



***ESTANDARIZACIÓN DE DICCIONARIOS
ELECTRÓNICOS: UN ESTUDIO DE CASO***

ANGIE XIOMARA RODRÍGUEZ URIBE

DIRECTORA: ANA FERNÁNDEZ-PAMPILLÓN

GRADO EN LINGÜÍSTICA Y LENGUAS APLICADAS

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FILOLOGÍA

FEBRERO 2017

ÍNDICE DE CONTENIDOS

1. Introducción	2
2. Objetivo del Trabajo.....	4
3. Método de trabajo.....	4
4. Estudio del Diccionario Didáctico Digital de latín	5
5. Revisión de estándares y guías para la construcción de diccionarios electrónicos	7
a. ISO Lexical Markup Framework.....	8
b. LEMON.....	14
c. UNE-EN 62605:2011	19
d. El Text Encoding Initiative (TEI).....	24
e. Síntesis de los estándares revisados.....	28
6. Selección del estándar para el Diccionario Didáctico Digital del Latín	30
7. Aplicación del estándar LMF al DDDL.....	30
7.1 Correspondencia.....	30
7.2 Codificación LMF-XML.....	35
7.3 Resultados y discusión	39
8. Resumen y conclusiones	40
Bibliografía	41
ANEXO 1	43
ANEXO 2.....	50
ANEXO 3.....	52
ANEXO 4.....	53
ANEXO 5.....	54

1. Introducción

Los diccionarios han jugado un papel fundamental dentro del área de aprendizaje, tanto en formato electrónico como en papel, ya que permiten al usuario ampliar su conocimiento sobre una lengua dada. De forma general, un diccionario puede definirse como “Repertorio en forma de libro o en soporte electrónico en el que se recogen, según un orden determinado, las palabras o expresiones de una o más lenguas, o de una materia concreta, acompañadas de su definición, equivalencia o explicación” (Real Academia Española, 2014, 23º ed.). En este trabajo se hará referencia a un tipo de diccionario: los diccionarios didácticos y en formato electrónico. Se considera que un diccionario didáctico es “un diccionario concebido para facilitar el proceso de enseñanza/aprendizaje de lenguas”(Azorín Fernández, 2000). El carácter didáctico, hace que tanto la macro como la microestructura del diccionario estén orientadas a las necesidades de los destinatarios.

Autores como Arnold et al.(1994) manifiestan que diversas teorías de representación lingüística, dan lugar a diferentes estrategias de implementación, provocando que, tanto los diccionarios en formato papel, como en formato electrónico sean diversos en términos de formatos, cobertura, nivel de detalle y formalismo preciso para la descripción léxica. Con el auge de la tecnología, esta idea de diversidad ha prosperado entre las herramientas de aprendizaje, como lo son los diccionarios electrónicos. El hecho de desarrollar estructuras de diseño y de aplicación de forma individual, ha provocado que la información recogida en múltiples diccionarios no sea reutilizable, portable, ni tampoco interoperable. La idea de crear y adoptar estándares se ha reforzado, sobre todo en el área de *e-learning*, donde se busca formalizar todos los recursos educativos existentes, mediante métodos y herramientas de calidad como los estándares. (Álvarez Álvarez, 2003).

El BSI (British Standard Institute, 2017) describe el término estándar como “formas acordadas de hacer algo, escritas como un conjunto de criterios precisos para que puedan usarse como reglas, guías o definiciones”. Por su parte, la Organización Internacional de Normalización (<http://www.iso.org>), define estándar como “un documento que proporciona requisitos, especificaciones, directrices o características, que se pueden utilizar de manera consistente para asegurar que materiales, productos, procesos y servicios sean adecuados para sus propósitos”. Desde el punto de vista tecnológico, los estándares aportan al mundo del aprendizaje reutilización¹,

¹ El contenido puede usarse y encajar en un contexto diferente al del origen.

interoperabilidad² y portabilidad³(Masie, 2003), y, en consecuencia, mejoran la durabilidad de la información.

En este trabajo se aborda la cuestión de cómo estandarizar la información de un diccionario didáctico electrónico. El por qué estandarizarla es una cuestión ya respondida, puesto que es la razón de ser de los estándares. Sin embargo, el cómo aplicar un estándar no es una cuestión trivial: (i) en primer lugar se necesita buscar y seleccionar qué estándar aplicar, (ii) en segundo lugar decidir cómo aplicarlo y, finalmente, (iii) documentar los resultados, proporcionando una descripción clara y precisa de la macro y microestructura del diccionario conforme los elementos del estándar. Resolver cómo estandarizar un diccionario es el primer paso en un proyecto de construcción de un diccionario y es clave puesto que guiará las fases posteriores de diseño e implementación del diccionario como sistema informático. Asimismo, será imprescindible para el mantenimiento durante la vida del diccionario y para la posible integración o reutilización del diccionario en otros sistemas informáticos (e.g. sistemas de aprendizaje electrónico).

En este Trabajo de Fin de Grado se aborda esta cuestión de forma empírica y cualitativa utilizando un estudio de caso real: el proyecto de construcción de un diccionario didáctico digital para el Latín y el Alemán y su integración en un sistema de aprendizaje de lenguas en línea (Fernández-Pampillón Cesteros, 2016). Se busca dar respuesta desde un punto de vista lingüístico a la descripción de la información léxica del diccionario⁴ para su codificación y la creación de herramientas que extraigan esta información.

El trabajo se estructura en siete secciones. En la primera, se definen los objetivos que se quieren alcanzar con este trabajo. En la segunda, se explica el procedimiento que se ha seguido. En la tercera parte, se analizan las características del estudio de caso, con el propósito de escoger un estándar que se adecue a la información que contiene el diccionario. En la cuarta parte, se hace una revisión y síntesis de cuatro estándares y guías para la construcción de diccionarios electrónicos. Posteriormente en la quinta parte, se escoge un estándar para aplicarlo al diccionario, explicando los motivos de esta elección. En la sexta parte, se aplica el estándar al diccionario y se discuten los

² El contenido de diferentes recursos puede mezclarse y combinarse. Comunicación, intercambio e interacción entre múltiples sistemas.

³ El contenido es independiente de la herramienta o plataforma. Posibilidad de acceder a un mismo contenido a través de plataformas diferentes.

⁴ Inicialmente sobre el diccionario didáctico digital para el Latín, pero en un futuro, sobre diccionarios de lenguas flexionadas, como el alemán.

resultados obtenidos. Por último, se realiza el resumen del trabajo y se presentan las conclusiones alcanzadas, las cuales servirán de apoyo para trabajos futuros.

2. Objetivo del Trabajo

El objetivo principal del trabajo es responder a la cuestión de cómo estandarizar la información de un tipo de diccionario electrónico, en este caso didáctico, con una estructura previamente definida de forma que su contenido sea portable, reutilizable e interoperable. Asimismo, se pretende identificar los posibles estándares para la construcción de diccionarios electrónicos, describir las ventajas y desventajas de cada estándar para el caso de estudio y obtener recomendaciones (pautas) provisionales sobre cómo podría organizarse un nuevo proyecto de estandarización de un diccionario. Esta cuestión tiene interés en primer lugar porque asegura que el diccionario no quedará obsoleto tecnológicamente o su uso quede limitado por el tipo de dispositivo de acceso. En segundo lugar, tiene interés por las posibilidades que ofrece para poder re-utilizar su contenido como fuente de información para otras aplicaciones software, por ejemplo, aplicaciones de aprendizaje en línea de lenguas como Babel (Albusac Jiménez, González Morcillo, & Jiménez Linares, 2007).

El diccionario sobre el que se va estudiar la cuestión de investigación es un Diccionario Didáctico Digital de Latín (en adelante DDDL). Este diccionario pretende suplir las carencias encontradas en los diccionarios de latín creados hasta ahora, es decir, “las descripciones del comportamiento gramatical incompletas, la falta de sistematización en la descripción de los datos de la microestructura de los lemas, y la falta de determinación de las relaciones entre los datos morfológicos o sintácticos con los semánticos” (Márquez & Chaves, 2016). Está dirigido a personas que se enfrentan al aprendizaje del latín por primera vez. Se pretende que el usuario adquiera un conocimiento básico de la sintaxis latina utilizando el diccionario propuesto (...) para aclarar no sólo los significados de las principales unidades de habla sino también las relaciones que estas unidades tienen entre ellas, especialmente entre verbos y sustantivos. (Márquez & Chaves, 2016).

3. Método de trabajo

El método seguido para responder esta cuestión está basado en un estudio de caso (Wiersma & Jurs, 2005, pp. 210-211). Este estudio se realizó a mediados de septiembre y diciembre del 2016. Se ha seguido el siguiente procedimiento: primero, se realizó un análisis de la estructura y los requisitos del diccionario didáctico digital utilizado como caso de estudio; segundo, se hizo una revisión y análisis de los estándares y guías para la construcción de diccionarios electrónicos, con el objetivo de realizar una síntesis que sirviera de guía para seleccionar el estándar o estándares más adecuado al diccionario y a su posible integración en otros sistemas software; en tercer lugar,

se seleccionó el estándar; y por último, se aplicó dicho estándar al diccionario, proporcionando resultados de los cuales se han extraído conclusiones.

4. Estudio del Diccionario Didáctico Digital de latín

Este diccionario forma parte del proyecto nº 269 de Innovación y Mejora de la Calidad Docente de la Universidad Complutense de Madrid de la rama de conocimiento de Artes y Humanidades del año 2016/2017. El objetivo principal del proyecto es “la creación de una versión digital en abierto de un diccionario basado en un modelo lexicográfico y cognitivo para mejorar el aprendizaje de la lengua latina y, el estudio preliminar de su generalización a otras lenguas”(Fernández-Pampillón Cesteros, 2016).

El diccionario sigue por una parte una perspectiva onomasiológica, es decir, parte del significado (conceptos) para llegar al significante (palabras), puesto que se representa con una imagen el concepto que recoge el lema. Por otra parte, una vez que se accede al lema para conocer el significado, se sigue una perspectiva semasiológica⁵. La macroestructura del diccionario se basa en una ordenación sistemática, permitiendo que el usuario pueda acceder a las entradas de distintas maneras: de forma didáctica, en la que las entradas se agrupan en función de la descripción de la valencia cuantitativa de los verbos y la caracterización léxica de los sustantivos; clasificada, siguiendo un sistema de clasificación lexicográfico: letra, categoría gramatical, lema, número de argumentos y ejemplo; secuencial, siguiendo el orden de los identificadores de las entradas; y a través del buscador. El diccionario contiene 533 entradas, en él se encuentran recogidos sustantivos, adjetivos, adverbios, conjunciones, preposiciones, pronombres y verbos más representativos y más usados en un nivel básico de latín.

Para usar este diccionario es necesario que previamente el usuario conozca algunos conceptos básicos de la lengua latina, por ejemplo, la morfología nominal y verbal, las funciones sintácticas y el orden de las palabras. Con este diccionario el usuario aprende a diseñar un mapa conceptual sintagmático y paradigmático que le ayude a entender el uso de unidades léxicas (principalmente unidades verbales y nominales, pero no exclusivamente), al relacionar los argumentos y los satélites verbales con los significados de las unidades. El aporte de este diccionario a la enseñanza es que su estructura de lemas principalmente verbales y nominales, puede ser

⁵ Partiendo de la palabra, es decir, el lema, para llegar a su significado.

beneficiosa a la hora de utilizar aplicaciones tecnológicas que fomenten el desarrollo cognitivo del usuario, facilitando un mayor y mejor conocimiento del idioma latino (Márquez & Chaves, 2016).

La microestructura de este diccionario, de forma novedosa, obedece básicamente a aspectos semánticos y morfológicos organizados entorno al verbo (marco predicativo⁶) según los criterios de la Gramática Funcional (Dik, 1997) y la Gramática Valencial o de dependencias (Tesnière, 1994). Es por esto que, se describe los marcos predicativos de los verbos y la caracterización léxica de sus argumentos.

Cada entrada está organizada mediante cuatro campos generales: identificador, lema, letra y categoría, siendo el más importante lema. Dependiendo de la categoría gramatical la estructura es la siguiente:

- En el caso de los verbos, se muestra primero la traducción equivalente en español, seguido de la descripción de la valencia cuantitativa: monovalente (el verbo tiene un argumento), bivalente (el verbo tiene dos argumentos) y trivalente (el verbo tiene tres argumentos) (Figura 1). A continuación, la caracterización morfológica de los argumentos verbales se presenta en función de los casos (nominativo, acusativo, dativo, ablativo) o del régimen proposicional (ab + ablativo, ex + ablativo). Después, se detalla la caracterización léxica de los argumentos mediante su descripción ontológica.

⁶ Un marco predicativo es el número de constituyentes obligatorios (complementos) del predicado (argumentos) y la función semántica de cada uno de ellos. La función semántica puede ser: agente, paciente, receptor, destinatario, dirección, lugar y cero.

Objeto Digital 646 ◀ Anterior / Siguiente ▶▶

Datos Recursos Metadatos

Identificador: 646

LEMA:
peto, is, ere, ii (ivi), itum (3ª)

LETRA: *p*

CATEGORÍA: *Verbo*

Significado 1: *Pedir algo a alguien*

Numero de Argumentos (Verbos): *TRIVALENTE*

Primer argumento : *Nominativo*
Caracterización Argumental 1: *+animado +humano*

Segundo argumento: *Acusativo*
Caracterización Argumental 1: *-animado +definido*
Caracterización Argumental 2: *-animado -definido*

Tercer argumento: *A /Ab+ABL /Ex+ABL*
Caracterización Argumental 1: *+animado +humano*

Ejemplo:
Poeta librum ex bibliothecario petit

Significado 2: *atacar*

Numero de Argumentos (Verbos): *BIVALENTE*

Primer argumento : *Nominativo*
Caracterización Argumental 1: *+animado +humano*

Segundo argumento : *Acusativo*
Caracterización Argumental 1: *+animado +humano*

Ejemplo:
Captivi romani Belgas petunt

Figura 1. Entrada de un verbo del Diccionario Didáctico Digital de Latín con dos marcos predicativos, uno trivalente y otro bivalente (Fuente: Diccionario Didáctico Digital del Latín http://repositorios.fdi.ucm.es/DiccionarioDidacticoLatín/view/paginas/view_paginas.php?id=1)

- En cuanto a los sustantivos, dentro del lema se especifica el género del sustantivo de forma abreviada (m.(masculino), f.(femenino), n.(neutro)). Lo siguiente que se muestra es su significado, seguido de la correspondiente caracterización léxica, empleando “la clasificación referencial la cual utiliza rasgos semánticos como [+/-concreto +/-humano +/- animado]. Estos rasgos se adjuntan a las unidades léxicas para describir su pertenencia a las clases semánticas” (Gross,1994).
- Por último, en los adjetivos, adverbios, preposiciones, pronombres y conjunciones sólo se muestra su significado.

