

Localización y terminología: efectos sinérgicos obtenidos a raíz de un experimento realizado en el aula

Detlef Reineke

**Facultad de Traducción e Interpretación
Universidad de Las Palmas de Gran Canaria**

Resumen

El avance de los contenidos digitales en la sociedad actual y la desafiante velocidad de los ciclos de innovación tecnológica exigen respuestas adecuadas por parte de la comunidad educativa en general, y de la universitaria en especial, con el fin de cerrar la brecha existente entre el aula y los requerimientos del mercado, sin perder de vista una formación universal sólida. Con el objetivo de ampliar la oferta de contenidos específicos a los estudiantes, la Facultad de Traducción e Interpretación de la Universidad de Las Palmas de Gran Canaria incorporó hace cuatro cursos la asignatura optativa "Localización de software". Dicha asignatura trata de familiarizar a los estudiantes con las principales herramientas, procedimientos y normas existentes en este ámbito. Un pilar importante en cualquier proyecto de localización es una gestión terminológica eficaz a través de toda la cadena de producción desde el desarrollo del software hasta su localización. En el marco de esta asignatura hemos intentado desarrollar un modelo de entrada terminológica específica para la localización de software con el objetivo de intentar superar las limitaciones de los enfoques tradicionales reinantes en este campo. Las actividades del aula se enmarcan en otras iniciativas (proyectos con otras universidades, investigación, colaboración con fabricantes de tecnología de localización), dado que solamente el esfuerzo común e interdisciplinario entre todos los estamentos interesados en la localización permite aumentar la calidad en los productos, procesos y condiciones laborales.

1. Introducción

1.1. Antecedentes

En el marco de la asignatura "Localización de software" hemos detectado que gran parte de las unidades objeto de localización no han podido ser contextualizadas (imagínense una unidad como "fFpPdDtTrRcCIL" sin más información), lo que desembocaba a menudo en una ineficaz búsqueda tipo *trial-and-error*, o que nos obligó a ponernos en contacto con el fabricante de software. A raíz de esta experiencia, decidimos dotar estas unidades con información adicional para que pudieran ser contextualizadas adecuadamente y sin pérdida de tiempo durante la localización. Un pequeño cálculo respecto a la necesidad de una mayor contextualización: supongamos que una agencia se dedica exclusivamente a la localización de software con un volumen anual de, por ejemplo, tres millones de palabras y que, a pesar de las instrucciones del cliente y después de la revisión del software original por parte del director del proyecto de la agencia, la localización de una de cada 100 unidades de

localización con una media de 10 palabras requeriría una búsqueda de solamente un minuto para ser adecuadamente traducida o localizada. Supongamos además que esta agencia localiza hacia diez lenguas (o *locales*). Con estos datos, la falta de contexto provocaría una pérdida de 500 horas, el equivalente a más de 60 días laborales.

Por otra parte, se debe señalar también que los glosarios integrados en las herramientas de localización para la reutilización de las unidades de localización en actualizaciones u otros proyectos son una mera ecuación de designaciones y no satisfacen las exigencias de una gestión terminológica adecuada. Es cierto que las bases de datos terminológicas de los fabricantes de software y de las grandes agencias de localización disponen de modelos de entrada terminológica mucho más elaborados, sin embargo dan poca importancia a las características estructurales y funcionales del software objeto de localización, lo que dificulta su contextualización. A esto hay que añadir que tanto los fabricantes de software como las agencias diseñan sus bases de datos terminológicas con formatos propietarios sin tener en cuenta métodos de modelación y formatos estandarizados, lo que impide o dificulta considerablemente el intercambio de datos.

Finalmente, nos encontramos con que predomina un enfoque diferenciado en la industria de la localización de software, según el cual una parte de las unidades objeto de localización se almacena preferiblemente en sistemas de gestión terminológica (menús y cuadros de diálogo), mientras que otra parte sustancial (unidades de localización de la categoría *string table*, es decir mensajes del sistema, texto flotante, etc.) se guarda en memorias de traducción. Dadas las características de las memorias de traducción, que solamente permiten documentar pocos datos administrativos, la localización adecuada de este tipo de unidades resulta a menudo muy ardua.

1.2. Objetivos y metodología

Ante este panorama hemos intentado poner a prueba a las herramientas específicas así como a los formatos y a los procedimientos estandarizados del ámbito de la localización con el fin de señalar puntos críticos y, si cabe, indicar soluciones a las empresas e instituciones correspondientes. En el experimento participó la Facultad de Lingüística Aplicada, Interpretación y Traducción de la Universidad del Saarland y contamos con la colaboración de la empresa alemana PASS Engineering GmbH que nos prestó gratuitamente la versión completa de su herramienta de localización Passolo 4.0 – Team Edition. Por otra parte, completamos el entorno de herramientas con el sistema de gestión terminológica MultiTerm iX de la empresa Trados.

Los objetivos del experimento consistieron en:

1. Discutir la conceptualidad de las unidades objeto de localización con la intención de justificar su introducción en una base de datos terminológica,
2. Desarrollar un modelo de entrada terminológica específico para la localización de software basado en estándares y principios reconocidos con el fin de facilitar el intercambio de datos sin pérdida de información,
3. Implementar el modelo de entrada en MultiTerm, que interactuaría con la herramienta de localización.

Tras la aclaración de la conceptualidad de las unidades de localización (ISO 1087, E DIN 2342), se han identificado los datos y procedimientos relevantes con el fin de plasmar esta información en un modelo de entrada terminológica básico, descriptivo y orientado a la localización de software. El desarrollo del modelo se ha basado en los métodos de modelación especificados en la ISO 16642 (2003) "Computer applications in terminology – Terminology markup framework (TMF)". Los datos identificados se han atribuido a las categorías de datos propuestas en la ISO CD 12620-2 (2001) "Computer applications in terminology – Data categories" que está estrechamente sintonizada con la ISO 16642. Para aquellos datos que no han podido ser documentados por las categorías de datos de la citada norma, se han especificado nuevas categorías de datos adecuadas atendiendo al método de descripción indicado en la misma norma. Donde ha sido posible, se ha tenido en cuenta el vocabulario del formato de intercambio XLIFF (XML Localisation Interchange File Format), dado que se espera una interfaz entre este estándar y el formato TBX (TermBase eXchange Format). En un segundo paso, las categorías de datos han sido estructuradas de acuerdo con el método de descripción de la ISO 16642 y representados conforme a las especificaciones TBX, un formato promovido por la LISA (Localisation Industry Association Standards) y el futuro estándar para el intercambio de datos terminológicos.

El desarrollo del modelo no ha sido sometido a combinaciones explícitas de lenguas. Lo que se ha intentado, en la medida de lo posible, es llegar a conclusiones generales a través de un enfoque supralingüístico.

2. Nociones básicas

2.1. Localización de software

Generalmente, la localización (abreviado L10N) designa un proceso en el que un producto o un servicio se adapta a las condiciones de uno o varios mercados meta. Una localización eficaz requiere que las variables específicas del mercado meta se separen de los elementos neutros. Las variables localizables pueden ser de índole lingüística (texto, formato, tipografía, ilustraciones), técnica (elementos de interacción, equipamiento), económica (sistemas de pago) y legal (responsabilidad

civil del fabricante). Se localizan productos de sectores industriales diversos como automóviles, electrodomésticos o maquinaria. No obstante, en la mayoría de los casos, el término 'localización' designa implícitamente la traducción y adaptación de un producto de software (ejecutable, ayuda, documentación, contenidos web). El término *localization* procede de la palabra *locale* (esquema regional) y define un conjunto de códigos e informaciones (conjunto de caracteres, formatos de fecha, hora y moneda, teclados, resoluciones de pantalla, orientación del texto y métodos de entrada) que se atribuyen a un determinado esquema regional como, por ejemplo, a la parte flamenca de Bélgica o los Países Bajos.

