la especificación de la LISA del año 2002. Dicha especificación sigue implementada en la versión 2014 de MultiTerm, a pesar de las modificaciones que ha sufrido TBX desde entonces y que han culminado en la publicación de la norma ISO 30042 en el año 2008. En cuanto a la conformidad de MultiTerm con el estándar TBX, el análisis podría concluir aquí, puesto que una comparación de las citadas versiones de TBX no concerniría, *sensu stricto*, a la aplicación MultiTerm. No obstante, las versiones TBX no distan mucho entre sí, por lo que cabe proceder a determinar hasta qué punto MultiTerm cumple con dichas especificaciones.

Como en muchas otras aplicaciones informáticas que operan con bases de datos, el procesamiento de datos está condicionado por la conjugación de dos elementos fundamentales:

- El sistema front-end, es decir, la interfaz a través de la que el usuario interactúa con la aplicación para introducir o acceder a datos – en caso de MultiTerm, por ejemplo, definir los modelos de datos, los modelos de entrada o los diseños, configurar las interfaces de exportación/importación, etc.
- El sistema back-end, en el que se procesan y se almacenan los datos procedentes del front-end.

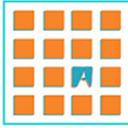
En el sistema front-end de MultiTerm, todo usuario, utilice plantillas de MultiTerm o estructuras de entradas terminológicas propias, debe definir la estructura de su entrada terminológica en torno a una estructura TBX propia de MultiTerm. La estructura, las categorías de datos y los propios datos son almacenados en una base de datos relacional en el sistema back-end. Desde aquí se exportan y se importan los datos mediante los pertinentes mecanismos de conversión. Cabe señalar que no sería preceptivo implementar una estructura de entrada terminológica en el sistema front-end con el fin de lograr un intercambio de datos conforme a las especificaciones del estándar TBX. Cualquier base de datos relacional del back-end, como una simple hoja de cálculo, podría ser mapeada a TBX (véase Reineke 2013a), eso sí, mediante formalismos de conversión algo más complejos. En este sentido, la imposición de una estructura de entrada terminológica en el sistema front-end no sólo simplifica el preprocesamiento de los datos terminológicos para su posterior conversión a TBX, sino que, también, facilita la adecuada interpretación de los datos terminológicos de las entradas terminológicas en el front-end por parte de los usuarios.

Por defecto, la entrada terminológica de MultiTerm consta de tres niveles: “Entry level”, “Index level” y “Term level”, tanto en las plantillas proporcionadas por MultiTerm (el glosario bilingüe, la base de datos terminológica multilingüe, etc.) como en los modelos de datos creados por el usuario. En todos los casos, cada una de las entradas terminológicas se transforma a XML y se almacena en una celda de una base de datos relacional (los comentarios en verde fueron añadidos por parte del autor):



```
<cG id="0"> <!-- Corresponde a "Entry level" en el front-end. -->
  <c>1 </c>
  <dG id="24">
    <d type="Campo de especialidad">propiedades de los materiales</d>
  </dG>
  <dG id="25">
    <d type="Definición">capacidad de un material de impedir el paso de la luz</d>
  </dG>
  <dG id="31">
    <d type="Fuente">
      <xref Ulink="http://www.glosariografico.com">http://www.glosariografico.com</xref>
    </d>
  </dG>
  <dG id="26" multimedia="1">
    <d type="Imagen">opacidad.png</d>
  </dG>
  <dG id="32">
    <d type="Fuente">
      <xref Ulink="http://docs.autodesk.com">http://docs.autodesk.com</xref>
    </d>
  </dG>
  <IG id="27"> <!-- Corresponde a "Index level" en el front-end. -->
    <l lang="ES" type="Español"/>
    <IG id="33"> <!-- Corresponde a "Term level" en el front-end. -->
      <t>opacidad</t>
      <dG id="34">
        <d type="Clase gramatical">sustantivo</d>
      </dG>
      <dG id="35">...</dG>
      <dG id="37">
        <d type="Fuente">
          <xref Ulink="http://www.glosariografico.com">http://www.glosariografico.com</xref>
        </d>
      </dG>
      <trG id="38">... </trG>
    </IG>
  </IG>
  <trG id="29">... </trG>
</cG>
<cG id="0">... </cG>
```

Figura 5. Estructura de entrada terminológica en el sistema back-end de MultiTerm.