5. Revisión de estándares y guías para la construcción de diccionarios electrónicos

A continuación, se hará una revisión y explicación de los estándares más adecuados para la codificación de un diccionario didáctico electrónico. Los estándares tratados son cuatro: ISO

24613:2008 Lexical Markup Framework (LMF)⁷, LEMON - The Lexicon Model for Ontologies⁸, la norma UNE-EN 62605:2011⁹ y el TEI¹⁰. Es preciso aclarar, que sólo la norma UNE-EN 62605:2011 está enfocada exclusivamente a diccionarios electrónicos como tal, mientras que el Lexical Markup Framework (LMF) y LEMON están orientados a recursos electrónicos léxicos en general, incluyendo los diccionarios electrónicos; y el TEI destinado a la representación de textos digitales en general e incluye un módulo para la descripción de diccionarios.

a. ISO Lexical Markup Framework

El Marco de Etiquetado Léxico (en adelante LMF) es un estándar de la Organización Internacional para la Estandarización (ISO) (<http://www.iso.org>). Su identificador completo es ISO 24613:2008 *Language resource management - Lexical Markup Framework* (LMF) y se define como:

Un metamodelo abstracto que proporciona un marco común y estandarizado para la construcción de lexicones computacionales. El LMF asegura la codificación de la información lingüística de una manera que permite la reutilización en diferentes aplicaciones y para diferentes tareas. Proporciona una representación común y compartida de objetos léxicos, incluyendo aspectos morfológicos, sintácticos y semánticos. (ISO 24613:2008).

Su objetivo principal es establecer un modelo común para la creación y uso de recursos léxicos electrónicos sin importar el tamaño, para que sea posible la gestión e intercambio de datos entre estos recursos, así como la unión de diferentes recursos electrónicos individuales, para formar extensos lexicones mundiales (ISO 24613:2008). El LMF utiliza Unicode («What is Unicode?», 2017) para la escritura y ortografía de las entradas léxicas, asimismo, emplea el ISO 12620 *Data Category Registry* (DCR) (Ide & Romary, 2004) para definir los rasgos que marcan la información léxica en forma de pares atributo-valor, como por ejemplo el género, el caso, información semántica, etc. Los modelos LMF son representados mediante clases UML (asociaciones entre clases) («Lenguaje unificado de modelado», 2017). La principal ventaja del LMF es su estructura modular interoperable formada por módulos que extienden un modelo base, permitiendo así que

⁷ <http://www.lexicalmarkupframework.org/>

⁸ <http://lemon-model.net/>

⁹ <http://www.aenor.es/aenor/normas/normas/fichanorma.asp?tipo=N&codigo=N0048269&PDF=Si#.WDn7vVzWErk>

¹⁰ <http://www.tei-c.org/index.xml>

la información lingüística pueda ir enriqueciéndose de forma incremental e independiente. Además, al ser un estándar facilita la portabilidad e interoperabilidad entre dispositivos de acceso y sistemas informáticos.

De forma más detallada, la estructura del LMF se divide en dos partes: la estructura básica (*core package*) y las extensiones (*extensions*) (Figuras 2 y 3 respectivamente).

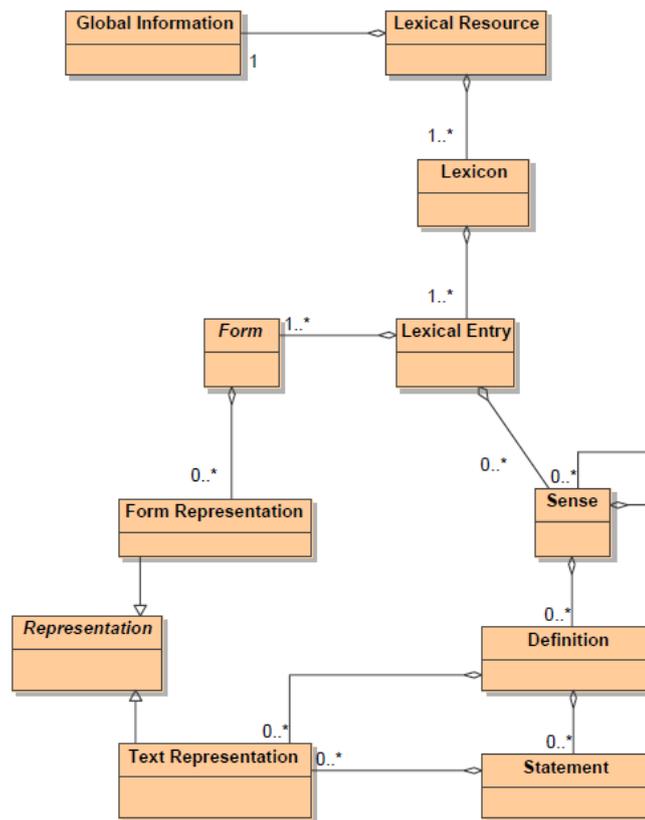


Figura 2. Estructura base del LMF (Fuente: ISO 24613:2008)

La estructura básica consta de once clases (o elementos estructurales):

1. Clase Recurso Léxico (*Lexical Resource class*): representa todo el recurso y consta de una o más instancias de la Clase Lexicon (*Lexicon class*).
2. Clase Información Global (*Global Información class*): contiene la metainformación del Recurso Léxico como código de las lenguas de los lexicones.
3. Clase Lexicon (*Lexicon class*): contiene todas las entradas léxicas de una lengua.
4. Clase Entrada Léxica (*Lexical Entry class*): contiene la información léxica de cada lexema. Dentro de esta clase se encuentra la Clase Forma (*Form*) y la Clase Significado (*Sense*). Por lo tanto, esta clase gestiona la relación entre formas y sus significados relacionados.

5. Clase Forma (*Form class*): representa un lexema, una variante morfológica de un lexema o un morfo. Agrupa una o más variantes ortográficas de la forma abstracta, así como el lema, pronunciación, etc. Contiene las subclases Clase Representación de la Forma (*Form Representation class*) y Clase Representación (*Representation class*).
6. Clase Significado (*Sense class*): representa un significado de una entrada léxica. Permite una organización jerárquica de significados dando a entender que un significado puede ser más específico que otro dentro de la misma entrada léxica.
7. Clase Definición (*Definition class*): representa una descripción narrativa de un significado. Una instancia de la Clase Significado puede tener cero o muchas definiciones. Cada instancia de la Clase Definición puede estar asociada con cero a muchas instancias del tipo Representación del Texto (*Text Representation*) que almacena texto Unicode.
8. Clase Declaración (*Statement class*): representa una descripción narrativa y redefine o complementa la Clase Definición. Una instancia de la Clase Definición puede tener de cero a muchas instancias de la Clase Declaración.

Las extensiones (*extensions*) del modelo LMF son subconjuntos de las clases de la estructura base. No pueden usarse para representar datos léxicos independientemente de la estructura base. Además, dependiendo del tipo de datos lingüísticos involucrados, una extensión puede depender de otra extensión. Las extensiones abarcan los ámbitos lingüísticos de la semántica, la morfología y la sintaxis, al igual que aplicaciones para el Procesamiento del Lenguaje Natural (PLN). Las extensiones previstas por el LFM son: extensión de morfología, extensión de semántica de PLN, extensión de sintaxis de PLN, extensión de diccionario legible por máquina (*Machine Readable Dictionary*), extensión de notación multilingüe de PLN, extensión de patrones morfológicos de PLN, extensión de expresión de restricción (ISO 24613:2008). En este trabajo se explican, por razones de espacio, sólo las dos primeras extensiones que son las relevantes para describir el DDDL.

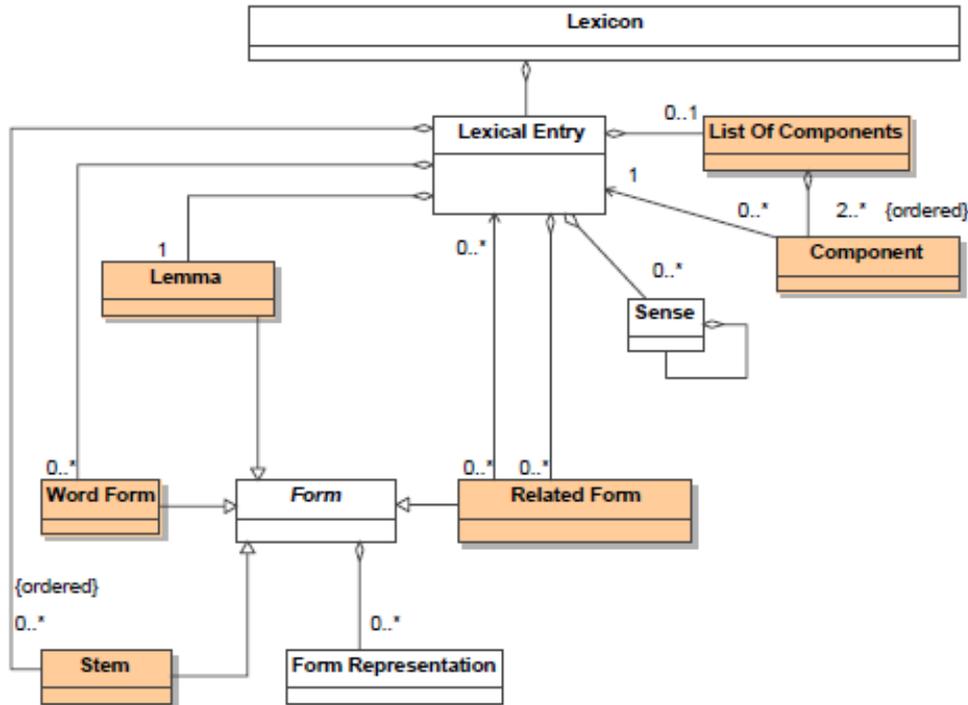


Figura 3. Modelo de la extensión de morfología (Fuente: ISO 24613:2008)

La extensión de morfología permite una descripción extensiva de la morfología de las entradas léxicas. El modelo de esta extensión además de tener las clases de la estructura base, añade un conjunto de subclases que, como se puede apreciar en la Figura 3, proceden de la Clase Forma. Estas subclases son las siguientes:

- Clase Lema (*Lemma class*): representa la forma de la palabra escogida para designar la entrada léxica.
- Clase Forma de Palabra (*Word Form class*): representa una forma que un lexema puede tomar cuando es usado en una frase o locución. Esta clase puede gestionar lexemas simples, compuestos o expresiones multipalabra.
- Clase Raíz (*Stem class*): representa un morfo. La asociación agregada entre una entrada léxica y una raíz es ordenada. De modo que la raíz organiza las partes de los sublexemas.
- Clase Forma Relacionada (*Related Form class*): representa la forma de una palabra o morfo que puede estar relacionado a la entrada léxica en una de las formas de variedad.

Según (Francopoulo et al., 2006) hay dos posibles estrategias para describir la morfología de una palabra con LMF. La primera es representar explícitamente todas las formas flexionadas como se presenta en la Figura 4:

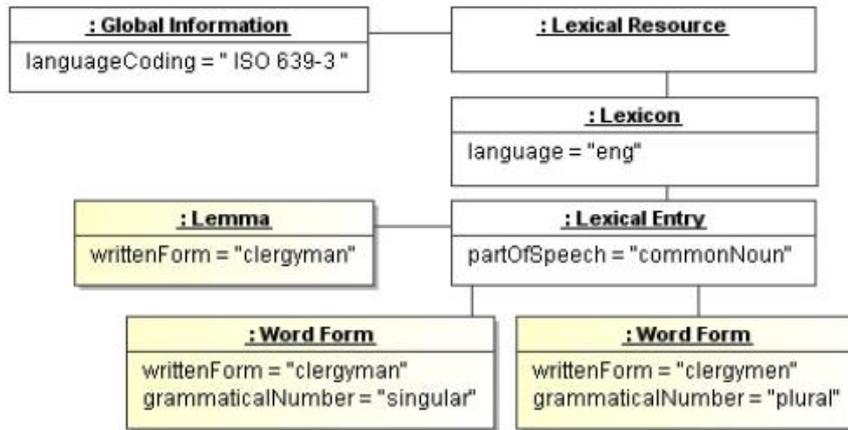


Figura 4. Formas flexionadas de una palabra (Fuente: ISO 24613:2008)

La segunda estrategia es usar un Paradigma Flexional (*Inflectional Paradigm*), el cual puede aplicarse, a su vez, de dos formas, la más simple es declarar que la flexión de la palabra es igual que otra palabra, en la Figura 5 se establece que la palabra *clergyman*(clérigo) se flexiona igual que la palabra *man*, empleando el Paradigma Flexional “asMan”:



Figura 5. Flexión simplificada de la palabra clergyman (Fuente: Francopoulo et al., 2006)

La segunda forma del Paradigma Flexional es mucho más elaborada ya que utiliza la extensión del paradigma flexional (Figura 6). En lenguajes cuya morfología sea simple no merece la pena utilizar esta segunda forma, pero en lenguajes como alemán o húngaro sería de gran utilidad para expresar la flexión, y su estructura sería así:

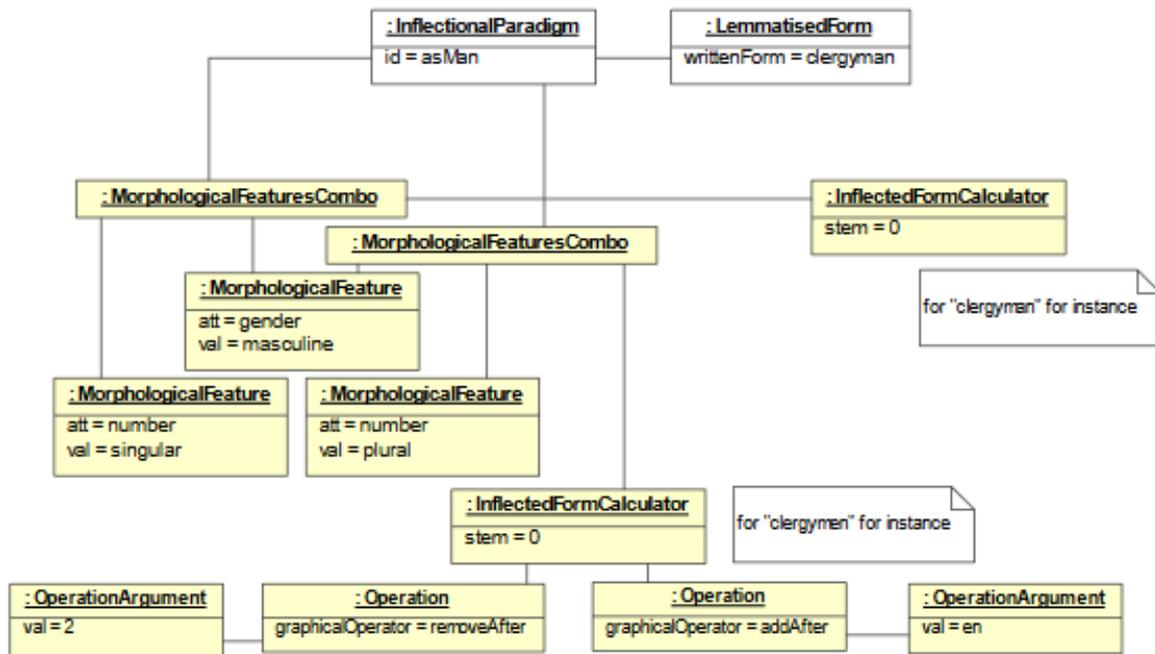


Figura 6. Flexión de la palabra *clergyman* usando la extensión del Paradigma Flexional (Fuente: Francopoulo et al., 2006)

En la Figura 6 se puede observar que la flexión de la palabra inglesa *clergyman* se realiza mediante dos operaciones, las cuales actúan sobre los dos argumentos/letras: la primera operación consiste en quitar los argumentos “an” (*removeAfter*), mientras que la segunda operación añade los argumentos “en” (*addAfter*) que forman el plural de *man*.