Si se recurre al concepto de traducción funcional (Reiß/Vermeer 1984, Höning/Kußmaul 1984, Schmitt 1999), los conceptos de *localización* y *traducción* serían sinónimos, dado que los funcionalistas no definen la traducción como una trascodificación de palabras o frases de un idioma a otro, sino como una acción compleja en la que se informa bajo condiciones funcionales y culturales nuevas en una situación nueva sobre un contenido de partida. Sin embargo, en la industria de la localización el proceso de la traducción, a menudo, se considera simplemente como una parte de la localización y hace referencia a la mera transferencia de texto (Esselink 2000, Kano 2000, Symmonds 2002). No obstante, desde el punto de vista funcionalista, entre la localización y la traducción no existe una relación 1:1; se trataría más bien de una inclusión. La traducción puede considerarse siempre como localización, pero la localización no siempre se puede equiparar a la traducción (por ejemplo, la adaptación constructiva de un vehículo a una cultura).

El término "software" tiene distintas acepciones. A continuación, "software" hace referencia a programas ejecutables y librerías dinámicas desarrolladas con el lenguaje de programación C++. El corpus del análisis constituyeron las categorías "menú", "cuadro de diálogo" y *string table*. Las unidades no verbales como bitmaps, iconos o cursores no se han tenido en cuenta.

2.2. Unidad de localización

En el lenguaje de los desarrolladores o programadores de software, el conjunto de informaciones que definen las interfaces del software, a través de las cuales el usuario comunica o interactúa con la máquina, se llaman recursos (inglés *resources*) distinguiéndose así del resto del código del programa invisible en tiempo de ejecución. Estos recursos se separan del resto del código fuente y son guardados de forma centralizada en archivos de recursos (RC, RES) que constan de categorías como menú, cuadro de diálogo o *string table*. Estas categorías, a su vez, contienen la información específica relativa a cada una de las opciones de menú, de los cuadros de diálogo o de los mensajes del sistema. Un botón con el texto de interfaz "OK" visualizado en tiempo de ejecución estaría acompañado, p. ej., de la siguiente información:



CONTROL "OK", 1026, "BUTTON", BS_PUSHBUTTON | WS_CHILD
| WS_VISIBLE | WS_TABSTOP, 298, 174, 50, 14

Gráfico 1. Unidad de localización "OK".

Dado que el término "interfaz del usuario" hace referencia a lo que el usuario ve o percibe en la pantalla, hemos optado por el término "unidad de localización", ya que contempla el hecho de que el traductor o localizador no trata directamente con las interfaces del usuario, sino que traduce o localiza una unidad en formato de código fuente que, en tiempo de ejecución, se *convierte* en una interfaz del usuario. Un ejemplo: no se localiza el mensaje del sistema "Can't open file test.doc", sino aquellos textos redactados por el programador en el código fuente, es decir el mensaje "Can't open file %s" y, si es necesario, los nombres de los archivos que pueden representar la variable del mensaje. Como ya se ha señalado anteriormente, el conjunto de estas unidades recibe el nombre de recursos. Ante la opacidad, sobre todo del singular "recurso", se ha creado el término "unidad de localización" por reflejar de forma más adecuada su concepto.

2.3. Método binario de localización

Generalmente, los archivos ejecutables se localizan en herramientas de localización como Passolo, Catalyst, RC-WinTrans o herramientas no comerciales desarrolladas por los fabricantes del mismo software objeto de localización. Estas herramientas permiten la extracción de las unidades de localización de sus correspondientes archivos binarios, su localización, así como la generación de la versión localizada (gráfico 2).

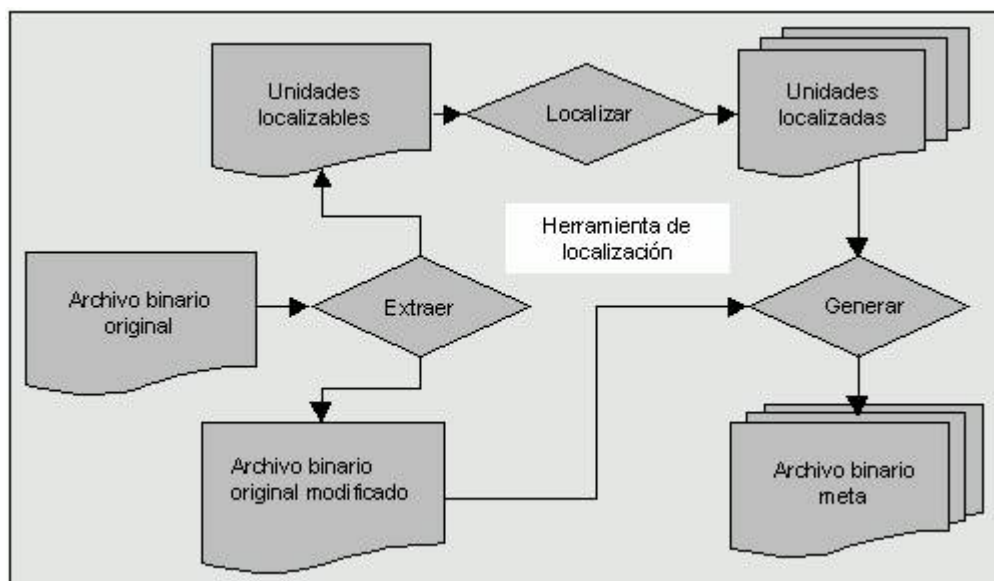


Gráfico 2. Método binario.

Entre las ventajas del método binario figura, entre otros, la posibilidad de trabajar en modo WYSIWYG, lo que permite el control y la adaptación inmediata de los elementos

gráficos de los cuadros de diálogo a las dimensiones del texto correspondiente. Como inconveniente hay que señalar la falta de integración con otras herramientas que apoyan la localización como los sistemas de gestión terminológica. En algunos casos, las herramientas de localización integran una interfaz que posibilita la interacción con memorias de traducción (RC-WinTrans para Translator's Workbench de Trados) o las herramientas de localización están dotadas con entornos de desarrollo, mediante las cuales se pueden programar dichas interfaces (Passolo).

3. Conceptualidad de las unidades de localización

Según la ISO 1087-1 (2000:2) "Terminology Work – Vocabulary. Theory and application" un concepto es "a unit of knowledge created by a unique combination of characteristics" obtenido mediante la abstracción, es decir la determinación de características comunes de una serie de objetos similares. La E DIN 2342 (enero 2003:4) "Begriffe der Terminologielehre" se distingue de la primera solamente en el *definiens*:

Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird.

En el caso de la opción de menú "Print...", por ejemplo, podríamos mencionar las características "lenguaje de programación", "entorno de desarrollo" o "aplicación", de la que la opción de menú forma parte. Sin embargo, la característica común al conjunto de objetos "Print..." sería la característica "Comando que inicia el proceso de la impresión".

Las unidades de localización representan generalmente operaciones y mensajes cortos y precisos así como texto estático breve. En el caso de las opciones de los menús y de los cuadros de diálogo, se trata de instrucciones que se verbalizan mediante designaciones simples o compuestas (File, Tabla, Datei öffnen), mientras que los mensajes constan de una o varias frases o fragmentos de ellas (Cannot create the %% file.\n\nMake sure that the path and filename are correct.). Las opciones de los menús y de los cuadros de diálogo se ajustan a la idea tradicional de concepto, sobre todo debido a la concisión de sus designaciones. Asimismo, una opción de menú "Print ..." podrá considerarse sin duda como concepto.

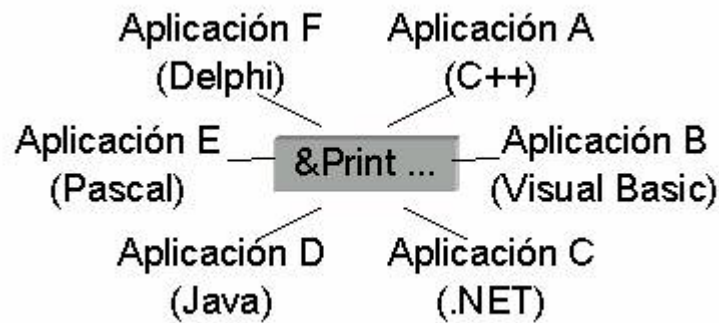


Gráfico 3. Unidad de localización "Print ..."