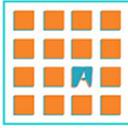
4.1. Exportar a TBX

Como demuestra el ejemplo, el sistema back-end conserva la estructura de la entrada terminológica. A la hora de exportar las entradas terminológicas a través de las distintas interfaces de exportación de MultiTerm, la información terminológica depositada en las entradas terminológicas es convertida a los correspondientes formatos de exportación. En el caso de la interfaz de exportación TBX, el resultado para el ejemplo ilustrado en la figura 5 es el siguiente:

```
<?xml version="1.0" encoding="utf-16"?>
<martif (declaraciones de espacios de nombres, referencia al modelo de contenido, indicación del tipo de documento y del idioma de trabajo del documento)>
  <martifHeader>...</martifHeader>
  <text>
    <body>
      <termEntry> <!-- Corresponde a "cG" en el back-end. -->
        <transacGrp>...</transacGrp>
        <transacGrp>...</transacGrp>
        <descripGrp>
          <descrip type="Campo de especialidad">propiedades de los materiales dato</descrip>
        </descripGrp>
        <descripGrp>
          <descrip type="Definición">capacidad de un material de impedir el paso de la luz </descrip>
        </descripGrp>
        <descripGrp>
          <descrip type="Imagen">opacidad.png</descrip>
        </descripGrp>
        <langSet xml:lang="ES"> <!-- Corresponde a "IG" en el back-end. -->
          <tig> <!-- Corresponde a "tG" en el back-end. -->
            <term>opacidad</term>
            <transacGrp>...</transacGrp>
            <descripGrp>
              <descrip type="Clase gramatical">sustantivo</descrip>
            </descripGrp>
            <descripGrp>...</descripGrp>
            <descripGrp>
              <descrip type="Fuente">http://www.glosariografico.com</descrip>
            </descripGrp>
            </tig>
          </langSet>
        </termEntry>
        <termEntry>...</termEntry>
      </body>
    </text>
  </martif>
```

Figura 6. Formato MultiTerm de TBX.

Como se puede apreciar a primera vista, en el proceso de la conversión a TBX se han eliminado, y sin previo aviso por parte de la aplicación, la fuente de la definición y de la imagen, respectivamente. Dichas categorías de datos se han definido en la estructura de la entrada terminológica como categorías de datos hijos de categorías asignadas al nivel "Entry level". Ese hecho es extrapolable a cualquier escenario en MultiTerm donde se pretenda definir una profundidad de nodo mayor que 1, tomando como referencia los niveles principales. Por consiguiente, el nivel "Term Component Section" del metamodelo TMF y, por



ende, los correspondientes nodos en TBX (elemento “termCompList” y sucesivos) tampoco son instanciables en la conversión a TBX.

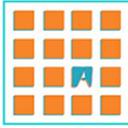
Por otra parte, en MultiTerm se asignan categorías de datos relativas a los términos, como “Clase gramatical”, “Género” o “Número” a la metacategoría “descrip”. Sin embargo, esa asignación contradice la especificación TBX de la LISA del año 2002 y la de la ISO 30042. En ambas, se determina que la metacategoría de datos “descrip” está reservada para la instanciación de categoría de datos que describen el concepto (LISA, 2002: 12). Lo correcto, en esos casos, sería utilizar la metacategoría “termNote” (ISO, 2008: 18). Aún así, cabe señalar que los modelos de contenidos de ambas especificaciones permiten instanciar la metacategoría de datos “descrip” al nivel del término, como se aprecia en las figuras 7 y 8:

```
<xs:group name="auxInfo">
  <xs:choice>
    <xs:element ref="descrip"/>
    <xs:element ref="descripGrp"/>
    <xs:element ref="admin"/>
    <xs:element ref="adminGrp"/>
    <xs:element ref="transacGrp"/>
    <xs:element ref="note"/>
    <xs:element ref="ref"/>
    <xs:element ref="xref"/>
  </xs:choice>
</xs:group>
...
<xs:element name="tig">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="term"/>
      <xs:choice minOccurs="0" maxOccurs="unbounded">
        <xs:element ref="termNote"/>
      </xs:choice>
      <xs:group ref="auxInfo" minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:attribute name="id" type="xs:ID" use="optional"/>
  </xs:complexType>
</xs:element>
```