Por su parte, la extensión de semántica de PLN del LMF describe un significado y su relación con otros significados que pertenecen al mismo lenguaje. Las clases que forman esta extensión son (Figura 7):

- Clase *Synset* (*Synset class*), que representa el conjunto de significados compartidos dentro del mismo lenguaje, una instancia de la Clase *Synset* puede enlazar significados de diferentes instancias de Entradas Léxicas con la misma categoría gramatical;
- Clase Relación de *Synset* (*Synset Relation class*), representa la relación orientada entre las instancias de la Clase *Synset*;
- Clase Relación Semántica (*Sense Relation class*), representa la relación orientada entre las instancias de la Clase Significado;
- Clase Ejemplo del Significado (*Sense Example class*), se usa para explicar el significado particular de una instancia de la Clase Significado;
- Clase Predicado Semántico (*Semantic Predicate class*), representa un significado abstracto junto con su asociación con la Clase Argumento Semántico;

- Clase de Representación Predicativa (*Predicative Representation Class*), representa el enlace entre las clases Significado y Predicado Semántico;
- Clase Argumento Semántico (*Semantic Argument class*), representa un argumento de un predicado semántico dado;
- Clase Relación de Argumento (*Argument Relation class*), representa una relación orientada entre las instancias de la Clase Argumento Semántico que pertenecen a la misma instancia de predicado;
- Clase SynSemArgMap (*SynSemArgMap class*), representa los enlaces entre un argumento semántico y un argumento sintáctico;
- Clase SynSemCorrespondence (*SynSemCorrespondence class*), representa un conjunto de instancias de la Clase SynSemArgMap para una instancia dada de la Clase Marco de Subcategorización;
- Clase Relación de Predicado (*Predicate Relation class*), representa la relación orientada entre las instancias de la Clase Predicado Semántico;
- Clase de Referencia Externa Monolingüe (*Monolingual External Reference class*), representa la relación entre una instancia de la Clase Significado o Synset y un sistema externo.

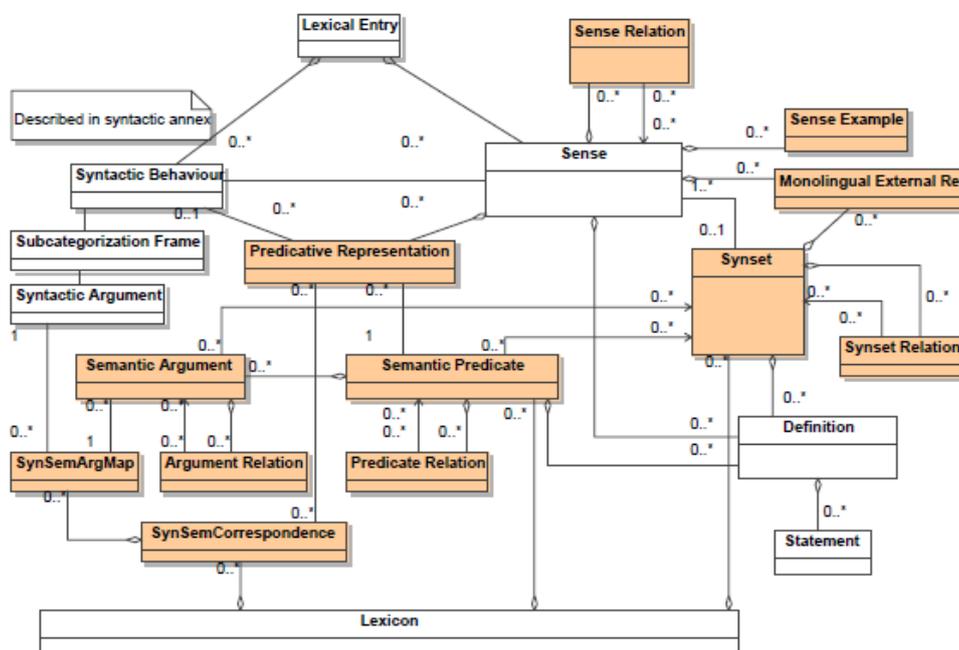


Figura 7. Extensión de semántica de PLN (Fuente: ISO 24613:2008)

b. LEMON

LEMON se define como un modelo, desarrollado en el proyecto Monnet, con el que se puede compartir información léxica en la web semántica. Se basa en RDF, con lo que permite una

mejor representación de los enlaces entre las diferentes secciones del léxico. Las principales ventajas de este estándar son su precisión, ya que sólo utiliza las clases y definiciones que sean necesarias; es descriptivo no prescriptivo, utiliza recursos externos para las definiciones, de modo tal que puede extenderse según su propósito; y es modular, se divide en un número de módulos, no es necesario aplicar el modelo entero para crear un léxico funcional (McCrae et al., 2008).

La estructura principal de este estándar (Figura 8) está compuesta de los siguientes elementos:

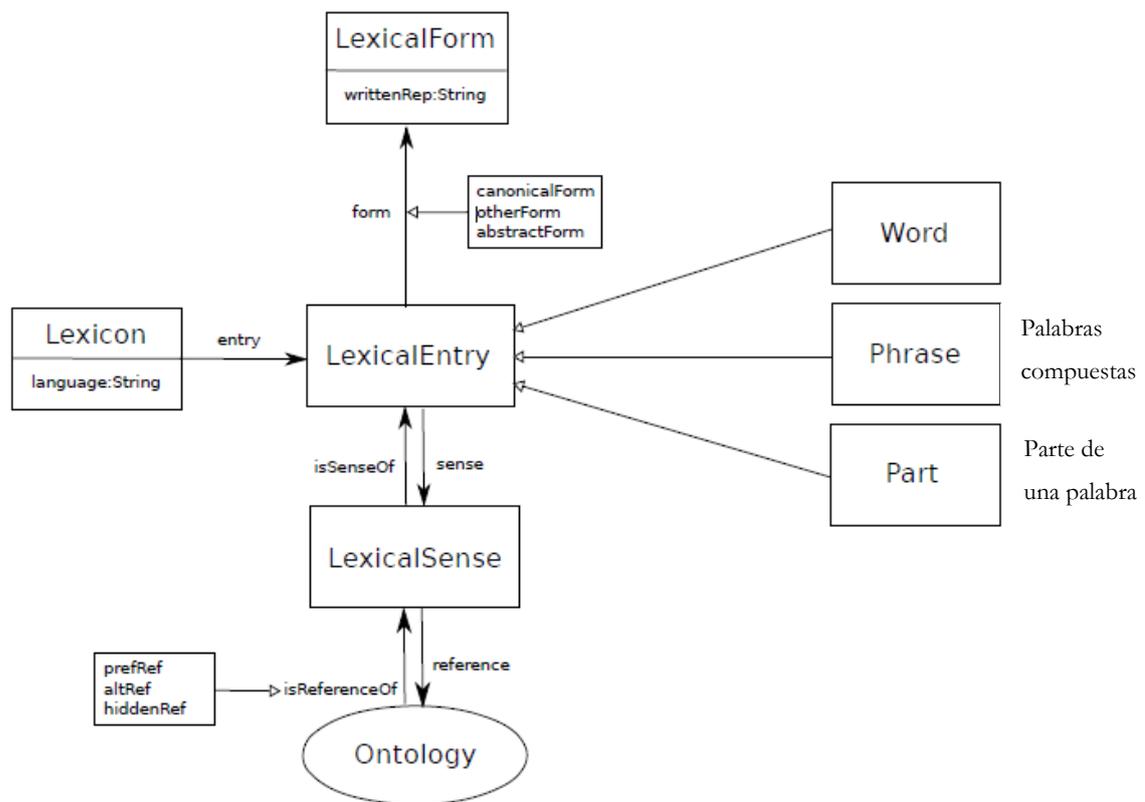


Figura 8. Estructura base del estándar LEMON (Fuente: McCrae et al., 2008).

El elemento *Lexicón* (*Lexicon*) contiene todos los elementos en un léxico. Cada léxico es marcado con un lenguaje en forma de una etiqueta del ISO639. El elemento *Entrada Léxica* (*LexicalEntry*) representa una entrada léxica dada. El elemento *Significado Léxico* (*LexicalSense*), representa la relación entre la entrada léxica y la entrada ontológica. El elemento *Forma Léxica* (*LexicalForm*) representa una realización superficial de una entrada léxica dada, típicamente una representación escrita. Además, LEMON permite que las variantes sintácticas se diferencien de las formas preferidas por sub-propiedades de la forma, a través de las formas canónicas (*canonical forms*) y lexicalizaciones preferidas. Cada entrada léxica no está desambiguada semánticamente y la referencia proporciona la semántica del término.

- La Forma Canónica (*canonicalForm*): es la forma estándar (como se escribe en el diccionario) de la entrada.
- Otra Forma (*otherForm*): variantes morfológicas de las palabras, por ejemplo, la -s del plural.
- Forma Abstracta (*abstractForm*): palabra raíz(*stem*) u otro morfema o forma no realizable.

Mediante cinco módulos (Figura9) la estructura principal puede ampliarse dependiendo del objetivo.

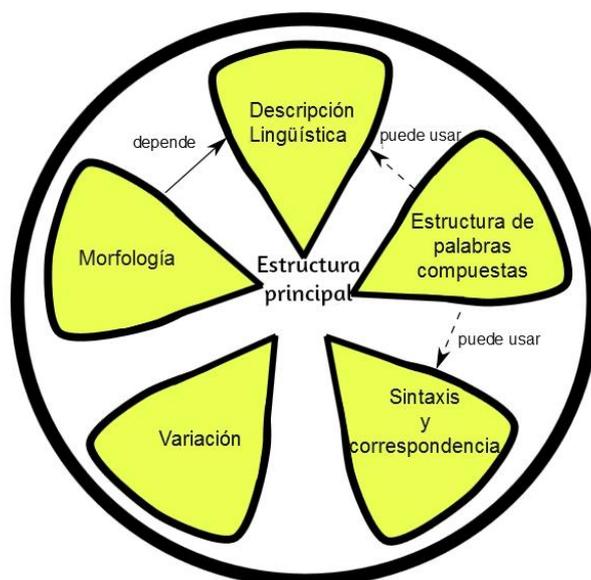


Figura 9. Módulos LEMON (Fuente: McCrae et al., 2008).

El módulo de Descripción Lingüística (Figura10) se usa para describir propiedades lingüísticas, comúnmente encontradas en los diccionarios, las cuales pueden añadirse a elementos del lexicon. Algunas propiedades pueden usarse desde recursos como ISOcat («ISOcat - a Data Category Registry», 2017). También puede describirse propiedades lingüísticas de las formas: como la descripción fonética y contextos.

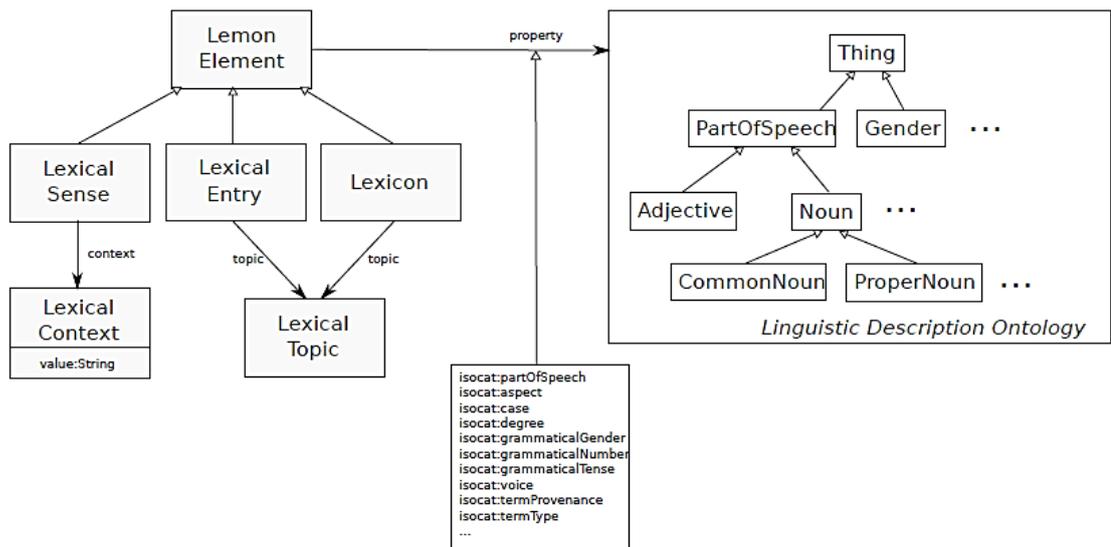


Figura 10. Módulo Descripción Lingüística (Fuente: McCrae et al., 2008).

El módulo de variación (Figura 11) se usa para describir relaciones entre objetos en lemon:lexicon. Hay tres relaciones: Relación de Significado (*senseRelation*), Variante Léxica (*lexicalVariant*) y Variante Forma (*formVariant*), las cuales describen relaciones entre acepciones, entradas y formas.

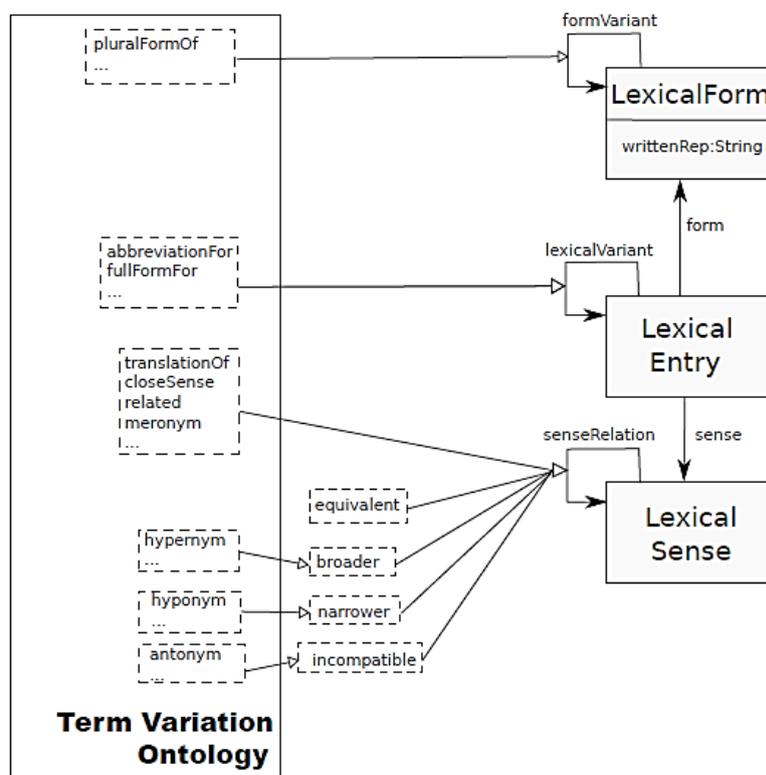


Figura 11. Módulo de variación (Fuente: McCrae et al., 2008).