Estas consideraciones pueden aplicarse también a las opciones de los cuadros de diálogo donde predominan designaciones mono- o pluriverbales cortas que representan un concepto bien definido y delimitado. Algo más problemático en cuanto a la relación representación-conceptualidad se presentan las unidades de localización pertenecientes a la categoría *string table* como el siguiente mensaje del sistema:

Cannot open the %% file. Make sure a disk is in the drive you specified.

La primera frase podría omitirse dada la redundancia del mensaje en el contexto de acción del usuario (todavía se acordará que quiso abrir un archivo hace unos instantes). Quedaría por lo tanto:

Make sure a disk is in the drive you specified.

Incluso la fórmula de cortesía 'Make sure' podría eliminarse, de forma que solamente quedaría la siguiente expresión:

No disk in drive you specified.

Llegado hasta aquí, nos acercamos a las nociones tradicionales de la terminología. Sería digno reflexionar sobre si se ha recurrido a representaciones lingüísticas más explícitas únicamente con el fin de ayudar al usuario (lego) a entender con más facilidad los procesos de la máquina, y que, a primera vista, no nos parecen conceptos. La importancia de la verbalización de los conceptos en el campo del software se refleja también en los siguientes ejemplos:

Open an existing document\nOpen
Quit the application; prompts to save documents\nExit

Estas unidades de localización constan de dos partes. La parte a la izquierda de "\n" se visualiza en la barra de estado cuando el usuario apunta la correspondiente opción del

menú con el ratón, que se encuentra a la derecha de "\n". Ambas partes representan *un* concepto. Su diferencia consiste únicamente en los medios lingüísticos utilizados para su representación. Por consiguiente, parece conveniente y cercano a las teorías de la terminología de tipificar las unidades de localización como conceptos y, por lo tanto, introducirlos en sistemas de gestión terminológica.

4. Modelación de datos según la ISO 16642

Tras la extracción de las unidades de localización y su adaptación, estas unidades se han introducido íntegramente en bases de datos terminológicas para su adecuada reutilización en actualizaciones o nuevos proyectos de localización. Debido a la implicación de un gran número de personas que participan en un proyecto de localización (desarrolladores, programadores, redactores, directores de proyectos, traductores, correctores, etc.) y dada la creciente descentralización de las tareas a través de la web, se imponen métodos, estructuras e interfaces estandarizados con el fin de minimizar las redundancias en el trabajo terminológico y de garantizar el intercambio de datos, su diseminación y la interoperabilidad entre diferentes sistemas sin pérdida de información. En este sentido, la ISO/TC 37 (Terminology and other language resources), Subcommittee SC 3 (Computer applications in terminology) ha elaborado y aprobado la norma ISO 16642 (2003) "Computer applications in terminology – Terminological Markup framework (TMF)" que propone un marco metodológico genérico para la descripción de todo tipo de recursos terminológicos y las reglas para la especificación de lenguajes de marcas terminológicos (Terminological Markup Language - TML).

4.1. Terminological Markup Framework (TMF)

La ISO 16642 se basa en un enfoque integrador para el análisis y la modelación de bases de datos terminológicos ya existentes o nuevos en sistemas relacionales o basados en texto. La descripción de la información terminológica en XML permite documentar los datos de forma estructurada, importarlos/exportarlos fácilmente y actualizarlos sin la necesidad de importarlos previamente en un sistema de gestión terminológica. Para la modelación de los datos terminológicos, la norma propone, en el nivel de abstracción superior, un metamodelo genérico que puede ser concretado por el diseñador en el nivel de abstracción inferior mediante modelos de datos específicos. Los modelos de datos descritos en XML se definen a través de un *Document Type Definition* (DTD) o un esquema XML. La implementación de perfiles específicos del metamodelo así como la concreción del modelo de datos en XML conducen a diferentes lenguajes de marcas terminológicos (TML). En un DTD/esquema XML se define:

1. Cómo se implementa el metamodelo,
2. Cuáles son las categorías de datos específicas utilizadas y cómo se relacionan con los niveles del metamodelo,

3. Cómo se representan estas categorías de datos en XML (*style*), y
4. Cómo se expresan, como elementos y atributos XML (*vocabularies*), los diferentes objetos de información.

4.2. Terminological Markup Language (TML)

Para la especificación de los lenguajes de marcas terminológicos (TML) son necesarios determinados recursos de conocimiento que están relacionados entre sí en diferentes niveles (gráfico 4). Es común a todos los TML el hecho que referencian *un* metamodelo, y que las categorías de datos utilizadas se extraen de *un* registro central de categorías de datos (DCR – Data Category Registry = ISO 12620). Se pueden definir perfiles específicos mediante especificaciones de categorías de datos (DCS – Data Category Specification) y dialectos (combinación de los niveles en el metamodelo y la forma de representación de las categorías de datos).

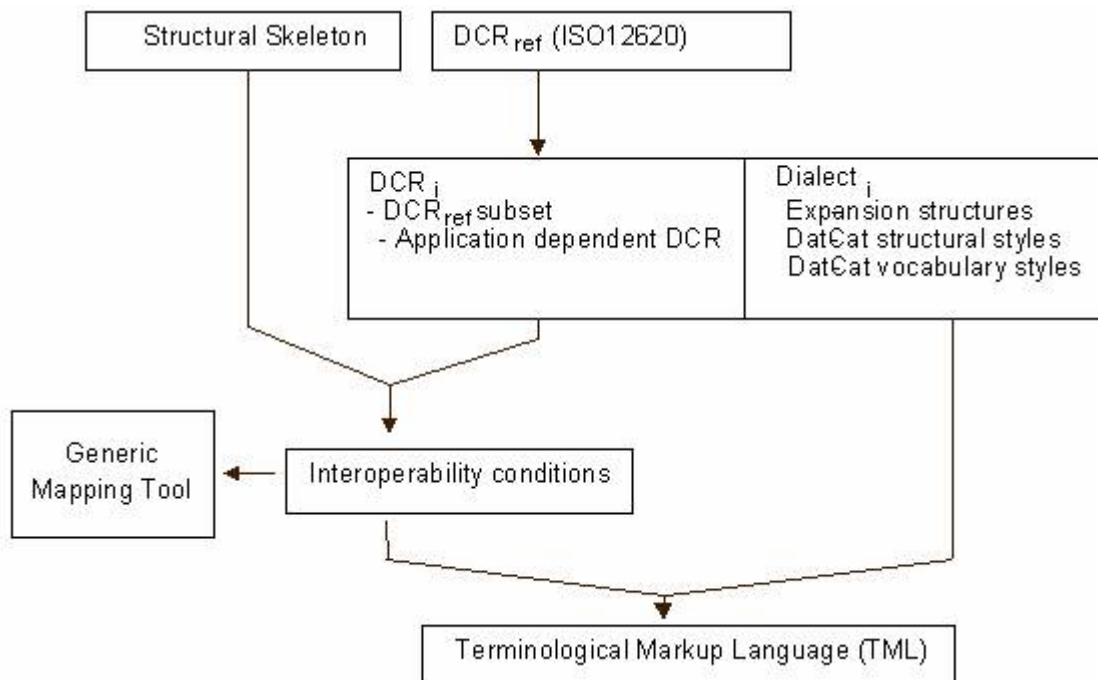


Gráfico 4. Recursos de conocimiento para la descripción de TML (ISO 16642:10).

La combinación del metamodelo con la especificación de las categorías de datos (DCS) es suficiente para determinar las condiciones de interoperabilidad entre recursos terminológicos de distinta estructuración y, por lo tanto, entre diferentes TML. Los TML pueden compararse por medio de su especificación y pueden ser modelados para ser intercambiados con la ayuda de herramientas genéricas (GMT - Generic Mapping Tool) conservando la totalidad de la información terminológica.

4.3. Metamodelo terminológico

El metamodelo (gráfico 5) definido en la ISO 16642 se basa en las recomendaciones de la ISO 704 (2000) "Terminology work - Principles and methods". Contiene una jerarquía de niveles de información a los que pueden ser atribuidas determinadas

categorías de datos. Las especificaciones del metamodelo pueden ser consultadas en la citada norma.