Figura 7. El elemento “tig” en el XML Schema correspondiente a la especificación del año 2002.

```
<!ENTITY % auxInfo '(descrip | descripGrp | admin | adminGrp | transacGrp | note | ref | xref)*'>
...
<!ELEMENT tig (term, (termNote)*, %auxInfo;)>
<!ATTLIST tig
  id ID #IMPLIED
```

Figura 8. El elemento “tig” en la DTD correspondiente a la estructura TBX (versión 2002 y 2008).



Esa vulneración del principio del metamodelo – la inclusión de información relativa al concepto al nivel del término – constituye un compromiso cara al intercambio de datos entre aplicaciones terminológicas y lexicográficas. De esa forma, las definiciones asignadas a lexemas podrán ser mapeadas con mayor facilidad a TBX, y viceversa.. Ahora bien, MultiTerm asigna de forma indiscriminada todas las categorías de datos a la metacategoría “descrip”, incumpliendo, así, ambas especificaciones TBX.

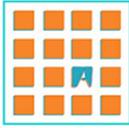
Otro defecto de MultiTerm consiste en que, de forma generalizada, se asigna el valor “source” (o “Fuente”, etc.) del atributo “type” a la metacategoría de datos “descrip”, mientras que TBX exige la asignación a la metacategoría “admin” (ISO, 2008: 20). Ese hecho sería justificable en estructuras de entradas terminológicas con categorías de datos definidas por el usuario, puesto que se precisarían componentes semánticos que interpreten el significado de esas categorías. Sin embargo, ni siquiera en las plantillas de MultiTerm que incluyen la categoría de datos “Source” se asigna esa categoría a la metacategoría “admin”.

Un aspecto que llama la atención es el hecho de que algunas de las plantillas de modelos de datos de MultiTerm definan la categoría de datos “Subject” como “Lista de selección”, mientras que en TBX el tipo de dato de esa categoría de datos es “plainText”, y no “picklist” (ISO 30042:2008: 19). No obstante, esa divergencia no afecta a la correcta validación de las instancias de documento TBX de MultiTerm porque: a.) la información del tipo de dato no es instanciada en las instancias de documento XML y b.) el pársers que valida las instancias de documento contra los modelos de contenido TBX no es forzado a interpretar valores asociados a esa categoría de datos.

Otro inconveniente de MultiTerm radica en que las categorías de datos y, donde procede, los valores de categorías de datos, no son mapeables en el proceso de exportación a los correspondientes identificadores de las categorías de datos definidos en TBX (por ejemplo, “Campo de especialidad” a “subjectField”, o “f.” a “feminine”). Esa circunstancia hace necesario llevar a cabo un mapeado adicional fuera de MultiTerm, una vez que se hayan exportado las entradas terminológicas. Un paliativo consistiría en utilizar los identificadores y los valores de TBX en el sistema front-end de MultiTerm. No obstante, esa variante puede incidir en la manejabilidad de la base de datos debido a la menor legibilidad de la escritura “camelCase”, y en aquellos casos donde el inglés no sea el idioma de trabajo.