Por otra parte, mediante el módulo Estructura de palabras compuestas (Figura 12) se puede abordar la descomposición de los términos, la estructura de las locuciones, las relaciones de dependencia y partes del sintagma nominal en referencia a la entrada léxica.

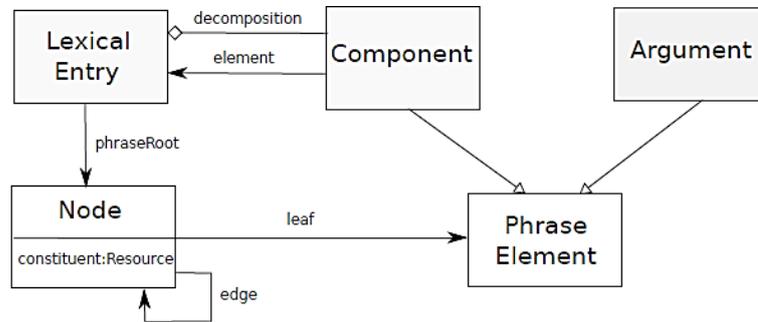


Figura 12. Módulo Estructura de palabras compuestas (Fuente: McCrae et al., 2008).

Con el módulo sintaxis y correspondencia (Figura 13), LEMON construye la sintaxis mediante marcos predicativos con la propiedad `lemon:synBehaviour` y cada argumento con `lemon:synArg`.

Mediante este marco se puede restringir la asignación entre los niveles sintácticos y semánticos. Por ejemplo, en alemán los verbos *essen* y *fressen* significan comer, pero se diferencian en que el primero se utiliza para personas y el segundo para animales. Con LEMON puede asignarse esta distinción de esta forma:

```

:essen lemon:sense [ lemon:reference ontology:eat ;
  lemon:propertyDomain ontology:Human ] .
:fressen lemon:sense [ lemon:reference ontology:eat ;
  lemon:propertyDomain ontology:NonHuman ] .

```

La asignación de adjetivos sólo puede hacerse de forma unitaria, por ejemplo, rojo, grande, y no de forma binaria como parecido a o los superlativos y comparativos. Con LEMON las formas binarias de los adjetivos deben estar especificadas en la ontología.

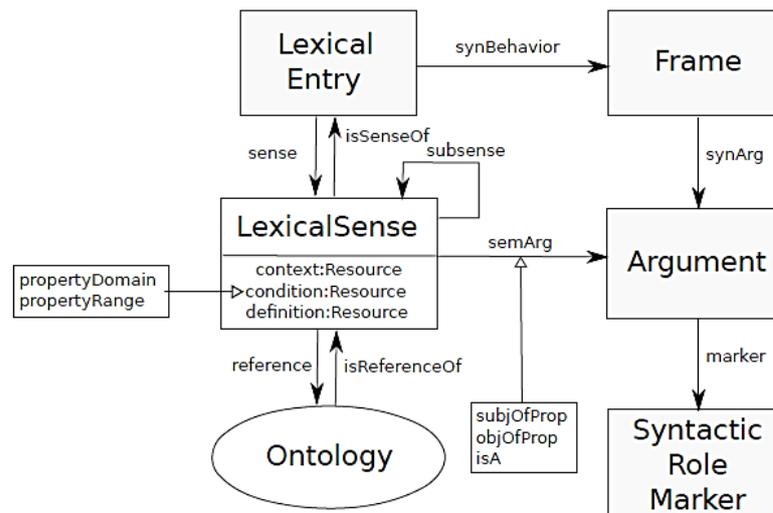


Figura 13. Módulo sintaxis y correspondencia (Fuente: McCrae et al., 2008).

El módulo de morfología (Figura 14) permite representar la flexión a través de un patrón, el cual está compuesto de un conjunto de transformaciones las cuales se realizan mediante reglas y producen una forma denominada prototipo. Aunque también puede representarse la aglutinación donde la forma de una palabra es dependiente de palabras vecinas.

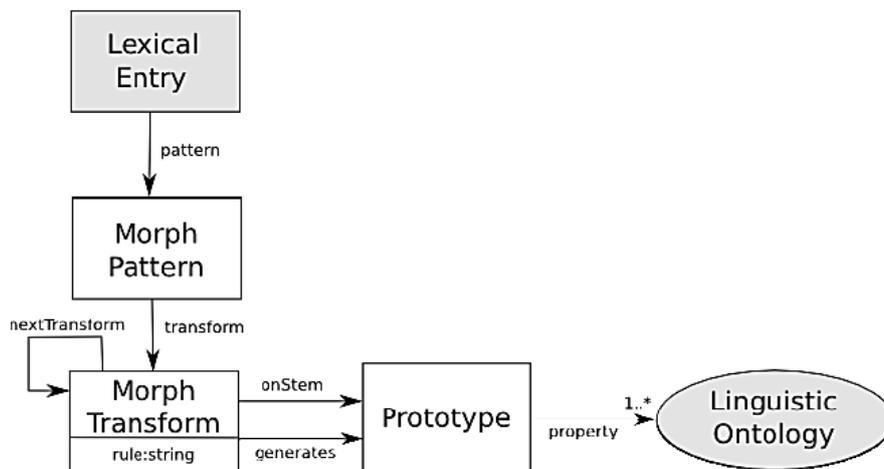


Figura 14 - Módulo de morfología (Fuente: McCrae et al., 2008).

c. UNE-EN 62605:2011

La norma UNE-EN 62605:2011 es una asimilación de la norma internacional IEC 62605:2011, ratificada por la Asociación Española de Normalización y Certificación (AENOR). Especifica el formato de intercambio de diccionarios electrónicos entre editores, creadores de contenido y fabricantes. Los formatos están clasificados en: formato de presentación (apoya una interacción entre autores y preparadores de datos), formato de intercambio (establece un formato de intercambio para los preparadores de datos y editores) y formato del lector (UNE-EN 62605:2011). Como ventajas de este estándar se pueden mencionar la posibilidad de ser utilizado

por cualquier lenguaje y en cualquier país, permite dar formato a los datos mediante atributos como color, tamaño de fuente, así como, añadir imágenes, sonidos y animaciones. Este estándar se basa en XMDF y LeXML, formatos de intercambio que, determinan el diseño de acuerdo con el tamaño de visualización del dispositivo del espectador, la fuente en uso, etc. En este estándar el contenido se denominarán contenido fluido, en oposición al contenido paginado.

En la Figura 15 se muestra la estructura que todo diccionario debe tener, según este estándar:

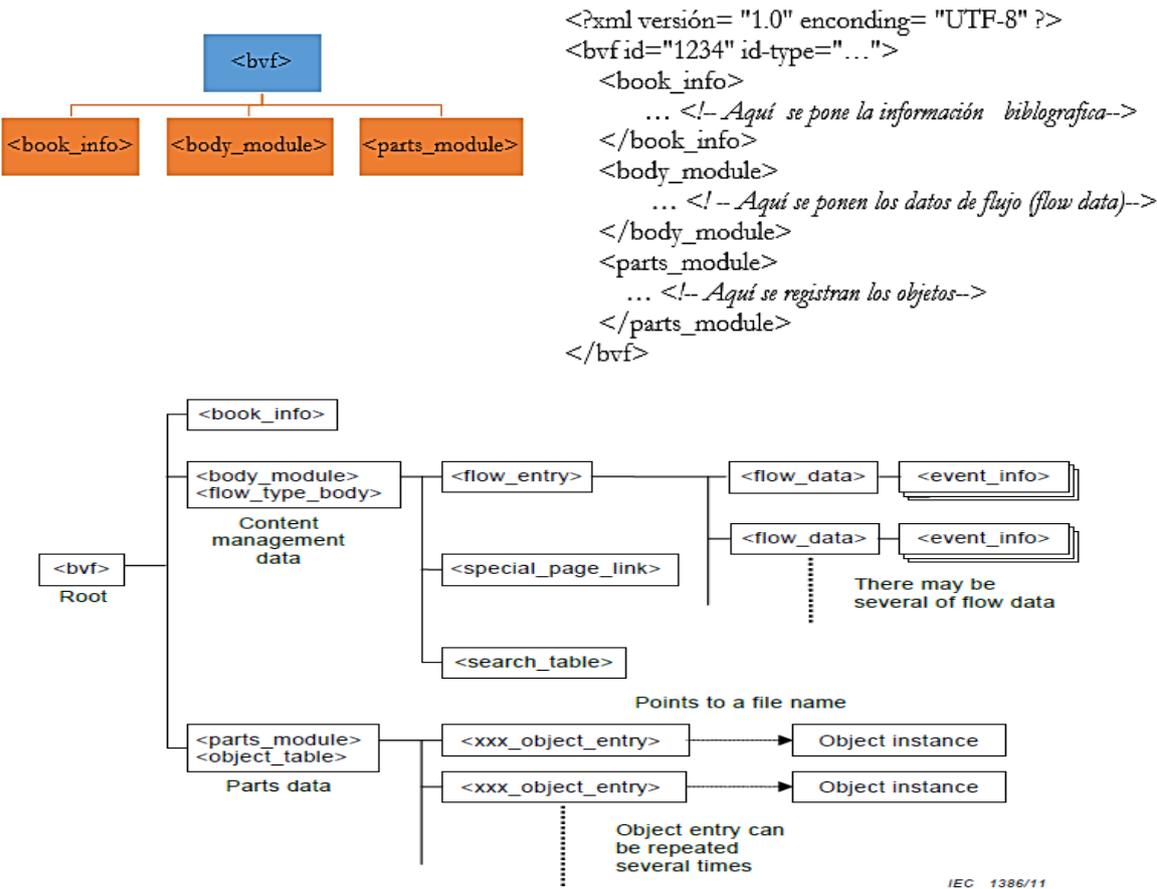


Figura 15. Estructura básica de la norma UNE-EN 62605 (Fuente: UNE-EN 62605:2011)

El módulo principal es el módulo de información de libro (*book information modules*) `<bvf>`. Sirve como una etiqueta raíz para todos los datos del diccionario, toda la información relacionada con el documento se almacena en su interior. Es un contenedor de los elementos `<book_info>`, `<body_module>` y `<parts_module>`. Los atributos que puede tener son: `id_types` (define qué tipo de número es almacenado en el atributo identificador único (`id`)), `id` (registra el número de identificación de este diccionario o e-book), `default_css` (establece el nombre del conjunto de caracteres estándar y caracteres extendidos utilizados en este estándar) y `display_size` (especifica el

tamaño de la pantalla que se asumió al crear el contenido, escrito en el formato de coordenadas estándar).

El siguiente elemento es el módulo datos bibliográficos <book_info> donde se almacena los datos bibliográficos como el autor, el título, información relacionada con el editor, el lugar de publicación, etc. Los elementos secundarios más importantes son <title_info> (almacena la información relativa al título. Puede tener los siguientes sub-elementos: <series_title> (título de la serie), <title> (título), <subtitle> (subtitulo), <edition_info> (información relativa a la revisión histórica del libro) y <title_other_info> (otra información relacionada con el título)) y <author_info> (almacena la información relacionada con el autor. Puede tener los siguientes sub-elementos: <personal_name> o <organization_name> (registra el nombre del autor. Si el autor es individual se usa la primera y si es una organización la segunda), <address_info> (dirección del autor), <author_other_info> (otra información relacionada como el autor)). En la Figura 16 puede observarse que este módulo puede tener más elementos secundarios.

```

<book_info>
  <title_info>
    <series_title>Dummy books</series_title>
    <title>the dummy book of nonsense</title>
    <edition_info>2000/01/01 first edition,
      2005/01/01 second edition</edition_info>
  </title_info>
  <author_info>
    <author role="author">
      <personal_name>
        <first_name>John</first_name>
        <last_name>Smith</last_name>
      </personal_name>
      <address_info>
        <mail_address>john.smith@abcd.com</mail_address>
        <website>http://www.abcd.com/~jsmith/</website>
      </address_info>
    </author>
  </author_info>
  <publisher_info>
    <publisher_office>
      <organization_name>abcd corporation</organization_name>
      <address_info>
        <postal_code>100-1000</postal_code>
        <address>1 main street, Fooobar city, Japan</address>
      </address_info>
    </publisher_office>
  </publisher_info>
  <book_id_info>
    <book_id type="ISBN">xxx-x-xxxx-xxxx-x</book_id>
    <book_id type="Japanese_ID_number">454745-7</book_id>
  </book_id_info>
  <classification_info>
    <classification type="Japanese_C_CODE">2143</classification>
  </classification_info>
  <rating adult="no" violence="no"/>
  <publication_place>jpn</publication_place>
  <publication_date_info>
    <publication_date type="publish">2005</publication_date>
  </publication_date_info>
  <net_price_info>
    <net_price country="jpn" unit="yen">1200</net_price>
  </net_price_info>
  <book_abstract>This book doesn't talk about anything special.</book_abstract>
  <front_cover_image type="image/png">xxx.png</front_cover_image>
  <spine_cover_image type="image/png">yyy.png</spine_cover_image>
  <keyword_list>
    <keyword>dummy</keyword>
    <keyword>nonsense</keyword>
  </keyword_list>
</book_info>

```

Figura 16. Estructura XML del módulo datos bibliográficos (Fuente: UNE-EN 62605:2011)

El siguiente módulo es el módulo de gestión de contenido <body_module>, el cual coordina los datos del contenido en la elaboración del documento real. Tiene un sub-elemento <flow_type_body> maneja los datos del contenido que fluye. Este sub-elemento contiene a su vez tres sub-elementos: <flow_entry> (registra los datos de flujo como contenido que fluye del texto principal. Tiene elementos los siguientes secundarios: <flow_default_attribute> (establece los valores predeterminados de los atributos que se utilizarán para mostrar cada dato de flujo) y <flow_data> (registra el objeto a ser utilizado como contenido del flujo, así como otra información como enlaces de página y eventos), <special_page_link> (datos de páginas especiales) y <search_table> (registra los datos necesarios para crear una tabla de búsqueda).

El módulo información del evento <event_info> registra los eventos de los datos de flujo, eventos entendidos como la reproducción de audio para reaccionar a los clics, enlaces a páginas u otras acciones activadas por el usuario.

El último módulo que forma parte de la estructura definida en el estándar UNE-EN 62605:2011 es el módulo datos de las partes <parts_module>, el cual se utiliza para almacenar y gestionar la información de las partes que son usadas para constituir los datos fluidos (*flow_data*). Tiene el elemento secundario <object_table> en el cual todos los objetos que son usados en el diccionario o libro son descritos. Cada objeto es registrado usando uno de los sub-elementos descritos a continuación:

- <dynamic_text_object_entry> registra objetos de texto sirviendo de base para el siguiente contenido de flujo.
- <sound_object_entry> registra sonidos.
- <search_page_object_entry> registra objetos de páginas de búsqueda.
- <movie_object_entry> registra objetos de película
- <dict_data_object_entry> registra objetos de datos de diccionario.

