

USO DINÁMICO DE ARTÍCULOS LEXICOGRÁFICOS: UNA APLICACIÓN DE XSLT

*Jorge Antonio Leoni de León**

1. Introducción

En el artículo Propuesta de formalización del manual de redacción del Nuevo Diccionario del Español de Costa Rica a partir del Lenguaje de Etiquetado Extensible (XML) (Leoni, 2003), mostramos la versatilidad del XML y su pertinencia en la labor lexicográfica. Sin embargo, en esa ocasión sólo pudimos exponer la capacidad del XML para describir contenidos lexicográficos, faltándonos mostrar su capacidad para transformar y manipular los contenidos descritos. Por este motivo, continuando la ruta de investigación iniciada ese artículo, ahora presentaremos una aplicación del Lenguaje de transformaciones de etiquetado extensible de hojas de estilo (XSLT, por sus siglas en inglés) en la redacción lexicográfica.

En Leoni (2003) se señala un cierto número de labores pendientes en relación con la creación de un Lenguaje de descripción lexicográfica (LDL), a saber:

- Definición de una Descripción de tipo de documento (DTD), la cual contendría las reglas del LDL.
- Creación de una o varias hojas de estilo para que los documentos codificados según LDL sean legibles por un lector humano; en este sentido, habíamos señalado dos opciones:
 - Hojas de estilo en cascada (CSS), las cuales, básicamente, indican cuáles

etiquetas son visibles y, de serlo, cuáles son sus características (color, tipo de letra, tamaño, etcétera).

- Lenguaje de transformaciones de etiquetado extensible de hojas de estilo (XSLT), el cual consiste en una serie de instrucciones para transformar los contenidos de cualquier documento XML, según las necesidades de quien utilice la definición XML por transformar.

XSLT, contrario a CSS, tiene la capacidad de generar nuevos documentos a partir del documento XML original. Por este motivo, lo encontramos particularmente conveniente para LDL, puesto que tiene el potencial necesario para seleccionar partes de cualquier artículo lexicográfico escrito según nuestra definición con el fin de modificarlas o publicarlas.

2. ¿Qué es el Lenguaje de transformaciones extensible de etiquetado de hojas de estilo (XSLT)?

XSLT es una herramienta de publicación de contenidos estructurados dentro del formato del *Lenguaje de etiquetado extensible* (XML). Este, a su vez, representa contenidos independientemente de la aplicación que puedan tener,

* Investigador del INIL, ELEXHICÓS, Universidad de Costa Rica. Doctorando en la Universidad de Ginebra, Suiza.

apoyándose en varios principios simples y generales para la representación y el intercambio de información. El XML permite:

- Designar partes del contenido marcándolas con etiquetas (nombres de elementos).
- Estructurar este contenido definiendo una jerarquía de elementos (definición del documento como un sistema, cuyas unidades, en consecuencia, responden a cierta organización).

Todo documento XML es representable, de manera abstracta, en la forma de un árbol que simboliza las dependencias del documento. Por tanto, el XML es un lenguaje orientado hacia la descripción de información y no hacia su publicación, para lo cual requiere de hojas de estilo.

El XSLT permite extraer contenidos de un documento XML y verterlos en otros formatos (incluyendo XML). Por ejemplo, es posible escoger los elementos de un documento XML y verterlos en un documento HTML, con la indicación exacta de las características de presentación y de los entornos en los que deben estar¹. Es por esto que se afirma que XML, como marco de trabajo, permite separar el contenido (XML) de su representación (XSLT).

Con el fin de comprender mejor el funcionamiento del XML y el propósito de LDL, recomendamos la lectura del artículo *Propuesta de formalización del manual de redacción del*

Nuevo Diccionario del Español de Costa Rica a partir del Lenguaje de Etiquetado Extensible (XML) (Leoni, 2003), en cuyas páginas se encuentran datos importantes para el mejor entendimiento de lo que sigue en esta presentación. Sin embargo, también es posible consultar los anexos A, B y C de este artículo, en los cuales hemos transcrito una exposición de las partes de una entrada lexicográfica, su representación en XML (nuestra definición de LDL) y una imagen arborescente parcial de sus dependencias, respectivamente; todas han sido tomadas del escrito citado.