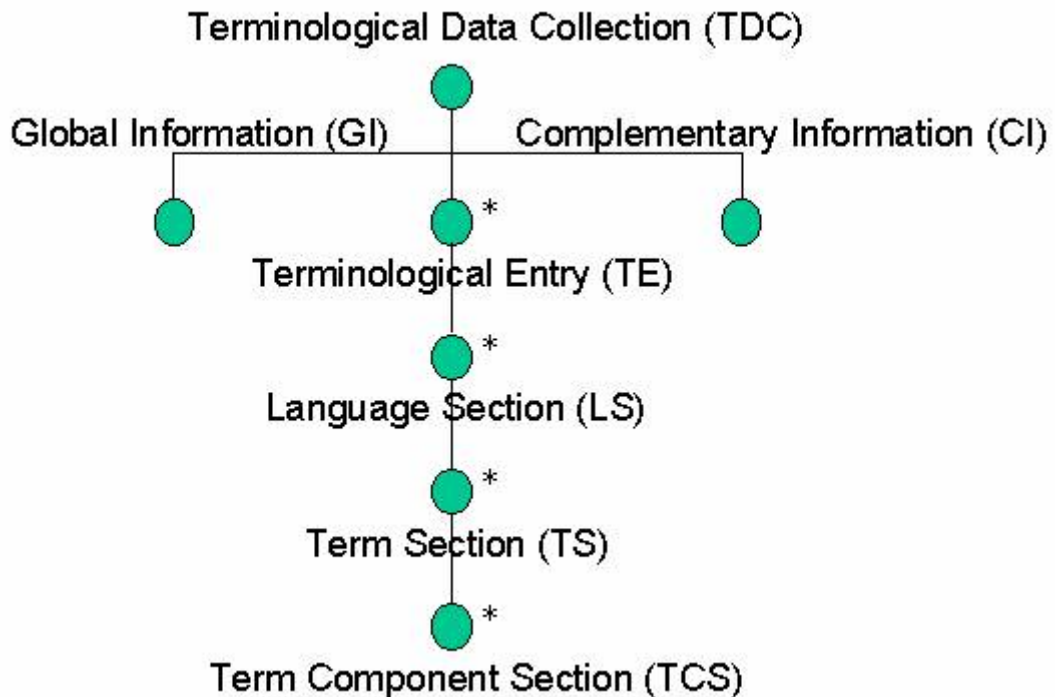


Gráfico 5. Niveles del metamodelo (ISO 16642:18).

4.4. MARTIF with Specified Constraints (MSC) y TBX

MSC, el lenguaje de marcas terminológico (TML) para la especificación del formato flexible de intercambio TBX (TermBase eXchange Format), está basado en el formato de intercambio MARTIF (Machine-readable Terminology Interchange Format) y contiene un subgrupo de datos extraídos (DCS) de la norma ISO 12620. Cada implementación del metamodelo especifica la forma en la que expanden los niveles del metamodelo y cómo se expresan las categorías de datos en XML. La instanciación de categorías de datos específicos se indica por medio de metacategorías de datos y el valor de un atributo *type* (<descrip type="definición">... </descrip>). Esta forma de representación permite una validación separada de la estructura y de las categorías de datos, lo que favorece particularmente a aquellos terminólogos entre los diseñadores de bases de datos que no están familiarizados con los mecanismos de validación en XML.

El máximo nivel de un documento TBX constituye el elemento *martif*, que a su vez consta de un elemento *martifHeader* y de un elemento *text* (gráfico 6). El elemento *martifHeader* contiene una descripción verbal del conjunto de los datos terminológicos (elemento *fileDesc*), informaciones sobre las categorías de datos utilizadas y sobre la codificación de los códigos (elemento *encodingDesc*), así como información acerca de revisiones efectuadas (*revisionDesc*), siendo opcionales los últimos elementos indicados por un signo de interrogación.

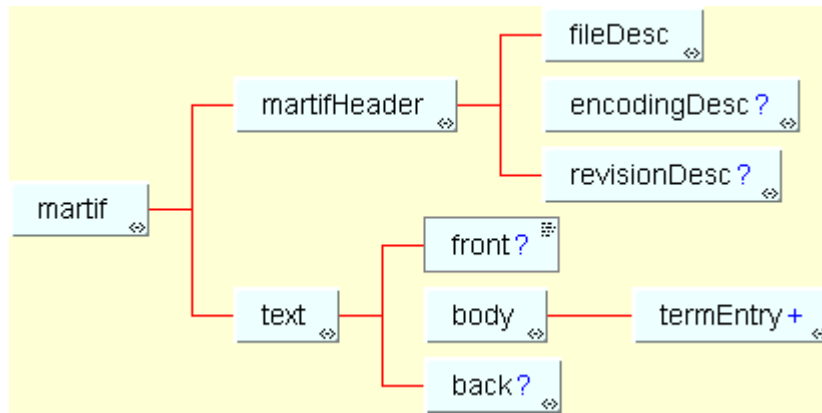
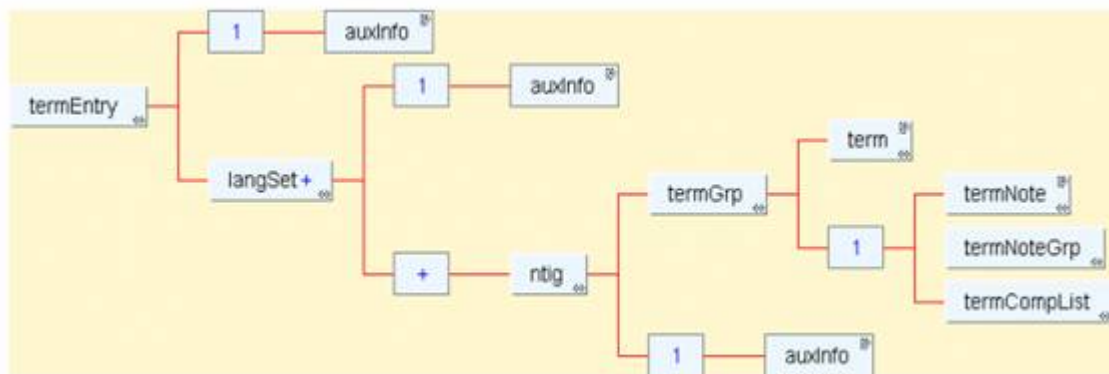


Gráfico 6. Niveles superiores en TBX (LISA 2002b:11).

El elemento *text* consta de los elementos *front*, *body* y *back*. Los elementos *front* y *back* acogen, por ejemplo, datos administrativos o bibliográficos, datos binarios (gráficos, vídeo o audio) o, incluso, enlaces hacia las informaciones mencionadas y hacia otros conjuntos de datos terminológicos. Del elemento *body* cuelgan las entradas terminológicas (*termEntry*) (gráfico 7).



entry level language level term level

Gráfico 7. Estructura de la entrada terminológica en TBX (LISA 2002b:12).

El elemento *auxInfo* contiene información descriptiva y administrativa relativa a la entrada, al concepto, a la lengua y a la denominación. Cada bloque de lenguas (*langSet*) contiene categorías de datos específicos de la lengua y el bloque de denominación (*ntig* – nested term information group) comprende la totalidad de la información relacionada con cada una de las denominaciones. Los elementos *termNote* y *termNoteGrp*, respectivamente, sólo pueden ser atribuidos al nivel *term* y a los niveles inferiores. En el elemento *termCompList* pueden especificarse los componentes de un grupo de palabras.

4.5. Modelo de entrada específico para la localización de software

El modelo de entrada terminológica básico recoge las especificadas de las unidades de localización. Se presenta como documento estructurado en formato XML y carece de valores específicos. En su lugar, se indican la clase del tipo de datos (o de la dirección del enlace) que puede ser asignado a una categoría de datos según las especificaciones TBX. La numeración continua en el margen izquierdo del modelo básico no forma parte de la estructura XML, sino que sirve como notas al pie del documento. Los comentarios detallados hacen referencia exclusivamente a las características específicas del modelo.