Debido al propósito de este trabajo, sólo se definirá el subelemento <dict_data_object_entry> el cual registra objetos de diccionario de texto, mediante el subelemento <permission_info> define los permisos relativos al objeto (Figura 17).

```
<parts_module>
  <object_table>
    <dict_data_object_entry src="dict1.xml" type="text/x-bvf-dict-data" object_id="OBD001">
      <permission_info>...</permission_info>
    </dict_data_object_entry>
    ...
  </object_table>
</parts_module>
```

Figura 17. Ejemplo del módulo datos de las partes (Fuente: UNE-EN 62605:2011)

Por último, se hablará de las instancias objeto (*object instances*), utilizado en este estándar para referirse a los objetos que son visualizados o reproducidos por el visor. Los tipos de instancia objeto son los siguientes:

- Instancia de objeto texto (*text object instance*).
- Instancia de objeto datos de diccionario (*dictionary data object instance*).
- Instancia de objeto imagen (*image object instance*).
- Instancia de objeto sonido (*sound object instance*).
- Instancia de objeto animación (*animation object instance*).

- Instancia de objeto páginas de búsqueda (*search page object instance*).
- Instancia de objeto película (*movie object instance*).

A continuación, sólo se comentará la de instancia de objeto datos de diccionario, pues está orientado a diccionarios electrónicos. Las instancias de objeto datos de un diccionario son almacenadas en archivos XML propios, con <dic_data> como elemento raíz. La etiqueta <dic_data> tiene dos elementos secundarios, uno para almacenar los atributos predeterminados para los datos del diccionario <dict_default_attribute>, y otro para almacenar los datos en sí mismos <dict_body>. Dentro de <dict_default_attribute> puede configurarse la orientación del texto, color de la fuente, el fondo de la página, etc. Por su parte el elemento <dict_body> tiene un sub- elemento de entrada <dic_item> que registra información para cada entrada individual. La información de tipo lingüística puede agruparse dentro de este sub-elemento, mediante el uso de los elementos secundarios:

- <gender> información gramatical de género.
- <pronuntiation> registra una cadena alfabética fonética.
- <psp>denota la categoría gramatical. Pueden incluirse aquí otras marcas gramaticales que indiquen el género y el caso, depende de la política del editor.
- <inflec> denota la flexión de la forma de la palabra.
- <slabel> denota la categoría de los términos. Por ejemplo, “filosofía”, “comercio” y sus abreviaturas.
- <guideword> también denota la categoría de los términos. Es utilizada para distinguir múltiples definiciones.
- <spellout> especifica la cadena repetida que se va a reemplazar por otras cadenas que denota omisión o simplemente repetir.
- <variant> denota variantes en una cadena de caracteres estándar.
- <etymology> denote información etimológica en una cadena de caracteres estándar.

d. El Text Encoding Initiative (TEI)

El Text Encoding Initiative (en adelante TEI) se ha desarrollado y se mantiene de forma colaborativa en el seno de un consorcio internacional, compuesto por instituciones académicas, proyectos de investigación e investigadores de todo el mundo, sostenido financieramente por sus miembros. No es un estándar oficial, es decir, no está acreditado por una institución constituida

legítimamente, ejemplo; IEEE, ISO o CEN, pero sí un estándar de facto¹¹ ampliamente utilizado en el ámbito de la Humanidades Digitales para la representación de textos en un formato digital. Su objetivo principal es proporcionar unas directrices, que especifican métodos de codificación para textos legibles por máquinas, principalmente en los campos de humanidades, ciencias sociales y lingüística. (TEI Consortium, 2016). Algunas de las ventajas que tiene el TEI son su independencia frente a cualquier entorno de software en particular, la importancia que da al significado del texto y no a su apariencia, además de, ser diseñado por y para la comunidad de investigación académica.

La estructura general de este estándar, está compuesta por módulos, en cada uno de ellos se explica de forma propia los elementos XML y sus atributos. Dentro de la explicación de un elemento se especifica su asignación a uno o más elementos *clases*. Y otra parte define su posible contenido y atributos con referencia a esas clases. En principio, un esquema TEI puede construirse usando cualquier combinación de módulos. Sin embargo, ciertos módulos TEI son de particular importancia, y deben incluirse en todas las circunstancias excepcionales. El módulo *tei* define clases, macros y tipos de datos los cuales son usados por todos los otros módulos. El módulo básico contiene explicaciones de elementos y atributos que es probable que se necesiten en casi cualquier tipo de documento, y además es recomendado para un uso global. El módulo *cabecera* contiene declaraciones para los elementos de metadatos y atributos que constituyen el encabezado TEI, un componente necesario para que un documento TEI este bien constituido. Mientras que el módulo *estructura* definida de texto describe los elementos básicos estructurales necesarios para la codificación de la mayoría de los objetos tipo libro.

El TEI define una clase como una forma de agrupar elementos que comparten una o varios atributos comunes o aparecen en la misma situación en un modelo de contenido. Existen dos tipos de clase: una clase atributo (*attribute class*) y clases modelo (*model class*). La primera agrupa miembros que comparten atributos comunes, se definen como *att.*, a menudo seguidas de un adjetivo. Una clase particular de esta clase es *att.global*, la cual es común en todos los módulos. La segunda clase es aquella en la que los miembros aparecen en las mismas situaciones, se define como *model.*, por ejemplo, *model.divPart* constituida por los elementos o clases de elementos que aparecen como contenido de un `<div>`.

¹¹ Existe con o sin autoridad de la ley. Se generan cuando una mayoría adopta ciertas reglas o especificaciones para utilizar una tecnología específica. (Álvarez Álvarez, 2003)

Seguidamente se explicará el módulo cabecera del TEI, el cual es necesario para una estructura conforme a este estándar, el módulo estructura definida de texto, y la sección dedicada a diccionarios, ya que es la parte que concierne a este trabajo. El módulo cabecera del TEI se define como <teiHeader> y consta de cinco elementos entre los que destacan una descripción del archivo <fileDesc> (contiene una descripción bibliográfica completa del archivo del ordenador) y una revisión histórica <revisionDesc> (contiene los cambios hechos durante el desarrollo del texto electrónico).

El módulo estructura definida de texto (Figura 18), especifica que un documento TEI debe tener los siguientes elementos:

- <TEI> el elemento raíz
- <teiHeader> contiene declaraciones para los elementos de metadatos y atributos.
- <text> contiene un solo texto de cualquier tipo, ya sea unitario o compuesto, por ejemplo, un poema, una colección de ensayos, una novela o un diccionario.
- <front> contiene cualquier asunto preliminar (encabezados, resúmenes, página de título, prefacios, dedicatorias, etc.) que se encuentra al principio de un documento.
- <body> (cuerpo de texto) contiene todo el cuerpo de un solo texto unitario, excluyendo cualquier parte frontal o posterior.
- <back> contiene cualquier apéndice, etc.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
  </teiHeader>
  <text>
    <front>
    </front>
    <body>
    </body>
    <back>
    </back>
  </text>
</TEI>
```

Figura 18. Estructura de un documento TEI escrito en XML (Fuente: TEI Consortium, 2016)

En cuanto a la sección dedicada a diccionarios del TEI, es importante destacar que debe estructurarse dentro del elemento <body> definido anteriormente. El elemento <body> puede contener alguno de estos elementos, específicos para la codificación de diccionarios: <entry> contiene una entidad estructurada unitaria de un recurso léxico, <entryFree> contiene una sola entrada no estructurada en cualquier tipo de recurso léxico, como un diccionario o un léxico, o <superEntry> agrupa una secuencia de entradas dentro de cualquier tipo de recurso léxico, que funciona como una sola unidad. Existen niveles jerárquicos que se especifican dentro de los elementos previamente mencionados, estos niveles jerárquicos son cinco, aunque los principales son: <form> agrupa información sobre la forma de la palabra, <hom> agrupa información relativa

a un homógrafo dentro de una entrada y <sense> agrupa toda la información relativa a un significado de palabra en una entrada de diccionario.

Los niveles jerárquicos pueden contener trece constituyentes, aunque los principales son: <orth> forma ortográfica de una entrada del diccionario; <pron> pronunciación de la palabra; <gramGrp> agrupa información morfosintáctica acerca del ítem léxico como el género de un ítem léxico <gen>, el número gramatical asociado a la forma <number> o <case> información de casos gramaticales para una forma determinada; y <def> definición de una entrada del diccionario.