2.1 ¿Cómo funciona XSLT?

Un programa XSLT es un documento XML con instrucciones de lectura y procesamiento de otro documento XML, que también incluye un modelo del documento que se creará y en el que deberá ser vertido el contenido.

El programa aprovecha la representación arborescente realizada por un procesador² de XSLT para orientarse en el documento XML como si se tratara de un mapa. Cada sección del documento original en XML (en nuestro caso, en LDL) es identificada por el nombre de la etiqueta (el elemento) y sus atributos, siguiendo esquemas configuracionales descritos en el *Lenguaje de caminos* (XPath). Una vez identificada y extraída la información, esta es vertida en el modelo y así se crea el documento meta.

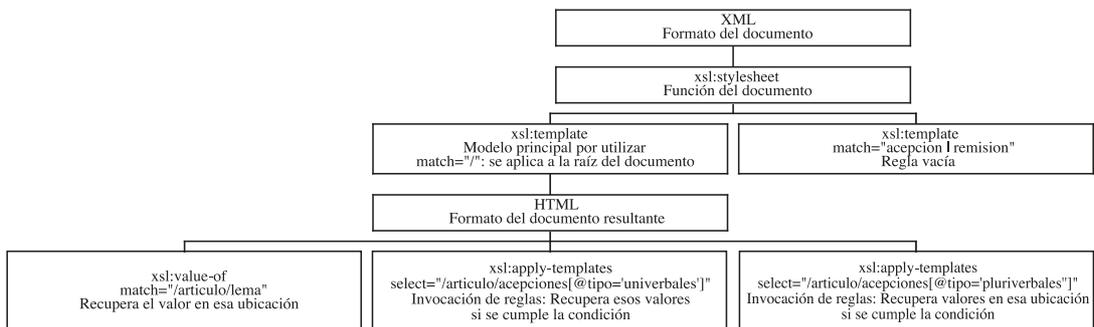


Ilustración 1

3. Reglas de transformación en XSLT para LDL

Para poder transformar un documento, primero es necesario conocer su estructura con el fin de poder describir recorridos por medio de patrones (configuracionales) o de identificar los nombres de elementos; luego es posible aplicar las reglas que crean un documento

con una estructura diferente, a partir de los datos extraídos de la fuente original. En la Ilustración 1 tenemos una muestra de tales patrones (nivel inferior) y de modelo en el cual se vierte la información (indicada como HTML).

En la Tabla 1, está el programa XSLT al cual corresponde la representación parcial de la Ilustración 1:

TABLA 1
Transformación de un artículo codificado en LDL a un documento HTML

```

<?xml version="1.0" encoding="UTF-8" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/">
<html>
<head>
<title>
<xsl:text>Artículo: </xsl:text>
<xsl:value-of select="/articulo/lema" />
</title>
</head>
<body>
<p>
<b>
<xsl:value-of select="/articulo/lema" />
</b>
<xsl:if test="/articulo/acepciones[@tipo='univerbales']">
<xsl:apply-templates select="/articulo/acepciones[@tipo='univerbales']" />
</xsl:if>
<xsl:if test="/articulo/acepciones[@tipo='pluriverbales']">
<xsl:text> * </xsl:text>
<xsl:apply-templates select="/articulo/acepciones[@tipo='pluriverbales']" />
</xsl:if>
</p>
</body>
</html>
</xsl:template>
<xsl:template match="acepcion | remision" />
</xsl:stylesheet>

```

Con el fin de brindar una mejor idea de la aplicación de XSLT en la lexicografía, en el Anexo D hay una versión preliminar de una DTD para LDL; en el Anexo E, la segunda hoja de estilo XSLT que utilizamos para producir el documento HTML de la tabla 2, cuyo código fuente está en el Anexo F.