Todas las instancias de documento TBX de MultiTerm hacen referencia a un modelo de contenido (TBXcsV02.xsd) que se ha derivado de su modelo homólogo expresado en SGML (TBXcdv04.dtd). Ambos modelos, que se encuentran en la carpeta de instalación de Multiterm, definen las reglas relativas a la estructura TBX con las que deben cumplir las instancias de documento TBX. Pero, en esos modelos no se especifican ni se referencian las categorías de datos y los valores permitidos, tal como exige el estándar TBX. Además, la mencionada DTD forma parte de la norma ISO 12200 que define el estándar MARTIF. Dicha norma está a punto de ser retirada por considerarse que la norma ISO 30042 ya le ha tomado el relevo. Es cierto que el modelo de contenido de la ISO 30042, relativa a la estructura TBX (TBXcoreStructV02.dtd), no dista mucho de su antecesor. Los cambios sustanciales consisten en:

- la retirada del elemento “front”,
- la modificación de los valores de atributo relativos a la indicación de la especificación de las categorías de datos,
- la sustitución del valor “terminologyManagementTransactions” por el valor “transacType”.



Por ello, la validación de la estructura de instancias de documentos TBX, sean resultados de la exportación de MultiTerm o procedentes de otras aplicaciones o escenarios, da positivo en la gran mayoría de los casos con sendos modelos. Pero, si se lleva a cabo una validación de las instancias de documento TBX de MultiTerm contra los recursos estándares vigentes señalados en el apartado 2, el resultado es negativo en todos los casos por los motivos anteriormente señalados.

4.2. Alternativas de exportación a TBX

Existen varias alternativas para generar instancias de documento conforme a las especificaciones de TBX, unas más complejas que otras. Por una parte, sería posible exportar hacia el formato XML de MultiTerm (“Default export”) y mapear el archivo a TBX en una aplicación de mapeado externa como, por ejemplo, Altova® MapForce®. Por defecto, las instancias de documento del formato MultiTerm XML no incluyen ninguna referencia a un modelo de contenido, modelo que es necesario para el mapeado. No obstante, la mencionada aplicación de mapeado puede generar un modelo de contenido a partir de la instancia de documento, al instante. La ventaja de pasar por MultiTerm XML reside, por ejemplo, en que profundidades de nodo igual a 2, en los nodos principales, son exportables. Ello permitiría conservar la información relativa a los componentes pluriverbales de un término (termCompList).

Una segunda opción consiste en exportar las entradas terminológicas al formato de texto, separado por tabuladores. Este escenario es un tanto más laborioso, ya que habría que importar el archivo resultante de la exportación, por ejemplo, en Microsoft® Excel®, crear un XML Schema para la exportación desde Microsoft® Excel® a XML y, posteriormente, mapear el archivo XML a TBX. Las demás interfaces de exportación de MultiTerm, como “HTML export”, “Word dictionary export” o “List”, podrían ser igualmente mapeadas a TBX, pero presentan un inconveniente adicional, dado que permiten exportar solamente una parte, y no la totalidad de la entrada terminológica.

Otra alternativa, a juicio del autor la más elegante, radica en definir una interfaz de exportación personalizada. En un primer paso, es preciso definir qué parte de información de la entrada terminológica se quiere exportar, para luego ir especificando el vocabulario y la sintaxis XML en torno a la estructura definida (véase figura 9). El resultado es una instancia de documento TBX que valida correctamente contra los modelos de contenidos estándares y la especificación de categorías de datos propuestos por el ISO TC 37.