```

<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Verbo ser</title>
      </titleStmt>
      <publicationStmt>
        <p>Ejemplo de codificación con TEI </p>
      </publicationStmt>
      <sourceDesc>
        <p>Este documento está sólo en formato electrónico</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <entry xmlid="674">
        <form type="lemma">
          <orth>sum,es,esse,fui</orth>
          <pos>verbo</pos>
        </form>
        <sense xmlid="significado1">
          <def>ser</def>
          <cit type="example">
            <quote>Pulla bona est</quote>
          </cit>
        </sense>
        <sense xmlid="significado2">
          <def>estar</def>
          <cit type="example">
            <quote>Marcus est in domum
            suam</quote>
          </cit>
        </sense>
      </entry>
    </body>
  </text>
</TEI>

```

Figura 19. Ejemplo de codificación TEI en XML de una entrada de un diccionario (TEI Consortium, 2016)

e. Síntesis de los estándares revisados

A modo de resumen, en la Tabla 1 se muestran las características de cada uno de los estándares:

Estándares	Características
LMF	<ul style="list-style-type: none"> ▪ Es un estándar internacional para recursos electrónicos léxicos. ▪ Estructura modular respecto a la información morfológica, sintáctica y semántica. ▪ Interconexión entre los elementos de los módulos de extensión (sintaxis, semántica, morfología, etc.) y el módulo de la estructura base. ▪ Permite la reutilización de los datos para diferentes aplicaciones y tareas software. ▪ Existe una gramática XML para codificar el recurso léxico como un documento XML.
LEMON	<ul style="list-style-type: none"> ▪ Es un estándar orientado a ontologías. ▪ Se basa en RDF. ▪ Uso de recursos externos para las definiciones (referencias a ontologías creadas previamente). ▪ Posibilidad compartir información léxica en la web semántica. ▪ Existe una gramática XML para codificar el recurso léxico como un documento XML.
UNE- EN 62605	<ul style="list-style-type: none"> ▪ Especifica el formato de intercambio de diccionarios electrónicos entre editores, creadores de contenido y fabricantes. ▪ Se basa en dos especificaciones estándares: XMDF y LeXML. ▪ Orientado a la visualización y el formato de los datos. ▪ Posibilidad de contenido dinámico: sonidos, imágenes, animaciones. ▪ No permite la representación de información semántica (rasgos semánticos). ▪ Existe una gramática XML para codificar el recurso léxico como un documento XML.
TEI	<ul style="list-style-type: none"> ▪ Es un estándar de facto para la representación de textos en un formato digital. ▪ Es válido para todo tipo de textos, de cualquier lengua y de cualquier época.

	<ul style="list-style-type: none"> ▪ Independencia frente a cualquier entorno de software. ▪ Diseñado por una comunidad científica muy activa. ▪ Existe una gramática XML para codificar el recurso léxico como un documento XML. ▪ Da importancia al contenido del texto y no a su apariencia.
--	---

Tabla 1- Análisis y síntesis de los estándares vistos

6. Selección del estándar para el Diccionario Didáctico Digital del Latín

El estándar que se ha elegido es el *Lexical Markup Framework* (LMF) porque es el que de forma más directa representa la información, fundamentalmente morfológica y semántica, del DDDL. Esta representación directa de la información léxica del DDDL facilita su intercambio y reutilización en diferentes aplicaciones software mediante documentos XML que contienen el diccionario (o cualquier diccionario del mismo tipo) codificado de forma sistemática y precisa.

Se ha descartado LEMON porque para usarlo sería necesario tener una ontología creada previamente, lo que complica y encarece su construcción y posible reutilización. La norma UNE-EN 62605 tampoco es adecuada para describir el DDDL porque se centra en la visualización y el formato del diccionario, dejando la descripción lingüística en un segundo plano. Por último, el módulo de diccionario del TEI no encaja con el contenido del DDDL y, aunque el TEI tiene un módulo de “Estructuras de Rasgos” que podría ser aplicable, se excluyó porque representaría la información de forma particular, sin un vocabulario estándar compartido, lo que disminuiría la interoperabilidad del DDDL.

7. Aplicación del estándar LMF al DDDL

El procedimiento que se ha seguido para la aplicación del estándar al diccionario ha sido el siguiente: primero, se estableció la correspondencia entre los elementos del estándar y el diccionario; segundo, se tomaron 7 muestras completas de cada tipo de entrada del diccionario para codificarlas; en tercer lugar, se evaluaron los problemas que surgieron tras la codificación, retroalimentándose de la correspondencia para solucionarlos.

El resultado de la aplicación del estándar LMF al DDDL son dos documentos: el Anexo 1 que muestra las tablas de correspondencia y el Anexo 5, un documento XML con la codificación de todas las entradas de muestra.

7.1 Correspondencia

Para la correspondencia, se ha usado sólo la estructura base y las extensiones de semántica de PLN y morfología de PLN del LMF, puesto que las entradas requieren este tipo de información.

Como se vio en la revisión del LMF, la estructura base tiene un elemento raíz, *LexicalResource*, el elemento *GlobalInformation*, y el elemento *Lexicon*, el cual contiene todas las entradas léxicas del diccionario. Estos elementos se corresponden con el diccionario entendido como un Recurso Léxico formado por un único Lexicón. Si se añaden nuevos diccionarios, cada uno de ellos sería un nuevo *Lexicon*. Puede verse la definición y atributos del DDDL detallados en la Tabla de correspondencia 1 del Anexo 1.

La muestra de entradas escogida comprende un verbo, un sustantivo, un adjetivo, un adverbio, una preposición, un pronombre y una conjunción. La correspondencia en cada caso se estableció de la forma siguiente:

- (i) en primer lugar, se escogió como muestra, un verbo con dos marcos predicativos, cada uno de ellos con una valencia cuantitativa diferente. En la Figura 20, mediante el uso de cuadros de diferentes colores, se puede apreciar la correspondencia entre los elementos del DDDL y los del LMF, para la entrada de dicho verbo.
- (ii) En segundo lugar, el sustantivo tomado como muestra es un sustantivo con dos significados diferentes. La correspondencia de esta entrada puede verse en la Figura 21.
- (iii) Por último, en las figuras 22, 23, 24, 25 y 26 se hace la correspondencia con un adjetivo, adverbio, preposición, pronombre y conjunción, siguiendo la misma estrategia de los cuadros de diferente color, para que sea mucho más ilustrativo. Asimismo, en el Anexo 1 podrá encontrarse las tablas de correspondencia de cada una de las entradas mencionadas.

<p>Identificador: 646</p> <p>LEMA: peto, is, ere, ii (ivi), itum (3ª)</p> <p>LETRA: p</p> <p>CATEGORÍA: Verbo</p> <p>Significado 1: Pedir algo a alguien</p> <p>Numero de Argumentos (Verbos): TRIVALENTE</p> <p>Primer argumento : <i>Nominativo</i> Caracterización Argumental 1: +animado +humano</p> <p>Segundo argumento: <i>Acusativo</i> Caracterización Argumental 1: -animado +definido Caracterización Argumental 2: -animado -definido</p> <p>Tercer argumento: <i>A /Ab+ABL /Ex+ABL</i> Caracterización Argumental 1: +animado +humano</p> <p>Ejemplo: Poeta librum ex bibliothecario petit</p> <p>Significado 2: atacar</p> <p>Numero de Argumentos (Verbos): BIVALENTE</p> <p>Primer argumento : <i>Nominativo</i> Caracterización Argumental 1: +animado +humano</p> <p>Segundo argumento : <i>Acusativo</i> Caracterización Argumental 1: +animado +humano</p> <p>Ejemplo: Captivi romani Belgas petunt</p>	<p>LexicalEntry: partOfSpeech: verb id: id646 Lemma: writtenForm: peto, is, ere, ii (ivi), itum (3ª)</p> <p>Sense: id: id646.1 SenseExample: text: Poeta librum ex bibliothecario petit Definition: text: Pedir algo a alguien PredicativeRepresentation: SemanticPredicate: id: marco646.1 label: ISOtrivalent SemanticArgument: label: ARG1 case: nominativeCase fs: semanticAnimacy: animate semanticType: human SemanticArgument: label: ARG2 case: accusativeCase fs: semanticAnimacy: inanimate semanticType: concrete fs: semanticAnimacy: inanimate semanticType: abstract SemanticArgument: label: ARG3 case: A/Ab+ABL/Ex+ABL fs: semanticAnimacy: animate semanticType: human</p>	<p>Sense: id: id646.2 SenseExample: text: Captivi romani Belegas petunt Definition: text: atacar PredicativeRepresentation: SemanticPredicate: id: marco641.2 label: ISObivalent SemanticArgument: label: ARG1 case: nominativeCase fs: semanticAnimacy: animate semanticType: human SemanticArgument: label: ARG2 case: accusativeCase fs: semanticAnimacy: animate semanticType: human</p>
---	---	---

Figura 20. Correspondencia de la entrada de un verbo del DDDL(izq.) y su representación en el estándar LMF(der.)

<p>Identificador: 345</p> <p>LEMA: <i>balteum, i (n.)</i></p> <p>LETRA: B</p> <p>CATEGORÍA: <i>Sustantivo</i></p> <p>Significado: <i>vaina</i></p> <p>Caracterización Léxica (Sustantivos): <i>-animado +definido</i></p> <p>Significado 2: <i>cinturón</i></p> <p>Caracterización Léxica: <i>-animado +definido</i></p>	<p>LexicalEntry:</p> <p>partOfSpeech: noun</p> <p>id: id345</p> <p>Lemma: writtenForm: <i>balteum, i(n.)</i></p> <p>Sense: id: id345.1</p> <p>Definition: text: <i>vaina</i></p> <p>fs: semanticAnimacy: inanimate semanticType: concrete</p> <p>Sense: id: id345.2</p> <p>Definition: text: <i>cinturón</i></p> <p>fs: semanticAnimacy: inanimate semanticType: concrete</p>
--	---

Figura 21. Correspondencia de la entrada de un sustantivo del DDDL(izq.) y su representación en el estándar LMF(der.)

<p>Identificador: 532</p> <p>LEMA: <i>pulcher, chra, chrum</i></p> <p>LETRA: P</p> <p>CATEGORÍA: <i>Adjetivo</i></p> <p>Significado: <i>hermoso</i></p>	<p>LexicalEntry:</p> <p>partOfSpeech: adjective</p> <p>id: id532</p> <p>Lemma: writtenForm: <i>pulcher, chra, chrum</i></p> <p>Sense: id: id532.1</p> <p>Definition: text: <i>hermoso</i></p>
---	---

Figura 22. Correspondencia de la entrada de un adjetivo del DDDL(izq.) y su representación en el estándar LMF(der.)

Identificador: 613 LEMA: <i>saepe</i> LETRA: S CATEGORÍA: <i>Adverbio</i> Significado: <i>a menudo, con frecuencia</i>	LexicalEntry: partOfSpeech: adverb id: id613 Lemma: writtenForm: saepe Sense: id: id613.1 Definition: text: a menudo, con frecuencia
--	---

Figura 23. Correspondencia de la entrada de un adverbio del DDDL(izq.) y su representación en el estándar LMF(der.)

Identificador: 614 LEMA: <i>in</i> LETRA: I CATEGORÍA: <i>Preposición</i> Significado: <i>a, hacia (ACU)</i> Significado 2: <i>en (ABL)</i>	LexicalEntry: partOfSpeech: preposition id: id614 Lemma: writtenForm: in Sense: id: id614.1 Definition: text: a, hacia(ACU) Sense: id: id614.2 Definition: text: en(ABL)
---	---

Figura 24. Correspondencia de la entrada de una preposición del DDDL(izq.) y su representación en el estándar LMF(der.)

Identificador: 606 LEMA: <i>is, ea, id</i> LETRA: I CATEGORÍA: <i>Pronombre</i> Significado: <i>este, esta, esto</i>	LexicalEntry: partOfSpeech: pronoun id: id606 Lemma: writtenForm: is,ea,id Sense: id: id606.1 Definition: text: este, esta, esto
--	---

Figura 25. Correspondencia de la entrada de un pronombre del DDDL(izq.) y su representación en el estándar LMF(der.)

Identificador: 611 LEMA: <i>atque, ac</i> LETRA: A CATEGORÍA: <i>Conjunción</i> Significado: <i>y</i>	LexicalEntry: partOfSpeech: conjunction id: id611 Lemma: writtenForm: atque, ac Sense: id: id611.1 Definition: text: y
---	---

Figura 26. Correspondencia de la entrada de una conjunción del DDDL(izq.) y su representación en el estándar LMF(der.)

7.2 Codificación LMF-XML

Para poder codificar la estructura argumental semántica del verbo y las caracterizaciones argumentales de los argumentos que los acompañan, se ha modificado la DTD¹² al igual que el proyecto Cornetto (Maks, van der Vliet, Görög, Vossen, & VU University Amsterdam, 2013). En la aplicación del estándar, se ha detectado un fallo en la DTD definida por el estándar. Este fallo hace referencia a los sub-elementos que puede tener el elemento *PredicativeRepresentation*, propio de la extensión de semántica, en la DTD se define que sólo puede tener como sub-elemento *feat* (característica), mientras que en el diagrama UML¹³ *PredicativeRepresentation* tiene un sub-elemento que es *SemanticPredicate*. Además, tal y como lo expresan Windhouwer, Petro, & Shayan (2014) el LMF tiene las siguientes carencias:

- Los enlaces desde las características (*features*) al ISOcat DCR no están suficientemente especificados.
- Otras representaciones alternativas de estructuras de rasgos no son compatibles con este estándar, e.g. las estructuras de características TEI/ISO.
- No todas las construcciones y restricciones UML pueden ser soportadas por una DTD y como resultado el esquema es más laxo que el modelo UML de estándares.

A continuación, se detallan los cambios que se realizaron en la DTD:

- Los sub-elementos que conforman el Recurso Léxico se han reducido a tres elementos: *LexicalResource* (*feat*, *GlobalInformation*, *Lexicon*), ya que los demás elementos no son necesarios en la codificación.
- El valor de la versión de la DTD se ha cambiado por “UCM” en lugar de “16”, debido a que es una versión concreta para el DDDL y en este caso, es un Trabajo de Fin de Grado de realizado para la Universidad Complutense de Madrid.
- Los sub-elementos que conforman el elemento *Lexicon*, se han simplificado de la siguiente forma: *Lexicon* (*feat*, *LexicalEntry*) porque los elementos descartados, pertenecen a otras extensiones del LMF, que no se han utilizado.
- Se han acotado los sub-elementos que pueden formar parte del elemento *LexicalEntry*: *LexicalEntry* (*feat*, *Lemma*, *WordForm*, *Stem*, *MorphologicalPattern*, *Sense*) ya que con estos elementos se ha codificado la información de las entradas del DDDL.

¹² Documento que define la estructura de un documento XML

¹³ Ver Figura 7

- Se ha añadido el elemento *fs* (*feature structure*-estructura de rasgos): *fs* (*feat**). Este elemento se ha tomado como ejemplo del TEI. Este cambio se debe a la carencia del estándar para representar la caracterización morfológica de los argumentos verbales y la caracterización léxica de los sustantivos.
- Se han disminuido los sub-elementos del elemento *Sense* y se añadió el elemento *fs*, de la siguiente forma: *Sense* (*feat*, *SenseExample*, *PredicativeRepresentation*, *Definition*, *fs*). Con el propósito de usar sólo los elementos necesarios para representar el significado de cada entrada.
- El elemento *Sense* sólo tiene el atributo ID, porque el otro atributo “synset” se emplea para hacer referencia a la clase *Synset*, que en este caso no se utiliza.
- Los sub-elementos que conforman los elementos *Definition*, *Lemma*, *WordForm* y *Stem* se han reducido de la siguiente manera: *Definition* (*feat*), *Lemma* (*feat*), *WordForm* (*feat*), *Stem* (*feat*), ya que los otros elementos no se han utilizado en la codificación
- Se ha corregido el defecto de la DTD proporcionada por el estándar: *PredicativeRepresentation* (*feat*, *SemanticPredicate*).
- El carácter obligatorio u opcional de aparición de los atributos del elemento *PredicativeRepresentation* se han cambiado de REQUIRED (obligatorio) a IMPLIED (voluntario). Esto se ha modificado porque al ser obligatorio es necesario utilizar la extensión de sintaxis que no se ha utilizado, este atributo es un enlace entre la extensión de semántica y sintaxis.
- Se ha añadido el atributo *label* al elemento *SemanticPredicate* para codificar la estructura argumental semántica del verbo y al elemento *SemanticArgument* para establecer el orden de los argumentos.

En los Anexos 3 y 4 pueden compararse la DTD original y la DTD ajustada para la codificación del DDDL.

Como se ha visto en la sección anterior todas las entradas deben ir dentro del elemento *Lexicon*, que a su vez es un sub-elemento de *Lexical Resource*, así que, primero se establecerá una codificación de lo que es el diccionario en general y después de forma particular la codificación de las entradas.

La codificación se hace en un documento XML, en el que es importante declarar las dos primeras líneas que se ven en la Figura 27, la primera especifica la versión XML que se usa, mientras que, la segunda, la DTD en la que se basa el documento XML en sí. Además, tiene un elemento raíz, *LexicalResource*, con un atributo “*dtd Version*” el cual determina la versión de la DTD; el

elemento *GlobalInformation*, con el atributo *languageCoding*, donde se define el estándar que se va a utilizar para codificar los nombres de los lenguajes dentro del recurso léxico, en este caso se utiliza el ISO 639-3¹⁴; el elemento *Lexicon*, contiene todas las entradas léxicas del diccionario, mediante el atributo “*language*” y el valor “*spa*” se especifica que todas las entradas léxicas están en español.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM "LMFdtd.dtd">
<LexicalResource dtdVersion="UCM">
  <GlobalInformation>
    <feat att="languageCoding" val="ISO639-3"/>
    <feat att="title" val="Diccionario Didáctico Latín"/>
    <feat att="URL" val="http://repositorios.fdi.ucm.es/DiccionarioDidacticoLatin"/>
  </GlobalInformation>
  <Lexicon>
    <feat att="language" val="spa"/>
    </LexicalEntry> (Aquí van cada una de las entradas del diccionario)
  </LexicalEntry>
</Lexicon>
</LexicalResource>
```

Figura 27. Codificación de la macroestructura diccionario en el estándar LMF

Cada entrada léxica se codifica usando el elemento *LexicalEntry*, de modo que, en adelante, cada entrada será un sub-elemento de *Lexicon*. La codificación del verbo “*peto, is, ere, ii(ivi), itum (3ª)*” en el estándar LMF es muy extensa, en consecuencia, se muestra en el Anexo 2. En la Figura 28 se presenta la codificación del sustantivo “*balteum, i (n)*”, que se corresponde a la estructura de la Figura 21.

¹⁴ ISO que codifica los lenguajes humanos conocidos («ISO 639-3», 2017)

```

<LexicalEntry id="id345">
  <feat att="partOfSpeech" val="noun"/>
  <Lemma>
    <feat att="writtenForm" val="balteum,i(n)"/>
  </Lemma>
  <Sense>
    <feat att="id" val="signifi"/>
    <Definition>
      <feat att="text" val="vaina"/>
    </Definition>
    <fs>
      <feat att="semanticAnimacy" val="inanimate"/>
      <feat att="semanticType" val="concrete"/>
    </fs>
  </Sense>
  <Sense>
    <feat att="id" val="id345.2"/>
    <Definition>
      <feat att="text" val="cinturón"/>
    </Definition>
    <fs>
      <feat att="semanticAnimacy" val="inanimate"/>
      <feat att="semanticType" val="concrete"/>
    </fs>
  </Sense>
</LexicalEntry>

```

Figura 28. Codificación XML del sustantivo *balteum* en el estándar LMF

La codificación de los adjetivos, adverbios, preposiciones, pronombres y conjunciones se realiza de forma similar, así que, posteriormente, en las figuras 29, 30, 31, 32 y 33 se expone la codificación en el estándar LMF de las categorías gramaticales mencionadas anteriormente.

```

<LexicalEntry id="id532">
  <feat att="partOfSpeech" val="adjective"/>
  <Lemma>
    <feat att="writtenForm" val="pulcher, chra, chrum"/>
  </Lemma>
  <Sense id="id532.1">
    <Definition>
      <feat att="text" val="hermoso"/>
    </Definition>
  </Sense>
</LexicalEntry>

```

Figura 29 – Codificación del adjetivo *pulcher*

```

<LexicalEntry id="id613">
  <feat att="partOfSpeech" val="adverb"/>
  <Lemma>
    <feat att="writtenForm" val="saepe"/>
  </Lemma>
  <Sense id="id613.1">
    <Definition>
      <feat att="text" val="a menudo, con frecuencia"/>
    </Definition>
  </Sense>
</LexicalEntry>

```

Figura 30 – Codificación del adverbio *saepe*

```

<LexicalEntry id="id614">
  <feat att="partOfSpeech" val="preposition"/>
  <Lemma>
    <feat att="writtenForm" val="in"/>
  </Lemma>
  <Sense id="id614.1">
    <Definition>
      <feat att="text" val="a_hacia(ACU)"/>
    </Definition>
  </Sense>
  <Sense id="id614.2">
    <Definition>
      <feat att="text" val="en(ABL)"/>
    </Definition>
  </Sense>
</LexicalEntry>

```

Figura 31 – Codificación de la preposición *in*

```

<LexicalEntry id="id606">
  <feat att="partOfSpeech" val="pronoun"/>
  <Lemma>
    <feat att="writtenForm" val="is_ea_id"/>
  </Lemma>
  <Sense id="id606.1">
    <Definition>
      <feat att="text" val="este, esta, esto"/>
    </Definition>
  </Sense>
</LexicalEntry>

```

Figura 32 – Codificación del pronombre *is, ea, id*

```

<LexicalEntry id="id611">
  <feat att="partOfSpeech" val="conjunction"/>
  <Lemma>
    <feat att="writtenForm" val="atque_ac"/>
  </Lemma>
  <Sense id="id611.1">
    <Definition>
      <feat att="text" val="y"/>
    </Definition>
  </Sense>
</LexicalEntry>

```

Figura 33 – Codificación de la conjunción *atque, ac*

7.3 Resultados y discusión

A pesar de que, el LMF no esté planteado específicamente para un diccionario de las características del DDDL (onomasiológico, morfo-semántico), se ha conseguido codificar de forma satisfactoria, entradas de categorías gramaticales diferentes, al igual que la información adyacente a ellas. El diccionario didáctico ahora es portable y, al estar codificado según el LMF, es legible tanto por humanos como por aplicaciones software, gracias al formato XML.