4. Aplicaciones

XML y XSLT tienen una amplia gama de aplicaciones; citamos las siguientes:

- Clasificación y actualización de información sobre obras de arte (películas, pinturas, etcétera).
- Distribución de *memorias de traducción* (TMX).
- Intercambio de información científica.
- Páginas electrónicas.
- Publicación en línea de obras literarias (Biblioteca Virtual Miguel de Cervantes, 1999).
- Publicación de información para ser leída desde teléfonos celulares (tecnología celular GSM).
- Revistas.

Fácilmente se puede adaptar cualquiera de ellas en la edición lexicográfica.

5. Impacto posible en la redacción y publicación de obras lexicográficas

La tecnología XML (y sus derivados, como el XSLT) puede tener numerosas ventajas para la lexicografía; por ejemplo, se puede:

- Compartir y actualizar información lexicográfica en el caso de métodos descentralizados de trabajo o de proyectos de investigación que requieran utilizar datos comunes generados separadamente por cada equipo investigador.
- Publicar, fácilmente, textos parciales lexicográficos (diccionarios de sinónimos, antónimos, tecnicismos, etcétera) a partir de base datos común (o *gran diccionario*).

- Darle mayor énfasis al contenido que al *estilo* del documento, con lo que la productividad de los lexicógrafos aumentaría; otro equipo podría encargarse de los detalles de presentación en caso de considerarse necesario.
- Aumentar la seguridad y la coherencia de los datos. Cada lexicógrafo, al ocuparse únicamente de sus datos, no necesita tocar la información de los colegas y menos aún del documento madre, por lo que disminuye el riesgo de alterar inadvertidamente la consistencia de los datos.

6. Conclusiones

Dentro del trabajo pendientes para nuestra investigación, podemos citar:

- Conveniencia de una *Descripción de tipo de documento* (DTD) o de un *Schema* (esquema).
- Tipo de base de datos por utilizar (de preferencia que sea basada en XML).
- Organización de un protocolo de trabajo.

Continuando con la propuesta de formalización del manual de redacción del Nuevo Diccionario del Español de Costa Rica a partir del *Lenguaje de etiquetado extensible* (XML); Mostramos cómo el *Lenguaje extensible de transformaciones de hojas de estilo* (XSLT) permite:

- Seleccionar y extraer datos de un documento lexicográfico en XML.
- Asignarles un nuevo formato o ubicarlos en un nuevo contexto.

Notas

- 1 Por ejemplo en tablas, entre etiquetas de negritas o de hipervínculos.
- 2 Algunos de los procesadores más conocidos son Xalan, Xerces, libxml2 y libxslt, xmllint y xsltproc. Estos son gratuitos, aunque también hay opciones comerciales con licencias de uso y contratos de atención al cliente. Algunas direcciones útiles son: <http://xml.apache.org> y <http://www.xmlsoft.org>.

- 3 Referencia a lemas ya definidos en el diccionario.
- 4 Como principio metodológico se recomienda que los nombres de las etiquetas se limiten a los caracteres ASCII, de forma que no aparezcan caracteres acentuados. Esto no aplica para todo lo que esté etiquetado.

Estudios de Lexicografía Hispánico-Costarricense (ELEXHICÓS). *Nuevo Diccionario del Español de Costa Rica*. Sin publicar.

Estudios de Lexicografía Hispánico-Costarricense (ELEXHICÓS) 2003. *Plan de desarrollo*. ELEXHICÓS: Universidad de Costa Rica.

Bibliografía

Biblioteca Virtual Miguel de Cervantes. *Biblioteca Virtual Miguel de Cervantes* – [página de internet] julio de 1999 <http://cervantesvirtual.com/> [12 de junio de 2003]

Bray, Tim *et al.* *El lenguaje extensible de marcas (XML) 1.0 - REC-xml-19980210* - [página de internet] febrero de 1998; <http://www.thefactory.com/ta/xmlspec/index.html> [5 de enero de 2002].