- [1]

```
<?xml version='1.0'?>
<!DOCTYPE martif SYSTEM ".\TBXcoreStructureDTD-v-1-0.DTD">
<martif type='TBX' xml:lang='en'>
  <martifHeader>
    <fileDesc><sourceDesc><p>noteText</p></sourceDesc></fileDesc>
    <encodingDesc><p type='DCSName'>TBXdefaultXCS-v-1-0.XML</p></encodingDesc>
  </martifHeader>

  <text> <body>
    <termEntry id="plainText">
```
- [2]

```
<transacGrp>
  <transac>picklist </transac>
  <date>date (ISO format) </date>
  <transacNote type="responsibility" target="respPersonId" >noteText </transacNote>
</transacGrp>
```
- [3]

```
<descrip type="subjectField">plainText </descrip>
<descrip type="conceptOrigin">picklist </descrip>
<descrip type="conceptPosition" target="conceptSysDescripId" >noteText </descrip>
<descrip type="superordinateConceptGeneric" target="entryId" >noteText </descrip>
<descrip type="coordinateConceptGeneric" target="entryId" >noteText </descrip>
<descrip type="subordinateConceptGeneric" target="entryId" >noteText </descrip>
<descrip type="superordinateConceptPartitive" target="entryId" >noteText </descrip>
<descrip type="coordinateConceptPartitive" target="entryId" >noteText </descrip>
<descrip type="subordinateConceptPartitive" target="entryId" >noteText </descrip>
<descrip type="sequentiallyRelatedConcept" target="entryId" >noteText </descrip>
```
- [4]

```
<langSet xml:lang="language identifier">
<descripGrp>
  <descrip type="definition">noteText </descrip>
  <admin type="sourceIdentifier" target="biblId" >noteText </admin>
</descripGrp>
```
- [5]

```
<ntig>
```
- [6]

```
<termGrp>
  <term>noteText </term>
  <termNote type="termType">picklist </termNote>
  <termNote type="partOfSpeech">picklist </termNote>
  <termNote type="grammaticalGender">picklist </termNote>
  <termNote type="grammaticalNumber">picklist </termNote>
  <termNote type="mainResourceType">picklist </termNote>
  <termNote type="resourceGroupId">plainText </termNote>
  <termNote type="resourceId">plainText </termNote>
  <termNote type="menuType">picklist </termNote>
  <termNote type="controlType">picklist </termNote>
  <termNote type="positionCoordinates">plainText </termNote>
  <termNote type="sizeCoordinates">plainText </termNote>
  <termNote type="stringType">picklist </termNote>
  <termNote type="stringVariableOptions">plainText </termNote>
  <termNote type="productSubset">noteText </termNote>
  <termNote type="environmentSubset">noteText </termNote>
  <termNote type="projectSubset">noteText </termNote>
  <termNote type="customerSubset">noteText </termNote>
```
- [7]

```
<termCompList type="termElements"><termCompGrp>
  <termComp>noteText </termComp>
  <termNote type="partOfSpeech">picklist </termNote>
  <termNote type="grammaticalGender">picklist </termNote>
  <termNote type="grammaticalNumber">picklist </termNote>
```

```

        </termCompGrp>
    </termCompList>
</termGrp>
[8] <descripGrp>
    <descrip type="explanation">noteText </descrip>
    <admin type="sourceIdentifier" target="biblId" >noteText </admin>
</descripGrp>
<descripGrp>
    <descrip type="context">noteText </descrip>
    <descripNote type="contextType" >picklist </descripNote>
    <admin type="sourceIdentifier" target="biblId" >noteText </admin>
</descripGrp>
<descripGrp>
    <descrip type="figure" target="binaryDataId" >noteText </descrip>
    <admin type="sourceIdentifier" target="biblId" >noteText </admin>
</descripGrp>
    <admin type="sourceIdentifier" target="biblId" >noteText </admin>
</ntig>
</langSet>
</termEntry>
</body> </text>
</martif>

```

[1] Véase subcapítulo 5.4

[2] Los datos administrativos sobre la fecha y la responsabilidad referente a la introducción y modificación de datos en la entrada terminológica se documentan en el elemento *transacGrp*. El tipo de la actividad terminológica (input, modification, etc.) está encapsulado en el elemento *transac*. El atributo *target* contiene un valor enlazado (respPersonId) que apunta a otros datos fuera de las entradas terminológicas.

[3] Al nivel TE se asignan, además de las categorías de datos administrativos anteriormente mencionadas, las categorías de datos relativas al concepto, que permiten introducir información sobre el origen del concepto, el campo especializado, la posición del concepto y las relaciones conceptuales.

[4] El bloque *langSet* (nivel LS) engloba toda la información específica de una lengua y puede repetirse cuantas veces sea necesario dentro de la entrada. A su vez, el nivel *langSet* acoge la definición para todas las posibles designaciones de un concepto en una lengua. La categoría de datos "definition" se ha asignado a este nivel, puesto que, en la práctica de la localización, una definición adecuada no existe siempre en la lengua pivot. Directamente subordinado al nivel LS se encuentra el elemento *ntig* – *nested term information group*, bloque que acoge todos los datos relacionados con una designación.

[5] El bloque de la designación (*ntig*) y las respectivas categorías de datos sólo figuran una vez por lengua en el presente modelo de entrada. Este bloque puede repetirse dentro del bloque *langSet* cuantas veces sea necesario.

[6] En el elemento *termGrp* se documentan las designaciones y los datos relacionados con la designación. Este grupo comprende categorías de datos relativas a la designación, que en la especificación TBX se definen como categorías de datos administrativos (project subset, environment subset, etc.). No obstante, Arntz/Picht (2002:236) y Schmitz (2000:141) argumentan que estas categorías pueden o incluso deben asignarse a la designación. La mayor parte de las nuevas categorías de datos figuran en el nivel *termGrp* como "main resource type" (categoría de recursos a la pertenece la unidad de localización), "control type" (tipo de elemento de una cuadro de diálogo), "string type" (tipo de unidad de localización perteneciente a la categoría de recursos *string table*) o "string variable options" (posibles valores de una variable de un mensaje de error, p. ej.), puesto que documentan una determinada unidad de localización de un determinado software.

- [7] Al elemento *termCompList* se asignan los elementos *termCompGrp*, que, a su vez, incorporan información sobre los componentes de una designación pluriverbal. El elemento *termCompGrp* puede repetirse cuantas veces sea preciso. La necesidad de gestionar estos componentes se da particularmente en lenguas analíticas (español, francés), donde circunstancialmente puede precisarse información gramatical para cada una de las palabras con el fin de poder traducir de forma adecuada. En el caso del componente, si se trata de otro concepto, puede enlazarse con la entrada correspondiente.
- [8] Las categorías de datos "explanation", "context", "figure" encapsuladas en los elementos *descripGrp* se definen en las especificaciones TBX como categorías de datos relativas al concepto. Para el presente modelo, estas categorías, junto con las respectivas fuentes, se han asignado al nivel *ntig*, dado que, p. ej., la categoría "explanation" documenta las explicaciones respecto a la generación de un determinado mensaje de error, la categoría "figure" o la categoría "context" acoge una captura de pantalla con la respectiva designación, etc.

Una vez desarrollado el modelo, procedimos a su implementación en un software comercial de gestión terminológica.

5. Implementación del modelo de entrada terminológica

Sin duda, Trados MultiTerm es uno de los sistemas de gestión terminológica más utilizados, y, por lo tanto, ha sido elegido, junto con Passolo, para nuestro experimento con el fin de dar la máxima proximidad a la realidad del mercado. Actualmente, Trados comercializa la versión MultiTerm iX. Sin embargo, la última versión 4.0 de Passolo solamente viene con una interfaz para exportar las unidades de localización hacia la versión anterior, MultiTerm 5, lo que obliga convertirlas por medio de la herramienta MultiTerm Convert de Trados al correspondiente formato XML para poder importarlas en MultiTerm iX.