Sin embargo, la codificación del DDDL a LMF no podría haberse realizado, sin el ajuste de la DTD, una cuestión a tener en cuenta para otros proyectos. Estos ajustes, obedecen criterios de concepción del diccionario o lexicón: el estándar LMF está centrado en la descripción de los aspectos morfo-sintácticos a partir de los cuales define los aspectos semánticos del léxico. El DDDL, como se describe en la sección 1, está centrada en los aspectos morfo-semánticos de los verbos que son las piezas clave del diccionario.

Asimismo, los elementos de las extensiones empleadas, han sido utilizados de forma sintética y enfocados a los requisitos del diccionario didáctico. Si fuese necesario ampliar la

información lingüística del diccionario, podría emplearse otras extensiones, generando una estructura mucho más compleja e interconexiones adicionales entre los elementos.

8. Resumen y conclusiones

Con este trabajo se ha tratado de responder a la cuestión de cómo estandarizar la información de un tipo de diccionario electrónico, en este caso didáctico, para facilitar su reusabilidad, portabilidad e interoperabilidad. De esta forma se mejora la rentabilidad de estos recursos léxicos que requieren un importante esfuerzo e inversión económica para su construcción.

Para ello, primero se ha realizado un estudio lingüístico de la estructura del diccionario, con el propósito de obtener sus requisitos lingüísticos y didácticos. En segundo lugar, se ha realizado un análisis y síntesis de los estándares que dio como resultado la selección del LMF como el más adecuado para desarrollar proyectos de diccionarios que necesitan únicamente representar de forma precisa la información léxica que contienen con el fin de que sea reutilizable e interoperable. En este Trabajo de Fin de Grado, la información del diccionario DDDL que se necesitaba describir era, fundamentalmente, léxica -morfológica y semántica- y primaba la posibilidad de reutilización de los datos para diferentes aplicaciones y tareas software. Finalmente se realizó una correspondencia entre los elementos del DDDL y el estándar. De esta fase es importante resaltar que fue necesario ajustar el esquema LMF -DTD- para ajustarlo al enfoque morfo-semántico del DDDL. Además, se ha visto que existen incongruencias en el DTD LFM con respecto a la documentación del estándar -dada en los diagramas UML- y que existe una imprecisión en los elementos del *Data Category Registry* (DCR) que deben usarse en este estándar.

En conclusión, este trabajo valida, mediante un estudio de caso, el uso de estándares como estrategia para mejorar la interoperabilidad, re-usabilidad y portabilidad de los diccionarios electrónicos. De este modo, la información del diccionario DDDL puede usarse con otros fines, como por ejemplo para la creación de herramientas informáticas para generar ejercicios en línea de aprendizaje de latín. Actualmente, en el marco de un Trabajo de Fin de Grado de la Facultad de Informática, se está construyendo una herramienta *e-learning* que usa el DDDL codificado con LMF y las pautas de estandarización resultado de este Trabajo de Fin de Grado. Asimismo, los resultados del trabajo también pueden aplicarse al desarrollo de otros diccionarios con estructura similar al Diccionario Didáctico Digital de Alemán actualmente en construcción, <http://repositorios.fdi.ucm.es/DiccionarioDidacticoAleman/>. Finalmente, el trabajo aporta un análisis y síntesis de los principales estándares y guías aplicables a la construcción de Diccionarios Electrónicos que pueden servir de punto de partida de nuevos proyectos relacionados con estos recursos léxicos.

Bibliografía

- Albusac Jiménez, J. A., González Morcillo, C., & Jiménez Linares, L. (2007). Babel; Plataforma e-learning para el aprendizaje colaborativo de idiomas. *Recuperado el 26/01/2017, 24*. Recuperado a partir de http://www.esi.uclm.es/www/jalbusac/doc/babel_cuiet.pdf
- Álvarez Álvarez, J. V. (2003). Uso de estándares e-learning en espacios educativos. *Revista Fuentes, 5, 122-142*. Recuperado a partir de <https://idus.us.es/xmlui/handle/11441/32081>
- Arnold, D., Balkan, L., Meijer, S., Humphreys, R. L., & Sadler, L. (1994). Chapter 5: Dictionaries. En *Machine Translation: an Introductory Guide*. London. Recuperado a partir de <http://www.essex.ac.uk/linguistics/external/clmt/mtbook/PostScript/ch5.pdf>
- Azorín Fernández, D. (2000). Los diccionarios didácticos del español desde la perspectiva de sus destinatarios. *ELUA. Estudios de Lingüística, N. 14 (2000); pp. 19-44*. Recuperado a partir de <http://rua.ua.es/dspace/handle/10045/6267>
- British Standard Institute. (2017). Information about standards. Recuperado 23 de enero de 2017, a partir de <https://www.bsigroup.com/en-GB/standards/Information-about-standards/>
- Diccionario Didáctico Digital de Latín. (2016). Recuperado 12 de enero de 2017, a partir de http://repositorios.fdi.ucm.es/DiccionarioDidacticoLatin/view/cm_view_virtual_object.php?idov=646&seleccion=1
- Dik, S. C. (1997). *The Theory of Functional Grammar. Part 1. The Structure of the Clause*; (2nd, ed. ed.). Berlin, Germany: Mouton de Gruyter.
- Fernández-Pampillón Cesteros, A. (2016). *Proyecto Innova Docencia 2016-2017: Construcción de un diccionario didáctico digital para la mejora del aprendizaje del latín y estudio preliminar de su extensión a otras lenguas* (p. 21). Madrid: Universidad Complutense (Innova Docencia).
- Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M., ... others. (2006). Lexical markup framework (LMF). En *Proceedings of LREC* (Vol. 6). Recuperado a partir de <https://hal.archives-ouvertes.fr/inria-00121468/document>
- Gross, G. (1994). Classes d'objets et description des verbes. *Langages, 115, 15-30*.
- Ide, N., & Romary, L. (2004). A Registry of Standard Data Categories for Linguistic Annotation. En *In Proceedings of the Fourth Language Resources and Evaluation Conference (LREC)*. Lisbon (pp. 135–139).
- ISO 24613:2008 Language resource management - Lexical markup framework (LMF)
- ISO 639-3. (2017, enero 28). En *Wikipedia, la enciclopedia libre*. Recuperado a partir de https://es.wikipedia.org/w/index.php?title=ISO_639-3&oldid=96506352
- ISOCat - a Data Category Registry. (2017). Recuperado 26 de enero de 2017, a partir de <http://www.isocat.org/>

LEMON - The Lexicon Model for Ontologies <http://lemon-model.net/index.php>

Lenguaje unificado de modelado. (2017, enero 24). En *Wikipedia, la enciclopedia libre*. Recuperado a partir de

https://es.wikipedia.org/w/index.php?title=Lenguaje_unificado_de_modelado&oldid=96416773

Maks, I., van der Vliet, H., Görög, A., Vossen, P., & VU University Amsterdam. (2013, mayo). Cornetto LMF. Lexical Resource for Dutch. CLARIN project CLARIN - NL - 11 - 020.

Márquez, C., & Chaves, Y. (2016). A Latin Functionalist Dictionary as a Self-Learning Language Device: Previous Experiences to Digitalization. *Education Sciences*, 6(3), 23. <https://doi.org/10.3390/educsci6030023>

Masie, E. (2003). *Making Sense of Learning Specifications & Standards: A Decision Maker's Guide to their Adoption* (p. 86). Saratoga Springs: The Masie Center. Recuperado a partir de http://www.immagic.com/eLibrary/ARCHIVES/GENERAL/MASIE_US/M031120M.pdf

McCrae, J., Aguado-de-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., ... Wunner, T. (2008). The lemon cookbook. Moonet. Recuperado a partir de <http://lemon-model.net/learn/cookbook.html>

Real Academia Española. (2014). Diccionario. En Diccionario de la lengua española (23.a ed.). Recuperado de <http://dle.rae.es/?id=DgIqVCc>.

TEI Consortium. (2016). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Recuperado a partir de <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>

Tesnière, L. (1994). *Elementos de sintaxis estructural*. Madrid: Gredos.

UNE-EN 62605:2011. Equipos y sistemas multimedia. Libros y publicaciones electrónicas multimedia. Formato de intercambio para diccionarios electrónicos.

What is Unicode? (2017). Recuperado 26 de enero de 2017, a partir de <http://www.unicode.org/standard/WhatIsUnicode.html>

Wiersma, W., & Jurs, S. G. (2005). *Research Methods in Education: An Introduction, 8th Edition* (8th Edition). Pearson.

Windhouwer, M., Petro, J., & Shayan, S. (2014). RELISH-LMF - Unlocking the Full Power of the Lexical Markup Framework. En Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014). Reykjavik, Iceland. Recuperado a partir de <https://github.com/TheLanguageArchive/RELISH-LMF>

ANEXO 1

TABLAS DE CORRESPONDENCIA

Tabla de correspondencia 1. Macroestructura del diccionario en el LMF

ELEMENTO LMF	ATRIBUTO	VALOR	DEFINICION
<Lexical Resource>			Es el elemento raíz
	dtdVersion		Especifica la de dtd en la que está basada el documento XML
		UCM	La versión de la dtd
<GlobalInformation>			Información administrativa y otros atributos generales de todo el recurso léxico
	languageCoding		Define el estándar utilizado para codificar los nombres de los lenguajes dentro del recurso léxico
		“ISO 639-3”	ISO que codifica los lenguajes humanos conocidos.
<Lexicon >			Contiene todas las entradas léxicas
	language		Especifica el lenguaje de las entradas léxicas
		spa	Spa se corresponde a español en el ISO 639-3

Tabla de correspondencia 2. Correspondencia verbos

ELEMENTO DDDL	ELEMENTO LMF	ATRIBUTO	VALOR	DEFINICIÓN
Entrada	<LexicalEntry>			
Categoría		partOfSpeech		Describe la categoría

				gramatical de la entrada léxica
			verb	Define que la entrada léxica es un verbo
Identificador		id		Identificador de la entrada léxica
			646	Marca el identificador de cada entrada léxica tomando como referencia el identificador de Oda
Lema	<Lemma>			Define el lema
		writtenForm		Especifica el lema
	<Sense>			Describe el marco predicativo
		id		Identificador del marco predicativo
Ejemplo	<SenseExample>			Ejemplo que explica el significado
		text		Ejemplo escrito
Significado	<Definition>			Define el significado del verbo
		text		Significado del verbo

	<PredicativeRepresentation>			Define el marco predicativo
	<SemanticPredicate>			Expresa el orden de los argumentos
		id		Identificador del predicado semántico
Número de argumentos		label		Define los argumentos que intervienen el predicado semántico
			ISObivalent	El verbo tiene dos argumentos
			ISOtrivalent	El verbo tiene tres argumentos
Argumentos	<SemanticArgument>			Define los argumentos semánticos
		label		Determina el orden de los argumentos
Primer argumento			ARG1	Primer argumento
Segundo argumento			ARG2	Segundo argumento
Tercer argumento			ARG3	Tercer argumento
Caracterización argumental	<fs>			Estructura de rasgos semánticos

		case		Caso del argumento semántico
			nominative Case	Caso nominativo
			accusative Case	Caso acusativo
			ablativeCase	Caso ablativo
			dativeCase	Caso dativo
	<feat>			Elemento necesario para usar los valores del DCR
		semanticAnimacy		Describe los rasgos semánticos del argumento
			animate	animado
			inanimate	no animado
		semanticType		Describe los rasgos semánticos del argumento
			abstract	abstracto
			concrete	concreto
			human	humano
			place	lugar
			nonhuman	no humano
			time	tiempo

Tabla de correspondencia 3. Correspondencia sustantivos

ELEMENTO DDDL	ELEMENTO LMF	ATRIBUTO	VALOR	DEFINICIÓN
Entrada	<LexicalEntry>			Entrada léxica. Este elemento recoge todos los elementos de una entrada léxica
Categoría		partOfSpeech		Describe la categoría gramatical de la entrada léxica
			noun	Define que la entrada léxica es un sustantivo
Identificador		id		Identificador de la entrada léxica
			345	Marca el identificador de cada entrada léxica tomando como referencia el identificador de Oda
Lema	<Lemma>			Define el lema
		writtenForm		Marca el lema
	<Sense>			Expresa el significado del sustantivo
		id		Identificador del significado del sustantivo
Significado	<Definition>			Define el significado del sustantivo
		text		Significado del sustantivo
Caracterización léxica	<fs>			Estructura de rasgos semánticos
	<feat>			Elemento necesario para usar los valores del DCR
		semanticAnimacy		Describe los rasgos semánticos del sustantivo

			animate	animado
			inanimate	no animado
		semanticType		Describe los rasgos semánticos del sustantivo
			abstract	abstracto
			concrete	concreto
			human	humano
			place	lugar
			nonhuman	no humano
			time	tiempo

Tabla de correspondencia 4. Correspondencia adjetivos, adverbios, preposiciones, pronombres y conjunciones

ELEMENTO DDDL	ELEMENTO LMF	ATRIBUTO	VALOR	DEFINICIÓN
Entrada	<LexicalEntry>			Entrada léxica. Este elemento recoge todos los elementos de una entrada léxica
Categoría		partOfSpeech		Describe la categoría gramatical de la entrada léxica
			adjective	Define que la entrada léxica es un adjetivo
			adverb	Define que la entrada léxica es un adverbio
			preposition	Define que la entrada léxica es una preposición
			pronoun	Define que la entrada léxica es un pronombre
			conjunction	Define que la entrada léxica es una conjunción

Identificador		id		Identificador de la entrada léxica
			XXX	Marca el identificador de cada entrada léxica tomando como referencia el identificador de Oda
Lema	<Lemma>			Define el lema
		writtenForm		Marca el lema
	<Sense>			Expresa el significado del adjetivo/adverbio /preposición/pronombre / conjunción
		id		Identificador del significado del adjetivo/adverbio/ preposición/pronombre/ conjunción
Significado	<Definition>			Define el significado del adjetivo/adverbio/ preposición/pronombre/ conjunción
		text		Significado del adjetivo/adverbio/preposición/pronombre/ conjunción