Bray, Tim *et al.* *Espacios de nombre en XML* - [página de internet] 14 de enero de 1999; <http://html.conclase.net/w3c/xml-names-es/> [12 de junio de 2003].

Clark, James y Steve de Rose. *Lenguaje de caminos XML (XPath)* - [página de internet] 19 de octubre de 2001; <http://www.sidar.org/recur/desdi/traduc/es/xml/xpath.html> [12 de junio de 2001].

Leoni de León, Jorge Antonio 2003. *Propuesta de formalización del manual de redacción del Nuevo Diccionario del Español de Costa Rica a partir del Lenguaje de Etiquetado Extensible (XML)*. Káñina, Revista de Artes y Letras, Universidad de Costa Rica. Vol. XXVII (2), 2003.

Ray, Erik 2001. *Learning XML*. O'Reilly: California.

Universidad de Augsburg. *Instrucciones para la redacción*. Sin publicar.

Anexo A

El artículo lexicográfico

En la ilustración siguiente tenemos un ejemplo de un artículo lexicográfico del *Nuevo Diccionario del Español de Costa Rica* (todavía en proceso de redacción):

andar *v 1 tr coloq* Llevar puesta una prenda de vestir. | **2 coloq** Llevar consigo algo de lo cual se puede disponer en cualquier momento. * ~**la** En el juego *la* → *anda*, ser en ese momento la persona que desempeña el papel de perseguidor, p.ej. *la anda Pedro* [llevarla]. || **las de** ~ *coloq euf* Cada una de las extremidades inferiores del cuerpo humano [*patriarcas*]. *Obs:* Es eufemismo de **patas**. || **meter las de** ~ **a** *coloq euf* Cometer un error [*embarcarse, irse con todo, jalarse una torta, meter la(s) pata(s), meter los escarpines, pelársela*]. | **b**) *coloq* Quedar una mujer soltera en estado de embarazo [*jalarse torta, meter la(s) pata(s)*]. *Obs:* Es eufemismo de *meter las* → *patas*.

Ilustración 2. Artículo lexicográfico

Las convenciones que hay que tener en cuenta a la hora de leer el artículo son las siguientes:

- El lema debe encabezar el artículo, se escribe en minúsculas y en negrita (**andar**, en este caso).

- Primero se escriben las unidades univerbales, luego las unidades pluriverbales.
- El asterisco (*) separa las unidades univerbales de las pluriverbales.
- La virgulilla (~) o tilde indica el lugar que ocupa el lema en las unidades pluriverbales.
- Las remisiones³ se representan con una flecha (→).
- Las definiciones de las unidades univerbales se separan con una sola línea vertical (|); las de las unidades pluriverbales, con dos líneas verticales (||). Sin embargo, una unidad pluriverbal con varias definiciones se separa con una sola línea vertical (es el caso de “meter las de ~ a) ... l b) ...”).
- Al empezar la definición, inmediatamente después del número de secuencia, se indican las marcas metalingüísticas en cursivas y por medio de abreviaturas.
- La secuencia de definiciones se indica con números arábigos en negrita (**112 ...**).
- Los sinónimos se escriben entre corchetes ([...]).
- Los ejemplos deben estar escritos en letra cursiva y precedidos de las siglas p.ej.
- Al final de la definición se pueden indicar observaciones, en cuyo caso se preceden de la abreviatura *Obs* en letra cursiva.