MultiTerm iX permite una gestión orientada al concepto, lo que garantiza la documentación de todas las designaciones referentes a un concepto en una entrada terminológica. Para cada entrada se puede establecer un número indeterminado de lenguas (o *locales*). Las categorías de datos pueden ser definidas sin restricciones, mientras que la estructura básica de la entrada terminológica está predefinida y consta por defecto de los tres niveles "Entry level", "Index level" y "Term level", jerarquía que corresponde al metamodelo definido en la ISO 16642 y en las especificaciones del formato TBX. MultiTerm iX está habilitado para XML, lo que flexibiliza el diseño y el intercambio de datos terminológicos, particularmente en el ámbito de las aplicaciones basadas en la web.

Todos los niveles del modelo desarrollado con sus correspondientes elementos e interdependencias han podido implementarse en MutliTerm iX. En MultiTerm, la definición de la base de datos se lleva a cabo en dos pasos. En primer lugar, se definen las categorías de datos. Estas categorías pueden ser de dos tipos: los campos de índice

(index fields) acogen las lenguas (o *locales*), mientras que los campos descriptivos (descriptive fields) documentan el concepto y las designaciones correspondientes. A estos últimos pueden ser asignados distintas propiedades (texto libre, valores predeterminados, datos externos, etc.). En segundo lugar, las categorías de datos se atribuyen a los tres niveles básicos predefinidos "Entry level", "Index level" y "Term level" (gráfico 8) o forman propios niveles (p. ej., a una categoría "term element" pueden asignarse categorías como "part of speech", "grammatical gender", etc.). Una vez definida la base de datos y la estructura de la entrada, MultiTerm permite la elaboración de un modelo de entrada o patrón con el fin de facilitar la introducción de datos terminológicos.

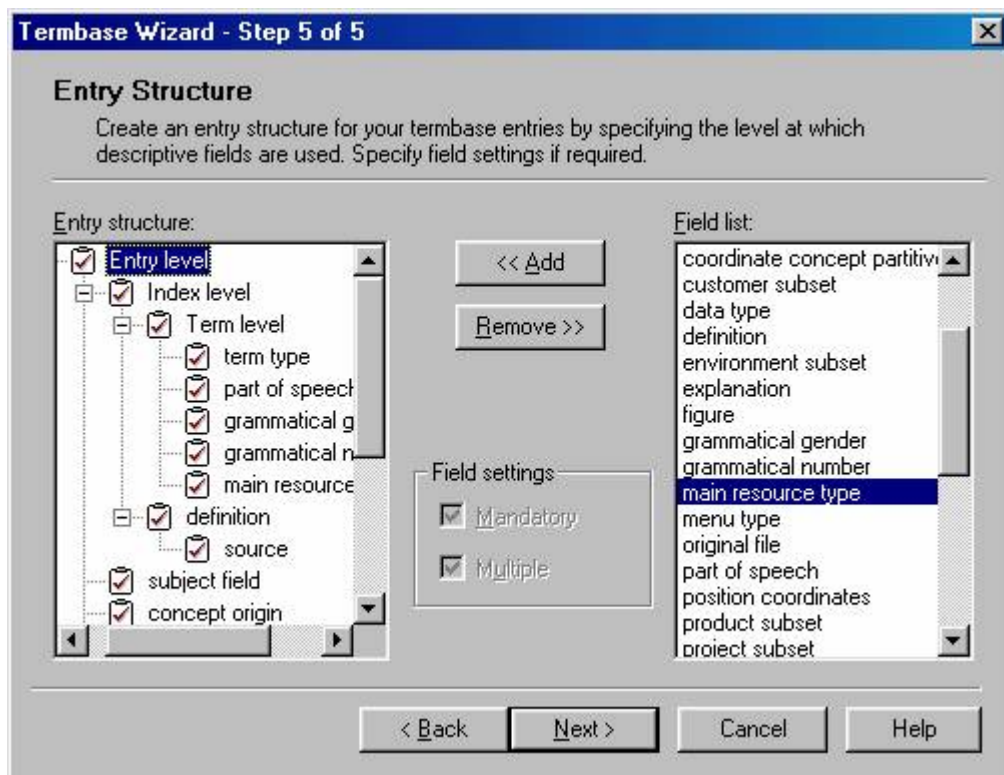


Gráfico 8. Definición del modelo de entrada en MultiTerm iX.

El modelo en formato TBX y el modelo implementado en MultiTerm no se distinguen con respecto a su contenido. Sin embargo, se constatan algunas diferencias en la manera de representar los datos en formato XML como es el caso de las etiquetas (<mtf> vs. <martif>, <concept> vs. <termEntry> o <language> vs. <langSet>) o él de la asignación de determinadas categorías de datos a metacategorías. Así, los datos exportados a través de la interfaz estándar de MultiTerm contienen las metacategorías <languageGrp> y <conceptGrp> para las respectivas categorías de datos, inexistentes en TBX. Por otro lado, en MultiTerm, las categorías de datos del nivel "Term" están atribuidos a las metacategorías "descrip", mientras que en TBX se están asignadas a la metacategoría "termNote". No obstante, dichas diferencias pueden amortizarse por

medio de la definición de interfaces personalizadas en MultiTerm o de rutinas de conversión externas.

6. Procesamiento de datos

6.1. Extraer, documentar y localizar las unidades

La extracción, la documentación y la localización de las unidades de localización se llevó a cabo con la herramienta Passolo 4.0 Team Edition. Passolo destaca por su sencilla e intuitiva interfaz del usuario e integra potentes funciones de localización (actualización de proyectos, alineación de archivos binarios, etc.). Permite localizar unidades contenidas tanto en archivos binarios como archivos fuente procedentes de los entornos de desarrollo más comunes (Microsoft Visual C++, Microsoft .NET, Borland Delphi/C++ Builder, etc.). Además, Passolo incorpora distintos *parser* para XML y otros formatos de textos con identificaciones unívocas como archivos INI, Java o bases de datos. Un proyecto abierto en Passolo tiene el siguiente aspecto:

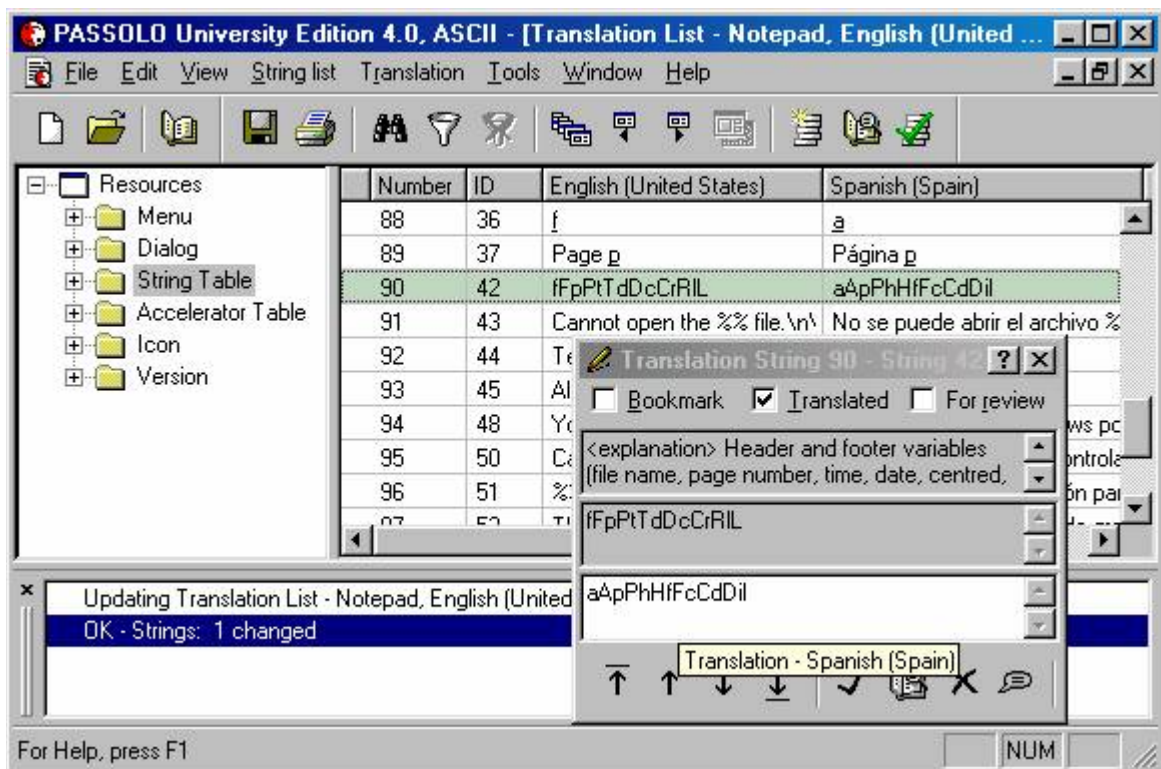


Gráfico 9. Distribución de ventanas en Passolo.

