

**OEG Publication**

Aguado de Cea G, Montiel-Ponsoda E, Ramos JA

***Multilingualidad en una aplicación basada en el conocimiento***

TIMM, Monográfico para la revista SEPLN  
2007  
España

# Multilingüalidad en una aplicación basada en el conocimiento

**Guadalupe Aguado de Cea**  
lupe@fi.upm.es

**Elena Montiel Ponsoda**  
emontiel@delicias.dia.fi.upm.es

**José Ángel Ramos Gargantilla**  
jarg@fi.upm.es

Ontology Engineering Group (OEG) – Universidad Politécnica de Madrid  
Campus de Montegancedo – Boadilla del monte (Madrid)

**Resumen:** La necesidad de dotar de multilingüalidad a sistemas basados en el conocimiento plantea un nuevo reto a la investigación en Inteligencia Artificial. La mayoría de las estrategias adoptadas en la actualidad para la representación de multilingüalidad en bases de conocimiento han optado por soluciones *ad hoc*. Partiendo, pues, de la inexistencia de planteamientos teóricos que hayan analizado todas las posibles soluciones, en este artículo tratamos de cubrir este vacío con un análisis detallado del fenómeno de la multilingüalidad en las aplicaciones basadas en el conocimiento, en concreto en las ontologías. En la primera parte, recogemos brevemente las estrategias de representación de multilingüalidad adoptadas por recursos léxicos existentes. A continuación, nos centramos en los tres niveles que pueden soportar multilingüalidad en las aplicaciones ontológicas, a saber, la interfaz, la representación de conocimiento propiamente dicha y los datos, y detallamos las principales posibilidades de representación. Asimismo se enumeran las ventajas y desventajas de cada representación, señalando las principales implicaciones técnicas de su implementación. Se dedica especial atención a la implantación de la multilingüalidad en el nivel de representación del conocimiento por los numerosos aspectos a considerar según la posición que ésta ocupe. Finalmente, se señala la posibilidad de hibridación de dichos modelos de representación.

**Palabras clave:** multilingüalidad, representación de multilingüalidad en sistemas basados en el conocimiento, ontologías multilingües, recursos léxicos multilingües

**Abstract:** The incipient need to provide knowledge based systems with multilingual information has given rise to a new challenge in Artificial Intelligence. Current strategies used to represent multilingüality in knowledge bases have relied on *ad hoc* solutions. Starting from the premise that there are no theoretical approaches to representing multilingüality and their implications, we aim at filling this gap by presenting a detailed analysis of the multilingüality phenomenon in knowledge based applications, specifically in ontologies. The first part of this article deals with multilingual representation strategies followed by relevant online multilingual lexical resources, applied in the domain of ontologies. The second part is devoted to the analysis of the three identified ontology levels that can support multilingüality, i.e., interface, knowledge representation as such, and data, in which we account for the main options of multilingual information representation in each level. After a detailed survey of the different options, we describe their main advantages and disadvantages, highlighting the technical implications of their implementation, or possible metamodel modifications in the case a new language is added to the ontology. Special attention has been given to multilingual representation at knowledge representation level. Finally, we include a section in which a combination of various multilingual representation models are proposed in what we have called hybrid systems.

**Keywords:** multilingüality, multilingual representation in knowledge based systems, multilingual ontologies, multilingual lexical resources

## 1 Introducción

Desde que la Inteligencia Artificial trabaja con sistemas basados en el conocimiento, los esfuerzos se han centrado en el desarrollo teórico de modelos y en la construcción tanto de motores de inferencia para explotarlos como de aplicaciones basadas en ellos.

La aparición de las ontologías como solución para la compartición de conocimientos consensuados, y la consiguiente inserción de éstas dentro de la Web, dando origen a la Web Semántica, ha supuesto la aparición del problema de la multilingüidad. El interés por los temas de multilingüidad ha estado presente en la comunidad científica desde diferentes perspectivas: búsqueda de información multilingüe, sistemas de búsqueda de respuestas, traducción automática, etc. (Peñas y Gonzalo, 2004). Sin embargo, aunque existen aplicaciones de Web Semántica y sistemas de conocimiento multilingüe, en general, se han adoptado soluciones específicas y particulares dirigidas a una implementación concreta y no como fruto de un estudio conceptual del problema.