ANEXO 2

CODIFICACIÓN LMF DEL VERBO PETO

```

<LexicalEntry id="id646">
  <feat att="partOfSpeech" val="verb"/>
  <Lemma>
    <feat att="writtenForm" val="peto, is, ere, ii(ivi), itum(3ª)"/>
  </Lemma>
  <Sense id="id646.1">
    <SenseExample>
      <feat att="text" val="Poeta librum ex bibliothecario petit"/>
    </SenseExample>
    <Definition>
      <feat att="text" val="Pedir algo a alguien"/>
    </Definition>
    <PredicativeRepresentation>
      <SemanticPredicate id="marco646.1" label="ISOtrivalent">

        <SemanticArgument label="ARG1">
          <feat att="case" val="nominativeCase"/>
          <fs>
            <feat att="semanticAnimacy" val="animate"/>
            <feat att="semanticType" val="human"/>
          </fs>
        </SemanticArgument>
        <SemanticArgument label="ARG2">
          <feat att="case" val="accusativeCase"/>
          <fs>
            <feat att="semanticAnimacy" val="inanimate"/>
            <feat att="semanticType" val="concrete"/>
          </fs>
          <fs>
            <feat att="semanticAnimacy" val="inanimate"/>
            <feat att="semanticType" val="abstract"/>
          </fs>
        </SemanticArgument>
        <SemanticArgument label="ARG3">
          <feat att="case" val="A/Ab+ABL/Ex+ABL"/>
          <fs>
            <feat att="semanticAnimacy" val="animate"/>
            <feat att="semanticType" val="human"/>
          </fs>
        </SemanticArgument>
      </SemanticPredicate>
    </PredicativeRepresentation>
  </Sense>
  <Sense id="id646.2">
    <SenseExample>
      <feat att="text" val="Captivi romani Belgas petunt"/>
    </SenseExample>
  </Sense>
</LexicalEntry>

```

```

<Definition>
  <feat att="text" val="atacar"/>
</Definition>
<PredicativeRepresentation>
  <SemanticPredicate "marco646.2" label="ISObivalent">

    <SemanticArgument label="ARG1">
      <feat att="case" val="nominativeCase"/>
      <fs>
        <feat att="semanticAnimacy" val="animate"/>
        <feat att="semanticType" val="human"/>
      </fs>
    </SemanticArgument>
    <SemanticArgument label="ARG2">
      <feat att="case" val="accusativeCase"/>
      <fs>
        <feat att="semanticAnimacy" val="animate"/>
        <feat att="semanticType" val="human"/>
      </fs>
    </SemanticArgument>
  </SemanticPredicate>
</PredicativeRepresentation>
</Sense>
</LexicalEntry>

```

ANEXO 3

DTD PROPORCIONADA POR EL LEXICAL MARKUP FRAMEWORK

```
<?xml version="1.0" encoding="UTF-8"?>
  <!-- LMF DTD -->
  <!-- ##### Core package -->
  <!ELEMENT LexicalResource (feat*, GlobalInformation, Lexicon+, SenseAxis*, TransferAxis*, ContextAxis*)>
  <!ATTLIST LexicalResource
    dtdVersion CDATA #FIXED "16">
  <!-- ##### for data category adornment: feat stands for feature -->
  <!ELEMENT feat EMPTY>
  <!-- att=constant to be taken from the DataCategoryRegistry -->
  <!-- val=free string or constant to be taken from the DCR -->
  <!ATTLIST feat
    att CDATA #REQUIRED
    val CDATA #REQUIRED>
  <!ELEMENT GlobalInformation (feat*)>
  <!ELEMENT Lexicon (feat*, LexicalEntry+, SubcategorizationFrame*, SubcategorizationFrameSet*,
    SemanticPredicate*, Synset*, SynSemCorrespondence*, MorphologicalPattern*, MWEPattern*, ConstraintSet*)>
  <!ELEMENT LexicalEntry (feat*, Lemma, WordForm*, Stem*, ListOfComponents?, RelatedForm*,
    TransformCategory*, Sense*, SyntacticBehaviour*)>
  <!ATTLIST LexicalEntry
    id #IMPLIED
    morphologicalPatterns IDREFS #IMPLIED
    mwePattern IDREF #IMPLIED>
  <!ELEMENT Sense (feat*, Sense*, Equivalent*, Context*, SubjectField*, PredicativeRepresentation*, SenseExample*,
    Definition*, SenseRelation*, MonolingualExternalRef*)>
  <!ATTLIST Sense
    id #IMPLIED
    synset IDREF #IMPLIED>
  <!ELEMENT Definition (feat*, Statement*, TextRepresentation*)>
  <!ELEMENT Statement (feat*, TextRepresentation*)>
  <!ELEMENT TextRepresentation (feat*)>
  <!-- ##### Package for Morphology -->
  <!ELEMENT Lemma (feat*, FormRepresentation*)>
  <!ELEMENT WordForm (feat*, FormRepresentation*)>
  <!ELEMENT Stem (feat*, FormRepresentation*, GrammaticalFeatures*)>
  <!ELEMENT FormRepresentation (feat*)>
  <!ELEMENT RelatedForm (feat*, FormRepresentation*)>
  <!ATTLIST RelatedForm
    targets IDREFS #IMPLIED>
  <!ELEMENT ListOfComponents (feat*, Component+)>
  <!ELEMENT Component (feat*)>
  <!ATTLIST Component
    entry IDREF #REQUIRED>
  <!-- ##### Package for Semantics -->
  <!ELEMENT PredicativeRepresentation (feat*)>
  <!ATTLIST PredicativeRepresentation
    predicate IDREF #REQUIRED
    correspondences IDREFS #REQUIRED>
  <!ELEMENT SemanticPredicate (feat*, Definition*, SemanticArgument*, PredicateRelation*)>
  <!ATTLIST SemanticPredicate
    id #REQUIRED
    semanticTypes IDREFS #IMPLIED>
  <!ELEMENT SemanticArgument (feat*, ArgumentRelation*)>
  <!ATTLIST SemanticArgument
    id #IMPLIED
    semanticTypes IDREFS #IMPLIED>
  <!ELEMENT ArgumentRelation (feat*)>
  <!ATTLIST ArgumentRelation
    targets IDREFS #IMPLIED>
  <!ELEMENT SynSemCorrespondence (feat*, SynSemArgMap*)>
  <!ATTLIST SynSemCorrespondence
    id #REQUIRED>
  <!ELEMENT SynSemArgMap (feat*)>
  <!ELEMENT PredicateRelation (feat*)>
  <!ATTLIST PredicateRelation
    targets IDREFS #IMPLIED>
  <!ELEMENT SenseExample (feat*)>
  <!ATTLIST SenseExample
    id #IMPLIED>
  <!ELEMENT Synset (feat*, Definition*, SynsetRelation*, MonolingualExternalRef*)>
  <!ATTLIST Synset
    id #IMPLIED>
  <!ELEMENT SynsetRelation (feat*)>
  <!ATTLIST SynsetRelation
    targets IDREFS #IMPLIED>
  <!ELEMENT MonolingualExternalRef (feat*)>
  <!ELEMENT SenseRelation (feat*)>
  <!ATTLIST SenseRelation
    targets IDREFS #REQUIRED>
```

ANEXO 4

DTD MODIFICADA PARA EL DICCIONARIO DIDÁCTICO DIGITAL DE LATÍN

```
<?xml version="1.0" encoding="UTF-8"?>
  <!-- LMF DTD -->
  <!--##### Core package-->
<!ELEMENT LexicalResource (feat*,GlobalInformation,Lexicon+)>
<ATTLIST LexicalResource
  dtdVersion CDATA #FIXED "UCM">
  <!--##### for data category adornment: feat stands for feature-
->
<!ELEMENT feat EMPTY>
  <!-- att=constant to be taken from the DataCategoryRegistry -->
  <!-- val=free string or constant to be taken from the DCR-->
<ATTLIST feat
  att CDATA #REQUIRED
  val CDATA #REQUIRED>
<!ELEMENT GlobalInformation (feat*)>
<!ELEMENT Lexicon (feat*, LexicalEntry+)>
<!ELEMENT LexicalEntry (feat*, Lemma, WordForm*, Stem*,MorphologicalPattern*,
Sense*)>
<ATTLIST LexicalEntry
  id ID #IMPLIED
  morphologicalPatterns IDREFS #IMPLIED>
  <!-- El element feature structure (fs) ha sido añadido para representar las
diferentes caracterizaciones(valencias) -->
<!ELEMENT Sense (feat*, SenseExample*, PredicativeRepresentation*, Defini-
tion*,fs*)>
<ATTLIST Sense
  id ID #IMPLIED>
<!ELEMENT Definition (feat*)>
  <!--##### Package for Morphology -->
<!ELEMENT Lemma (feat*)>
<!ELEMENT WordForm (feat*)>
<!ELEMENT Stem (feat*)>
  <!--##### Package for Semantics -->
```

```
<!ELEMENT PredicativeRepresentation (feat*,SemanticPredicate*)>
<!-- predicate and correspondences han sido cambiado de #REQUIRED a #IMPLIED
porque este diccionario no tiene información sintáctica -->
<ATTLIST PredicativeRepresentation
  predicate CDATA #IMPLIED
  correspondences IDREFS #IMPLIED>
<!ELEMENT SemanticPredicate (feat*, Definition*, SemanticArgument*)>
<!-- El atributo label ha sido añadido para codificar la estructura argumental semántica
del verbo -->
<ATTLIST SemanticPredicate
  id ID #REQUIRED
  label CDATA #IMPLIED>
<!-- fs, feature structure (estructura de rasgos), ha sido añadido para codificar las dife-
rentes caracterizaciones argumentales(valencias) del elemento semanticArgument -->
<!ELEMENT SemanticArgument (feat*,fs*)>
<ATTLIST SemanticArgument
  id ID #IMPLIED
  label CDATA #REQUIRED>
<!ELEMENT fs (feat*)>
<!ELEMENT SenseExample (feat*)>
  <!--##### Package for morphological patterns -->
<!ELEMENT MorphologicalPattern (feat*, TransformSet*)>
<!ELEMENT TransformSet (feat*, Process*, GrammaticalFeatures*)>
<!ELEMENT GrammaticalFeatures (feat*)>
<!ELEMENT Process (feat*)>
```

Final del documento ■

ANEXO 5
CODIFICACIÓN DEL DICCIONARIO DIDÁCTICO DIGITAL ELECTRÓNICO EN
EL ESTÁNDAR LMF

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM "LMFdtd.dtd">
<LexicalResource dtdVersion="UCM">
  <GlobalInformation>
    <feat att="title" val="Diccionario Didáctico Latín"/>
    <feat att="URL" val="http://repositorios.fdi.ucm.es/DiccionarioDidacticoLatin"/>
    <feat att="languageCoding" val="ISO639-3"/>
  </GlobalInformation>
  <Lexicon>
    <feat att="language" val="spa"/>
    <LexicalEntry id="id646">
      <feat att="partOfSpeech" val="verb"/>
      <Lemma>
        <feat att="writtenForm" val="peto,is,ere,ii(ivi),itum(3ª)"/>
      </Lemma>
      <Sense id="id646.1">
        <SenseExample>
          <feat att="text" val="Poeta librum ex bibliothecario petit"/>
        </SenseExample>
        <Definition>
          <feat att="text" val="Pedir algo a alguien"/>
        </Definition>
        <PredicativeRepresentation>
          <SemanticPredicate id="marco646.1" label="ISOtrivalent">
            <SemanticArgument label="ARG1">
              <feat att="case" val="nominativeCase"/>
              <fs>
                <feat att="semanticAnimacy" val="animate"/>
                <feat att="semanticType" val="human"/>
              </fs>
            </SemanticArgument>
            <SemanticArgument label="ARG2">
              <feat att="case" val="accusativeCase"/>
              <fs>
                <feat att="semanticAnimacy" val="inanimate"/>
                <feat att="semanticType" val="concrete"/>
              </fs>
              <fs>
                <feat att="semanticAnimacy" val="inanimate"/>
                <feat att="semanticType" val="abstract"/>
              </fs>
            </SemanticArgument>
            <SemanticArgument label="ARG3">
              <feat att="case" val="A/Ab+ABL/Ex+ABL"/>
              <fs>
```

```

        <feat att="semanticAnimacy" val="animate"/>
        <feat att="semanticType" val="human"/>
    </fs>
    </SemanticArgument>
</SemanticPredicate>
</PredicativeRepresentation>
</Sense>
<Sense id="id646.2">
    <SenseExample>
        <feat att="text" val="Captivi romani Belgas petunt"/>
    </SenseExample>
    <Definition>
        <feat att="text" val="atacar"/>
    </Definition>
    <PredicativeRepresentation>
        <SemanticPredicate id="marco646.2" label="ISObivalent">
            <SemanticArgument label="ARG1">
                <feat att="case" val="nominativeCase"/>
            </fs>
            <feat att="semanticAnimacy" val="animate"/>
            <feat att="semanticType" val="human"/>
        </fs>
    </SemanticArgument>
    <SemanticArgument label="ARG2">
        <feat att="case" val="accusativeCase"/>
    </fs>
        <feat att="semanticAnimacy" val="animate"/>
        <feat att="semanticType" val="human"/>
    </fs>
    </SemanticArgument>
    </SemanticPredicate>
</PredicativeRepresentation>
</Sense>
</LexicalEntry>
    <LexicalEntry id="id345">
        <feat att="partOfSpeech" val="noun"/>
        <Lemma>
            <feat att="writtenForm" val="balteum,i(n.)"/>
        </Lemma>
        <Sense id="id345.1">
            <Definition>
                <feat att="text" val="vaina"/>
            </Definition>
            </fs>
            <feat att="semanticAnimacy" val="inanimate"/>
            <feat att="semanticType" val="concrete"/>
        </fs>
        </Sense>
        <Sense id="id345.2">
            <Definition>

```

```

        <feat att="text" val="cinturón"/>
    </Definition>
    <fs>
        <feat att="semanticAnimacy" val="inanimate"/>
        <feat att="semanticType" val="concrete"/>
    </fs>
</Sense>
</LexicalEntry>
<LexicalEntry id="id532">
    <feat att="partOfSpeech" val="adjective"/>
    <Lemma>
        <feat att="writtenForm" val="pulcher,chra,chrum"/>
    </Lemma>
    <Sense id="id532.1">
        <Definition>
            <feat att="text" val="hermoso"/>
        </Definition>
    </Sense>
</LexicalEntry>
<LexicalEntry id="id613">
    <feat att="partOfSpeech" val="adverb"/>
    <Lemma>
        <feat att="writtenForm" val="saepe"/>
    </Lemma>
    <Sense id="id613.1">
        <Definition>
            <feat att="text" val="a menudo, con frecuencia"/>
        </Definition>
    </Sense>
</LexicalEntry>
<LexicalEntry id="id606">
    <feat att="partOfSpeech" val="pronoun"/>
    <Lemma>
        <feat att="writtenForm" val="is,ea,id"/>
    </Lemma>
    <Sense id="id606.1">
        <Definition>
            <feat att="text" val="este, esta, esto"/>
        </Definition>
    </Sense>
</LexicalEntry>
<LexicalEntry id="id614">
    <feat att="partOfSpeech" val="preposition"/>
    <Lemma>
        <feat att="writtenForm" val="in"/>
    </Lemma>
    <Sense id="id614.1">
        <Definition>
            <feat att="text" val="a,hacia(ACU)"/>
        </Definition>
    </Sense>
</LexicalEntry>

```

```
</Sense>
<Sense id="id614.2">
  <Definition>
    <feat att="text" val="en(ABL)"/>
  </Definition>
</Sense>
</LexicalEntry>
<LexicalEntry id="id611">
  <feat att="partOfSpeech" val="conjunction"/>
  <Lemma>
    <feat att="writtenForm" val="atque,ac"/>
  </Lemma>
  <Sense id="id611.1">
    <Definition>
      <feat att="text" val="y"/>
    </Definition>
  </Sense>
</LexicalEntry>
</Lexicon>
</LexicalResource>
```