Anexo B

Representación de un artículo lexicográfico por medio de LDL

Representación, en XML, del artículo lexicográfico del Anexo A⁴:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<articulo>
  <lema>
    andar
  </lema>
  <acepciones tipo="univerbales">
    <acepcion gramaticalidad="verbo" transitividad="transitivo" pronominalidad="no"
    restrccion="coloq">
      <definicion>
        Llevar puesta una prenda de vestir
      </definicion>
      <ejemplo>
        Ella andaba una suéter roja.
      </ejemplo>
    </acepcion>
    <acepcion gramaticalidad="verbo" transitividad="transitivo" pronominalidad="no"
    restrccion="coloq">
      <definicion>
Llevar consigo algo de lo cual se puede disponer en cualquier momento.
      </definicion>
      <ejemplo>
No andaba suficiente plata.
      </ejemplo>
    </acepcion>
  </acepciones>
  <acepciones tipo="pluriverbales">
```

```

<acepcion gramaticalidad="función sustantiva" transitividad="transitivo">
  <sublema>~la</sublema>
  <definicion>
    En el juego
    <remision tipo="variedad de lengua">
      la
      <remitido>
        anda
      </remitido>
    </remision>
    , ser en ese momento la persona que desempeña el papel de perseguidor,
    <ejemplo tipo="en línea">
      la anda Pedro
    </ejemplo>
  </definicion>
  <sinonimia>
    <sinonimo>
      llevarla
    </sinonimo>
  </sinonimia>
</acepcion>
<acepcion gramaticalidad="función sustantiva" restriccion="coloq"
actitud-del-hablante="euf">
  <sublema>
    las de ~
  </sublema>
  <definicion>
    Cada una de las extremidades inferiores del cuerpo humano
  </definicion>
  <sinonimia>
    <sinonimo>
      patriarcas
    </sinonimo>
  </sinonimia>
  <observaciones>
    Es eufemismo de
    <remision tipo="léxico general">
      patas
    </remision>
  </observaciones>
</acepcion>
<acepcion gramaticalidad="función verbal" actitud-del-hablante="euf"
restriccion="coloq">
  <sublema>
    meter las de ~
  </sublema>
  <definicion>
    cometer un error
  </definicion>

```

```
</definicion>
<sinonimia>
  <sinonimo>embarcarse</sinonimo>
  <sinonimo>irse con todo</sinonimo>
  <sinonimo>jalarse una torta</sinonimo>
  <sinonimo>meter la(s) pata(s)</sinonimo>
  <sinonimo>meter los escarpines</sinonimo>
  <sinonimo>pelársela</sinonimo>
</sinonimia>
</acepcion>
<acepcion gramaticalidad="función verbal" restriccion="coloq">
  <sublema>meter las de ~</sublema>
  <definicion>Quedar una mujer soltera en estado de embarazo.</definicion>
  <sinonimia>
    <sinonimo>jalarse torta</sinonimo>
    <sinonimo>meter la(s) pata(s)</sinonimo>
  </sinonimia>
  <observaciones>
    Es eufemismo de
    <remision tipo="léxico general">
      meter las
    <remitido>patas</remitido>
    </remision>
  </observaciones>
</acepcion>
</acepciones>
</articulo>
```

El conjunto de etiquetas de esta representación, más la expresión clara y determinada de sus relaciones, es lo que conforma el *Lenguaje de descripción lexicográfica* (LDL).

Anexo C

Representación arborescente parcial de un artículo lexicográfico

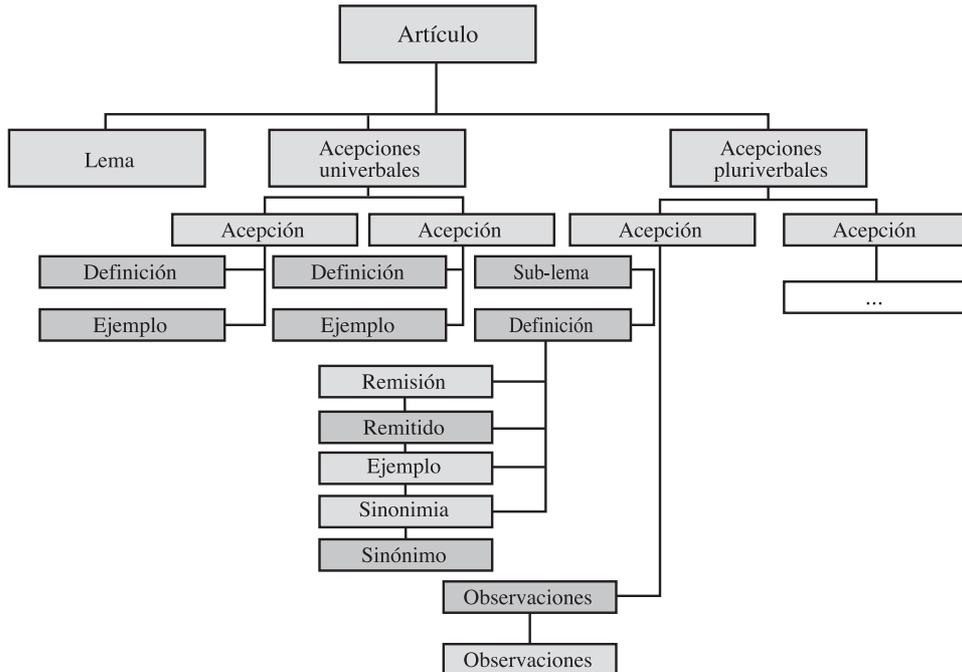


Ilustración 3: Representación gráfica parcial del artículo lexicográfico a partir de su descripción en XML.

Anexo D

Descripción de tipo de documento (DTD) de LDL