El Grupo de Ingeniería Ontológica (OEG) de la Universidad Politécnica de Madrid lleva más de diez años trabajando con ontologías y con aplicaciones basadas en ellas, dentro y fuera de la Web Semántica. En este tiempo ha tratado las aplicaciones y sistemas basados en ontologías en todos sus aspectos teóricos y prácticos, cubriendo todas las actividades del ciclo de vida de las ontologías. Por eso ahora, ante la creciente necesidad de tratar con elementos de conocimiento multilingües, ha estudiado a fondo y de manera teórica la implicación de la multilingüidad dentro de las aplicaciones basadas en el conocimiento.

En este artículo abordamos el análisis de los distintos niveles en que la multilingüidad se puede presentar dentro de estos sistemas, y describimos las posibles soluciones identificadas. Asimismo, para que el estudio sea verdaderamente útil, se completa con la evaluación de las ventajas y desventajas de cada una de las soluciones posibles en cada caso y nivel.

Las figuras que aparecen en este artículo tienen como misión que las argumentaciones se comprendan de una manera más intuitiva con el apoyo de una representación gráfica. Todas ellas son ejemplos docentes pero no se han de

tomar como cánones puesto que no siguen estándares, ni contienen conocimiento consensuado.

## 2 Nociones previas

Antes de entrar en materia propia del artículo, conviene tener presentes y claras una serie de nociones que faciliten la comprensión.

A la hora de representar un conocimiento tenemos tres niveles: datos, modelo y metamodelo. Los datos es el nivel inferior y son los individuos (o instancias) en sí. El modelo es el nivel intermedio y representa la estructura de los datos. El metamodelo es el nivel superior y representa la estructura del modelo.

Veamos un ejemplo de una ontología sobre animales de ficción (información de dominio): Mickey es un individuo, Minnie es otro individuo y Pluto es un tercer individuo. Estos tres individuos son el conjunto de instancias Disney, los datos. Mickey y Minnie son instancias del concepto Ratón ficticio, Pluto es instancia del concepto Perro ficticio y hay una relación son amigos entre Perro ficticio y Ratón ficticio. Ratón ficticio y Perro ficticio son conceptos de la ontología Animales de ficción, el modelo. Concepto y Relación son los componentes de la representación de ontologías, el metamodelo.

Un *mapping* es una relación entre elementos de conjuntos diferentes. Estos conjuntos pueden ser dos ontologías, una ontología y una base de datos, etc.

Siguiendo con el ejemplo anterior, imaginemos que tenemos otra ontología (información de dominio) Animales reales, con un conjunto de instancias llamado Zoo Madrid. En este conjunto de instancias hay una instancia del concepto Ratón real que se llama Madroño. Existe un *mapping* que se llama tienen igual color y que relaciona Madroño y Mickey. Este *mapping* es el dato de información de *mappings*. El modelo de *mappings* viene representado por Elemento1, Elemento2 y Relación. El modelo de la información de *mappings* está al mismo nivel que el modelo de la información de dominio (ontología).

Por otra parte, partiendo de las mismas dos ontologías de antes, podemos tener un *mapping* que relaciona el concepto Ratón ficticio

de la ontología *Animales de ficción* y el concepto *Ratón real* de la ontología *Animales reales*. Este *mapping* se llama toma como ejemplo a. En este caso, los datos de la información de *mappings* (el *mapping* toma como ejemplo a) está al mismo nivel que los modelos de la información del dominio (las ontologías *Animales de ficción* y *Animales reales*). Y, por tanto, el modelo de la información de *mappings* está al mismo nivel que el metamodelo de la información de dominio (el metamodelo de las ontologías).

Hay que tener presente esta separación entre la distinta información que se va a tratar (de dominio, lingüística y de *mappings*), ya que, como en el último caso de *mappings* ejemplificado, no estarán al mismo nivel los datos, modelos o metamodelos.