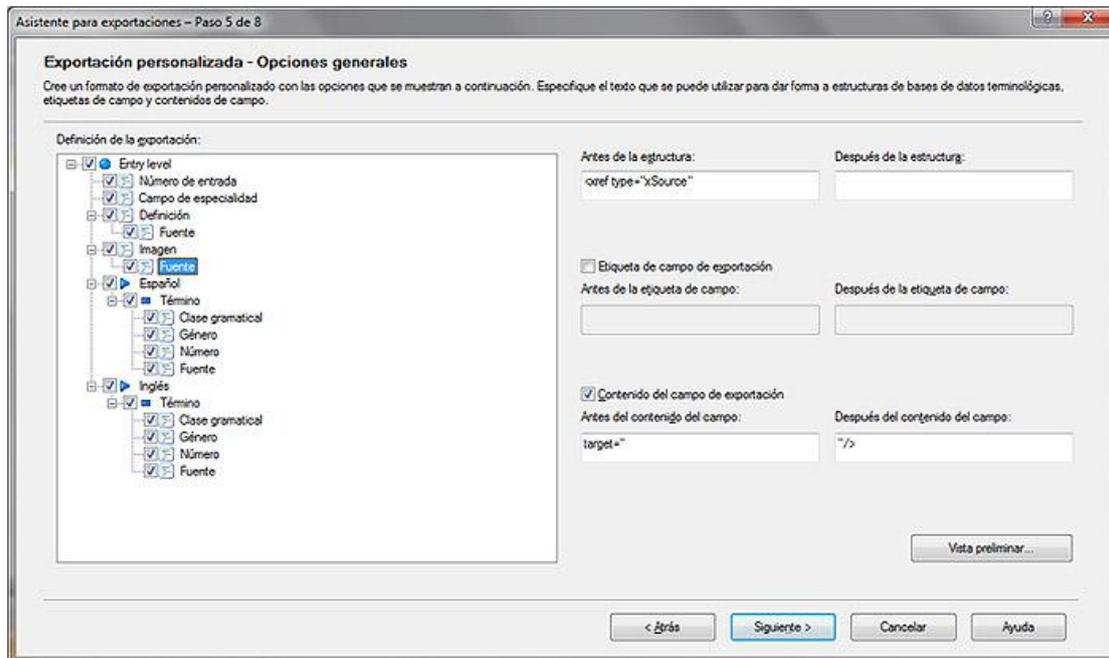
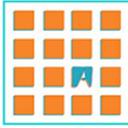


Figura 9. Exportación personalizada en MultiTerm.

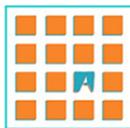
4.3 Importar desde TBX

MultiTerm es un paquete de aplicaciones, entre las que se encuentra MultiTerm Convert. Dicha aplicación permite convertir determinados formatos (hojas de cálculo, OLIF XML, MultiTerm 5, TBX, etc.) al formato MultiTerm XML, que es el único formato de importación procesable por MultiTerm. En lo referente a TBX, una conversión sin más de instancias de documento conforme a TBX conduce a la exclusión de aquellos datos que no coinciden con el formato MultiTerm XML (por ejemplo, la categoría de datos “source” del ejemplo ilustrado en la figura 4, por ser asignada a la metacategoría de datos “admin”). Con el fin de evitar la pérdida de datos es recomendable, por lo tanto, hacer un mapeado al formato TBX de MultiTerm antes de proceder a la transformación en MultiTerm Convert.

El mapeado en MultiTerm Convert ofrece una serie de opciones de control y de manipulación de datos. Por ejemplo, es posible validar las instancias de documento contra un XML Schema, definido por el usuario. Además, el usuario puede mapear los datos con la ayuda de un asistente para adaptar categorías de datos, valores, tipos de datos, y la asignación a los niveles del modelo de datos de la base de datos a la que se pretende importar la instancia de documento.

5. Conclusiones

Como se ha podido demostrar, TBX no está implementado, de forma completa, en MultiTerm y obliga al usuario a definir alternativas personalizadas con el fin de lograr procesos que desemboquen en resultados validables contra los recursos estándares. El asistente de mapeado en MultiTerm Convert es, sin duda, de gran ayuda, pero se echa en falta otro asistente de mapeado para el proceso de exportación. En cierto modo, la alternativa de exportación personalizada presentada en el apartado 4.2. constituye una especie de mapeado en código, aunque requiere sólidos conocimientos en TBX.