```

<!ELEMENT definicion (#PCDATA | remision | ejemplo)*>
<!ELEMENT sinonimo (#PCDATA)>
<!ELEMENT sinonimia (#PCDATA | sinonimo)*>
<!ELEMENT remitido (#PCDATA)>
<!ELEMENT acepciones (#PCDATA | acepcion)*>
<!ATTLIST acepciones tipo CDATA #IMPLIED>
<!ELEMENT lema (#PCDATA)>
<!ELEMENT articulo (#PCDATA | lema | acepciones)*>
<!ELEMENT sublema (#PCDATA)>

```

```

<!ELEMENT ejemplo (#PCDATA)>
<!ATTLIST ejemplo tipo CDATA #IMPLIED>
<!ELEMENT acepcion (#PCDATA | definicion | ejemplo | sublema | sinonimia | observaciones)*>
<!ATTLIST acepcion pronominalidad CDATA #IMPLIED>
<!ATTLIST acepcion transitividad CDATA #IMPLIED>
<!ATTLIST acepcion actitud-del-hablante CDATA #IMPLIED>
<!ATTLIST acepcion gramaticalidad CDATA #IMPLIED>
<!ATTLIST acepcion restriccion CDATA #IMPLIED>
<!ELEMENT observaciones (#PCDATA | remision)*>
<!ELEMENT remision (#PCDATA | remitido)*>
<!ATTLIST remision tipo CDATA #IMPLIED>

```

Anexo E

XSLT: segunda hoja de estilo para transformar LDL

Esta hoja de estilo es, en realidad la primera en ser aplicada. Contiene una invocación al primer programa (articulo-00.xsl) en el cual se encuentra el molde con el código HTML:

```
<?xml version="1.0" encoding="UTF-8" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://
  www.w3.org/1999/XSL/Transform">
<xsl:import href="articulo-00.xsl" />
<xsl:template match="acepcion">
<xsl:if test="sublema">
<b>
<xsl:value-of select="sublema" />
</b>
<xsl:text> </xsl:text>
</xsl:if>
<xsl:if test="sublema">
</xsl:if>
<xsl:if test="@gramaticalidad">
<i>
<xsl:value-of select="@gramaticalidad" />
<xsl:text> </xsl:text>
</i>
</xsl:if>
<xsl:if test="@transitividad">
<i>
<xsl:value-of select="@transitividad" />
<xsl:text> </xsl:text>
</i>
</xsl:if>
<xsl:if test="@pronominalidad">
<i>
<xsl:value-of select="@pronominalidad" />
<xsl:text> </xsl:text>
</i>
</xsl:if>
<xsl:if test="@restriccion">
<i>
<xsl:value-of select="@restriccion" />
<xsl:text> </xsl:text>
</i>
</xsl:if>
<xsl:value-of select="definicion/text()" />
```