### 3 Estado de la cuestión

El interés de los investigadores de Inteligencia Artificial por el lenguaje natural se centró primeramente en la comunicación del usuario con la máquina, con lo que las interfaces fueron los primeros componentes de las aplicaciones inteligentes construidas en lenguaje natural (Abascal y Moriyón, 2003). Así lo explica Moreno (2000): “De hecho, el interés por los sistemas de *Natural Language Processing* basados en el conocimiento surgió a raíz de la necesidad de construir interfaces en lenguaje natural para sistemas expertos”. Las ontologías tienen como principales características: el ser un modelo consensuado por la comunidad del dominio, la compartición de su conocimiento y la computabilidad del mismo. Todo esto ha convertido a las ontologías en las representaciones más utilizadas en los nuevos desarrollos y, sobre todo, han pasado a ser la piedra fundamental de la creciente Web Semántica. Al mismo tiempo, debido a su uso generalizado, no pueden escapar al fenómeno de la globalización que exige ahora que la información en lenguaje natural sea además multilingüe.

Las ontologías se enfrentan, pues, al fenómeno de la multilingüidad -que ya se manifiesta de forma patente en diferentes recursos léxicos computacionales- en lo que se ha denominado el proceso de “localización de ontologías”. La localización de un recurso léxico se entiende como el proceso de

adaptación de ese recurso a una comunidad lingüística y cultural concreta. Hasta ahora, el término ‘localización’ se ha venido aplicando al software (Esselink, 2000). Pero, si lo extrapolamos al dominio de las aplicaciones basadas en ontologías, nos encontraremos con que, en un sistema basado en el conocimiento, ese proceso de adaptación tiene que darse en tres niveles diferentes:

- Interfaz
- Representación del conocimiento
- Datos

En la actualidad, pocas son las aplicaciones basadas en ontologías disponibles en más de una lengua natural. La biblioteca de ontologías OntoSelect<sup>1</sup> contiene únicamente 28 ontologías multilingües de las 1.117 que recoge, es decir, menos de un 3%. Además, se ha podido constatar que estas ontologías multilingües muestran, en su mayoría, importantes carencias en aquellas lenguas que no son la original de creación de la ontología.

Como resultado de la búsqueda de ontologías multilingües, conviene destacar la casi nula información pública disponible sobre la representación de la multilingüidad para los diferentes recursos multilingües.

Esto puede ser debido a que, hasta ahora, la representación de la multilingüidad no ha sido objetivo primordial de la investigación sino que el interés se centraba meramente en su existencia. Es decir, que en los recursos y sistemas multilingües la multilingüidad era una propiedad secundaria, más enfocada a la explotación de los mismos que a la conformación del propio recurso. Sólo en casos contados la multilingüidad ha sido tratada con la importancia que le confiere la explotación actual desde la perspectiva de la globalización. No hay que olvidar en ningún momento que las características básicas de las ontologías, que les han convertido en uno de los sistemas de representación más extendido y con mayor auge en los últimos años, son el consenso de su conocimiento y la reutilización del mismo. Y es para lograr esta reutilización y este consenso para lo que es básica la multilingüidad. En los siguientes subapartados recorreremos algunos de los sistemas que incorporan multilingüidad