En la ventana a la izquierda, se visualizan las categorías de recursos parecidos a la estructura de MS-Windows Explorer. En la ventana inferior se indican los resultados de las distintas operaciones y la ventana a la derecha contiene una tabla con las unidades de localización extraídas, que se editan en un pequeño cuadro de diálogo. En la fila superior de este cuadro pueden añadirse comentarios respecto a la unidad de localización seleccionada. Los menús y cuadros de diálogo pueden ser editados en

módo gráfico, lo que, en el caso de los cuadros de diálogo, permite la adaptación de los elementos gráficos (botones, cajas de texto, etc.) a las dimensiones del texto.

Aparte de los formatos internos (TRX, TXT, GLO) Passolo ofrece interfaces de exportación hacia archivos CSV y formatos propietarios (Trados, STAR). Además, un entorno compatible con el lenguaje script VBA permite crear interfaces interactivas entre Passolo y TRADOS Workbench, Excel o Word.

6.2. Exportar y adaptar las unidades de localización

Las unidades de localización extraídas vienen acompañadas de la siguiente información (véase también gráfico 9): número de unidad en Passolo (number), identificación dentro del software del que ha sido extraído (ID) y el texto original. En un primer paso, se han preparado las unidades de localización asignándoles atributos (sólo lectura, ocultar) y añadiendo un comentario para su contextualización, si procede. Posteriormente, se han introducido los textos meta. Con el fin de evitar un volumen prohibitivo de trabajo manual en la adaptación del archivo exportado al formato de importación de MultiTerm, se ha elaborado una guía de estilo para sintonizar los comentarios con las categorías de datos definidas en el modelo de entrada desarrollado.

Basándonos en el enfoque integrado, según el cual todas las unidades de localización deberían almacenarse en un sistema de gestión terminológica, intentamos exportar la totalidad de las unidades a través de la interfaz de MultiTerm. No obstante, nos encontramos con que la interfaz no permite exportar unidades que comprenden más de 10 palabras (gráfico 10). Esta restricción se basa en la suposición de que las unidades con más de 10 palabras no representan conceptos. Sin embargo, ejemplos como *véhicule électrique alimenté par circuits inductifs à haute fréquence placés sous le chaussée* (Goffin, 1978 cit. s. Arntz/Picht/Mayer 2002:118) demuestran que la longitud de la designación no debería constituir un elemento definidor para los conceptos.

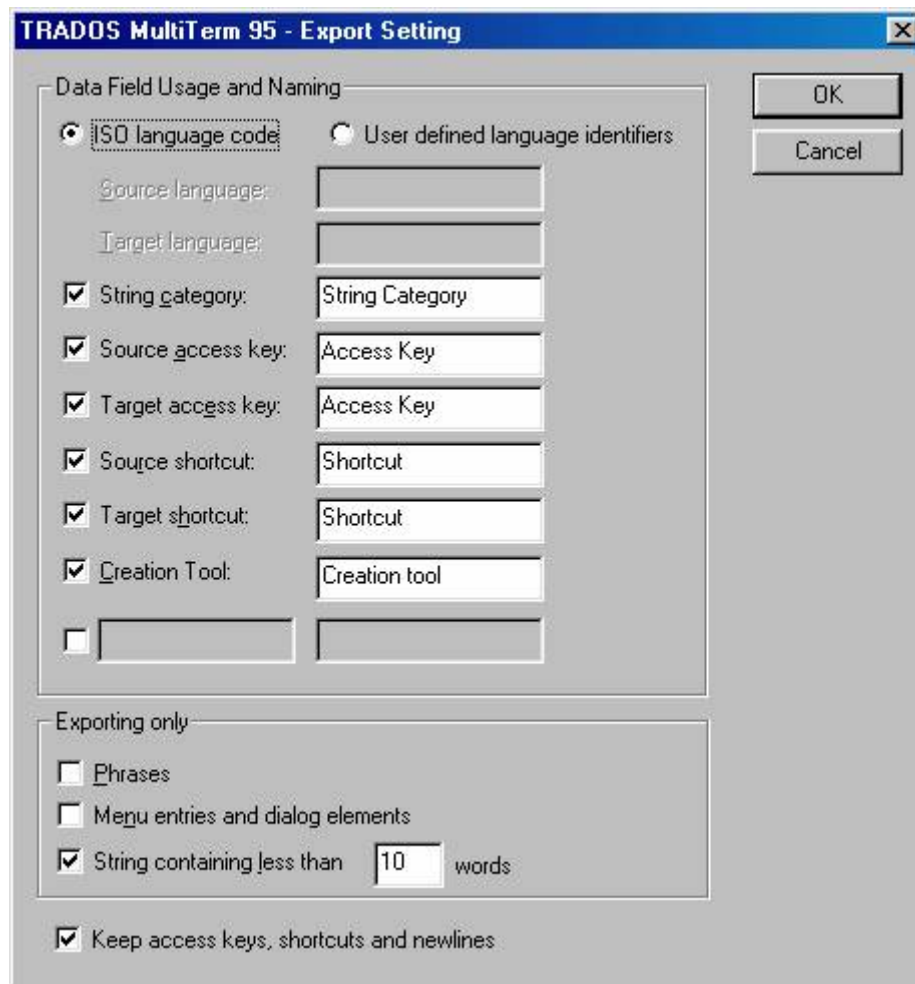


Gráfico 10. Interfaz de exportación de MultiTerm en Passolo 4.0.

Por este motivo, hemos tenido que optar por el formato de texto específico de Passolo que garantiza la exportación de la totalidad de los datos relevantes (gráfico 11).

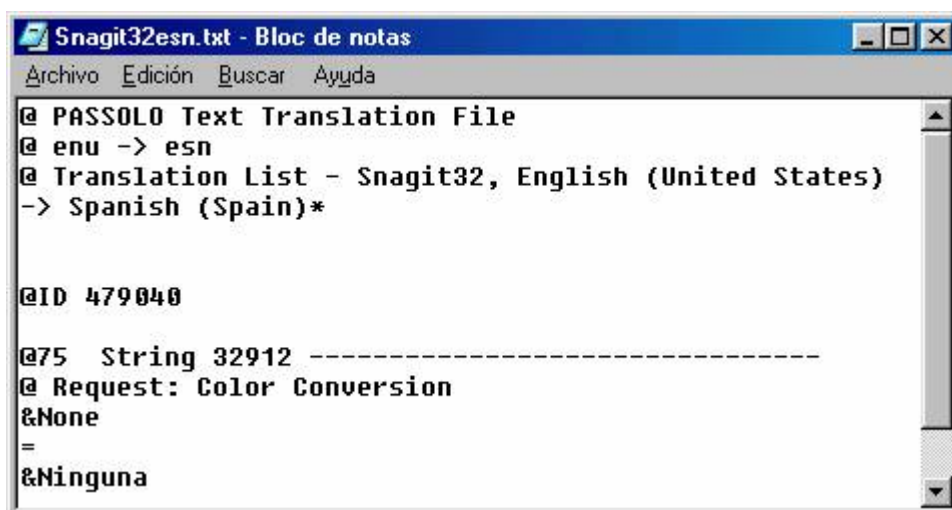


Gráfico 11. Formato de exportación de Passolo.

A continuación, el archivo de texto plano ha tenido que ser adaptado mediante rutinas de conversión a la estructura (contenido) y a la representación de datos (forma) exigidos por las especificaciones TBX y para su importación en MultiTerm iX.