```
<xsl:if test="definicion/remision">
<xsl:value-of select="definicion/remision/
  text()" />
<xsl:if test="definicion/remision/remitido">
<xsl:text>→</xsl:text>
<b>
<xsl:value-of select="definicion/remision/remi-
  tido" />
</b>
<xsl:value-of select="definicion/text()
  [position()=2]" />
</xsl:if>
</xsl:if>
<xsl:text> </xsl:text>
<xsl:if test="ejemplo">
<xsl:text>p. ej. </xsl:text>
<i>
<xsl:value-of select="ejemplo" />
</i>
</xsl:if>
<xsl:apply-templates select="sinonimia" />
<xsl:text> </xsl:text>
<xsl:if test="observaciones/remision">
<i>
<xsl:text>Obs: </xsl:text>
</i>
<xsl:value-of select="observaciones/text()" />
<xsl:value-of select="observaciones/remision/
  text()" />
<xsl:if test="observaciones/remision/remiti-
  do">
<xsl:text> </xsl:text>
<xsl:value-of select="observaciones/remision/
  remitido" />
<xsl:value-of select="observaciones/text()[2]"
  />
</xsl:if>
</xsl:if>
</xsl:template>
<xsl:template match="sinonimia">
<xsl:text>[</xsl:text>
<xsl:choose>
<xsl:when test="count(sinonimo) > 1">
<xsl:for-each select="sinonimo">
<xsl:value-of select="." />
<xsl:text>, </xsl:text>
</xsl:for-each>
</xsl:when>
```

```

<xsl:otherwise>                                <xsl:text>] </xsl:text>
<xsl:value-of select="sinonimo" />            </xsl:template>
</xsl:otherwise>                               </xsl:stylesheet>
</xsl:choose>

```

Anexo F

Resultado de las transformaciones

El documento HTML siguiente es el resultado de la aplicación de las dos hojas de estilo en XSLT:

```

<html>
  <head>
    <meta http-equiv="Content-Type content="text/html; charset=UTF-8">
    <title>Artículo: andar </title>
  </head>
  <body>
    <p>
      <b>andar</b>
      <i>verbo</i><i>transisivo </i><i>no</i><i>coloq</i>
      Llevar puesta una prenda de vestir p. ej. <i>Ella andaba una suéter rojas</i>
      <i>verbo</i><i>transisivo </i><i>no</i><i>coloq</i>
      Llevaba consigo algo de lo cual se puede disponer en cualquier momento p. ej. <i>No andaba suficiente plata.</i>
      *
      <b>la</b>
      <i>función sustantiva</i><i>transisivo</i>
      En el juego la →
      <b>anda</b>
      , ser en este momento la persona que desempeña el papel de perseguidor,
      [llevarla]
      <b>las de ~</b>
      <i>función sustantiva</i>
      <i>coloq</i>
      Cada una de las extremidades inferiores del cuerpo humano [patriarcas]
      <i>Obs: </i>
      Es un eufemismo de patas
      <b>meter las de ~</b>
      <i>función verbal</i>
      <i>coloq</i>
      cometer un error [embarcarse, irse con todo, jalarsse una torta, meter la(s) pata(s), meter los escarpines, pelársela]
      <b>meter las de ~</b>
      <i>función verbal</i>
      <i>coloq</i>
      Quedar una mujer soltera en estado de embarazo. [jalarsse torta, meter la(s) pata(s), ]
      <i>Obs: </i> Es eufemismo de meter las → patas.
    <p>
  </body>
</html>

```