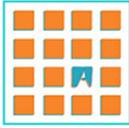


A pesar de las deficiencias señaladas, MultiTerm sigue siendo una de las aplicaciones con mayor grado de conformidad. En cuanto a otras herramientas de gestión terminológica de relevancia, como TermStar de STAR, crossTerm de Across Systems, TermWeb de Interverbum Technologies, qTerm de Kilgray o el componente de gestión terminológica de Déjà Vu, se aprecian igualmente defectos mayores o menores en la implementación de TBX (Reineke, 2013b). Pero ninguna de las herramientas mencionadas permite la configuración personalizada de las interfaces de exportación/importación de TBX.

Todo ello evidencia que aún queda un gran trabajo de concienciación por delante para que TBX llegue a tener la aceptación que experimenta, por ejemplo, el estándar TMX. Actualmente, el comité ISO TC 37 está revisando la norma ISO 30042 que concierne a cuestiones de forma y de fondo (la desaparición de la LISA, co-propietaria del estándar, la sustancial modificación de la norma ISO 12620, la creación de dialectos normalizados como TBX-Basic, TBX-Glossary, etc.).

6. Bibliografía

- BYU Translation Research Group. Translation, Theory & Technology. <<http://www.ttt.org>>. Última actualización 2011. Última consulta: 31 de agosto de 2013.
- GALA (Globalization & Localization Association). (2013). Term Base eXchange (TBX) (identical to ISO 30042:2008). Actualización 2013. <<http://www.gala-global.org/lisa-oscar-standards>>. Última consulta: 31 de agosto de 2013.
- ISO 12200. (1999). Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange. Ginebra: ISO.
- ISO 16642. (2003). Computer applications in terminology -- Terminological markup framework. Ginebra: ISO.
- ISO 30042. (2008). Systems to manage terminology, knowledge and content -- TermBase eXchange (TBX). Ginebra: ISO.
- ISO TC 37. (2013). Terminology and Other Language and Content Resources. ISOCat - Data Category Registry. Actualización 2013. <<http://www.isocat.org>>. Última consulta: 31 de agosto de 2013.
- LISA (Localization Industry Standards Association). (2002). TBX specification. Ginebra: LISA.
- Reineke, Detlef; Sánchez-León, Elizabeth (2005). "Perfil laboral y formación de – una encuesta a proveedores de servicios". En: Detlef Reineke (ed.), Traducción y localización. Mercado, gestión y tecnología. Las Palmas de Gran Canaria: Anroart.
- Reineke, Detlef. (2010). "Terminologieaustausch mit TBX". En: Felix Mayer, Detlef Reineke, Klaus-Dirk Schmitz (eds.). Best Practices in der Terminologearbeit. Munich, Colonia: SDK, p. 221-229.
- Reineke, Detlef. (2013a). "Von und nach TBX: Austausch terminologischer Daten (relativ einfach gemacht", Lebende Sprachen, 4/2013.
- Reineke, Detlef. (2013b). "Analyse der Möglichkeiten des Austausches terminologischer Daten gängiger Redaktions-, Übersetzungs- und Terminologieverwaltungssysteme", Informe del proyecto de investigación realizado en el Institut für Informationsmanagement (IIM) de la Universidad de Ciencias Aplicadas de Colonia. Bonn: Servicio Alemán de Intercambio Académico Alemán (DAAD).
- RELAX NG home page. <<http://relaxng.org>>. Última consulta: 31 de agosto de 2013.



Schmitz, Klaus-Dirk; Straub, Daniela. (2010). Erfolgreiches Terminologiemanagement im Unternehmen: Praxishilfe und Leitfaden: Grundlagen, Umsetzung, Kosten-Nutzen-Analyse, Systemübersicht. Gesellschaft für Technische Kommunikation e.V., tekomp. Stuttgart.

Senge, Sonja. (2010). Terminologieverwaltung ohne Terminologieverwaltungssystem – Alternative Möglichkeiten der Terminologieverwaltung. Master Thesis, Department für Wissens- und Kommunikationsmanagement, Donau-Universität Krems.