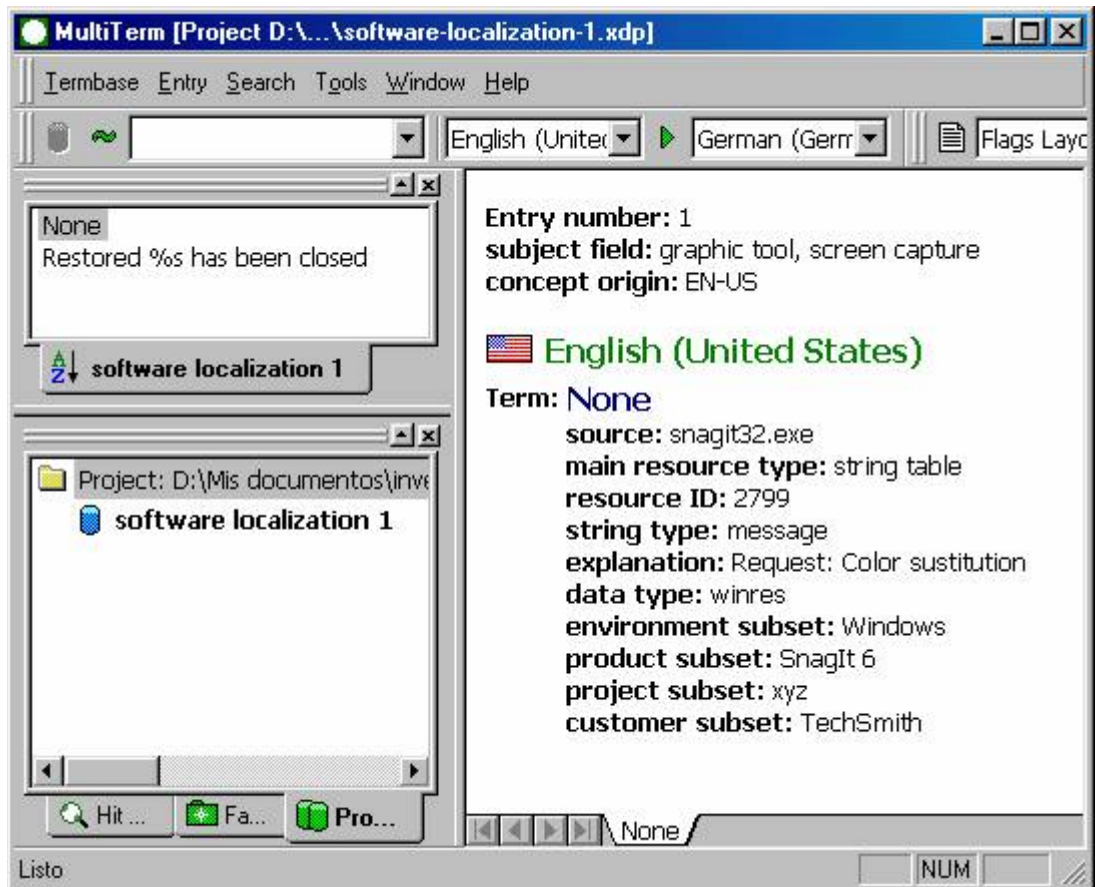


Gráfico 12. Unidad de localización en MultiTerm iX.

7. Conclusión

Nuestro experimento ha dejado patente algunas carencias en cada uno de los ámbitos de la localización (los fabricantes de software, las agencias de localización y los fabricantes de tecnología de localización). Por una parte, hemos podido constatar que el diseño de software objeto de localización se hace a menudo exclusivamente desde el punto de vista de la funcionalidad y bajo la premura de desarrollar aplicaciones que ocupen el menor espacio posible. Asimismo, se echa en falta, por un lado, una documentación más explícita de las unidades de localización para facilitar su adecuada contextualización por parte de los traductores o localizadores, y, por otro lado, un trabajo terminológico proactivo que, además, evite inconsistencias. Dado que las informaciones para una mejor contextualización no pueden ser extraídas por ser inexistentes, las agencias de localización no las incluyen en sus bases de datos terminológicos, ya que supondría un trabajo manual y no rentable desde el punto de vista económico. En cuanto a las herramientas de traducción y de localización, se percibe una mayor implementación de formatos e interfaces estandarizadas y basados

en XML como XLIFF y TMX, lo que reducirá la fragmentación de procesos y, por lo tanto, los costes de conversión. Sin embargo, queda todavía camino por recorrer en cuanto a una mayor penetración de formatos estandarizados como TBX en el campo de los métodos y de las aplicaciones relativas a la terminología. Además, sería deseable que se crearan interfaces entre XLIFF y TBX con el fin de agilizar el flujo de datos sin tener que pasar por herramientas intermediarias.

Bibliografía

- Arntz, Reiner, Picht, Heribert y Mayer, Felix (2002) *Einführung in die Terminologearbeit*, Hildesheim, Zürich, New York: Olms.
- Budin, Gerhard (2002) "Der Zugang zu mehrsprachigen terminologischen Ressourcen - Probleme und Lösungsmöglichkeiten" en Mayer, Felix, Schmitz, Klaus-Dirk y Zeumer, Jutta (eds.), *eTerminology - Professionelle Terminologearbeit im Zeitalter des Internet*, Actas del Symposio "Deutscher Terminologie-Tag e.V. Köln", 12-13 de abril de 2002, Köln: Deutscher Terminologie Tag e.V., 185-200.
- E DIN 2342 (Januar 2003) *Begriffe der Terminologielehre*, Berlin, Köln: Beuth.
- Esselink, Bert (2000) *A Practical Guide to Localization*, Amsterdam, Philadelphia: John Benjamins Publishing Company.
- ISO 704 (2000) *Terminology work - Principles and methods*, Geneva: ISO.
- ISO 1087-1 (2000) *Terminology Work - Vocabulary - Part 1: Theory and application*, Geneva: ISO.
- ISO CD 12620-1 (2001) *Computer applications in terminology - Data Categories - Part 1: Defining parameters for specifying data categories for terminology collections and other language resources*, Geneva: ISO.
- ISO 16642 (2003) *Computer applications in terminology - Terminological markup framework (TMF)*, Geneva: ISO.
- LISA (2002a) "TBX specifications, Working Draft, 5 May 2002". Disponible en URL <http://www.lisa.org/tbx>. Fecha de consulta: 15 de enero de 2004.
- LISA (2002b) "TMX 1.4a Specification, 10 July 2002". Disponible en URL <http://www.lisa.org/tmx>. Fecha de consulta: 15 de enero de 2004.
- OASIS (= Organization for the Advancement of Structured Information Standards) (2003) "XLIFF 1.1 Specification. Committee Specification, 31 Octubre 2003". Disponible en URL <http://www.oasis-open.org/committees/xliff>. Fecha de consulta: 15 de enero de 2004.
- Reinke, Uwe (2002) "Terminologiemanagement in modernen Übersetzungsumgebungen: Translation Memories und Lokalisierungstools" en Mayer, Felix, Schmitz, Klaus-Dirk y Zeumer, Jutta (eds.) *eTerminology - Professionelle Terminologearbeit im Zeitalter des Internet*, Actas del Symposio "Deutscher Terminologie-Tag e.V. Köln", 12-13 de abril de 2002, Köln: Deutscher Terminologie Tag e.V., 215-238.
- Schmitz, Klaus-Dirk y Wahle, Kirsten (eds.) (2000) *Softwarelokalisierung*, Tübingen: Stauffenburg.
- Wright, Sue E. y Budin, Gerhard (Hrsg.) (1997) *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*, Amsterdam, Philadelphia: John Benjamins Publishing Company.

- Wright, Sue E. y Budin, Gerhard (Hrsg.) (2001) *Handbook of Terminology Management. Volume 2: Application-Oriented Terminology Management*, Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Wüster, Eugen (1993) "Die Allgemeine Terminologielehre - ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften" en Laurén, Christer y Picht, Herbert (eds.) *Ausgewählte Texte zur Terminologie*, Wien: International Network for Terminology (TermNet), 331-376.