---

<sup>1</sup> <http://olp.dfki.de/OntoSelect/>

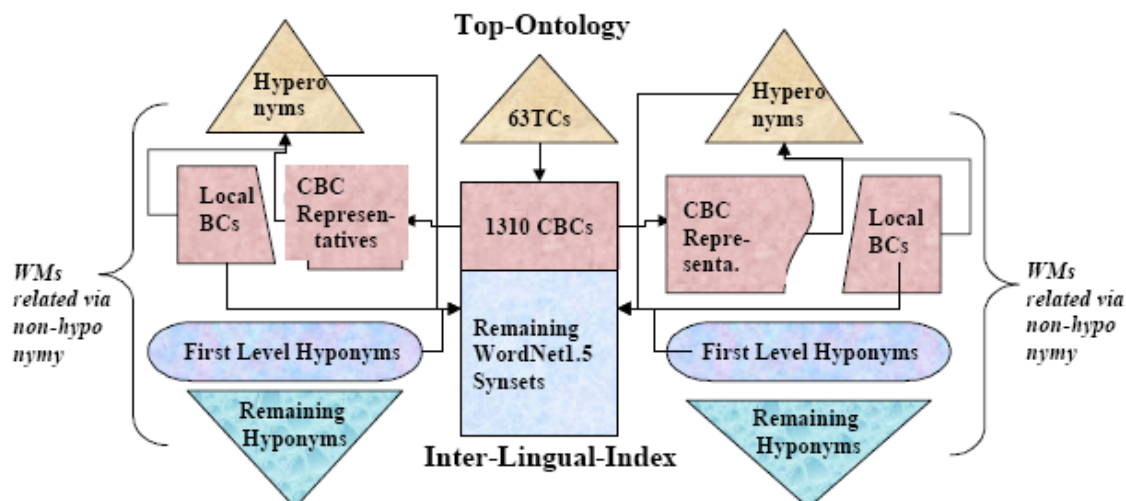


Figura 1: Esquema general de dos *wordnets* mapeados al ILI (Vossen, 2002)

### 3.1 EuroWordNet

EuroWordNet<sup>2</sup> (EWN) (Vossen, 2002 y 2004) fue un proyecto europeo de tres años de duración (1996-1999) y cuyo resultado principal fue un lexicón multilingüe. Este recurso léxico se basaba en la estructura de la base de datos léxica WordNet, desarrollada en la Universidad de Princeton (Miller et al., 1990). El objetivo del proyecto consistía en el desarrollo de un lexicón multilingüe con *wordnets* para ocho lenguas europeas (holandés, español, italiano, inglés, francés, alemán, checo y estonio). El enfoque general seguido para la construcción de EWN se basaba en la reutilización de recursos léxicos existentes en cada lengua.

Al igual que en WordNet, la información sobre nombres, verbos, adjetivos y adverbios se organiza en *synsets* (conjunto de sinónimos, es decir, palabras con la misma categoría gramatical, e intercambiables en ciertos contextos). Los *wordnets* en EWN se consideran “ontologías autónomas monolingües”, que están interconectadas a través de un *Inter-Lingual-Index* (índice interlingüístico o ILI). Dicho ILI consta de una lista de conceptos sin estructura, procedentes de WordNet y relacionados con dos ontologías, una general y otra de dominio (*Top-Ontology* y *Domain-Ontology*, respectivamente). Una selección de conceptos del ILI forma el núcleo de las diferentes ontologías monolingües, que

se amplía y complementa con conceptos específicos de cada lengua, como se muestra en la Figura 1.

### 3.2 GENOMA-KB

La Base de Conocimientos sobre el Genoma Humano (GENOMA-KB<sup>3</sup>) (Cabré et al., 2004a y 2004b) fue un proyecto desarrollado en el Instituto de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra en Barcelona (2001-2003). Este proyecto persigue la construcción de un recurso esencial para la recuperación de información, con control terminológico, en el dominio del genoma humano, en inglés, castellano y catalán. Esta base de conocimiento se divide en cuatro módulos interrelacionados: el módulo ontológico, el módulo terminológico, el módulo del corpus y el módulo de las entidades, como se puede observar en la Figura 2.

El módulo ontológico se desarrolló con la herramienta de gestión terminológica OntoTerm<sup>4</sup>, porque permite la construcción de la ontología y la base de datos terminológica de forma paralela. Esta herramienta, utilizada también en el proyecto OncoTerm<sup>5</sup>, proporciona un núcleo ontológico formado por 21 conceptos básicos (ALL, OBJECT, EVENT, PROPERTY, etc.). A continuación, a estos conceptos básicos se les añadió una lista de 100 conceptos propuestos por expertos en el

<sup>2</sup> <http://www.illc.uva.nl/EuroWordNet/>

<sup>3</sup> <http://genoma.iula.upf.edu:8080/genoma/>

<sup>4</sup> <http://www.ontoterm.com/>

<sup>5</sup> <http://www.ugr.es/~oncoterm/>

dominio, y el resto de conceptos de la ontología se recuperó a partir de información textual especializada, con la ayuda de recursos léxicos. De esta forma, la información lingüística pasó a estar integrada en un modelo separado de la ontología, con lo que se obtenían dos modelos,

el ontológico y el terminológico, relacionados entre sí. Esta distribución posibilita la inclusión de una cantidad considerable de información lingüística en la base terminológica, sin que por ello la ontología tenga que sufrir ningún tipo de modificación.

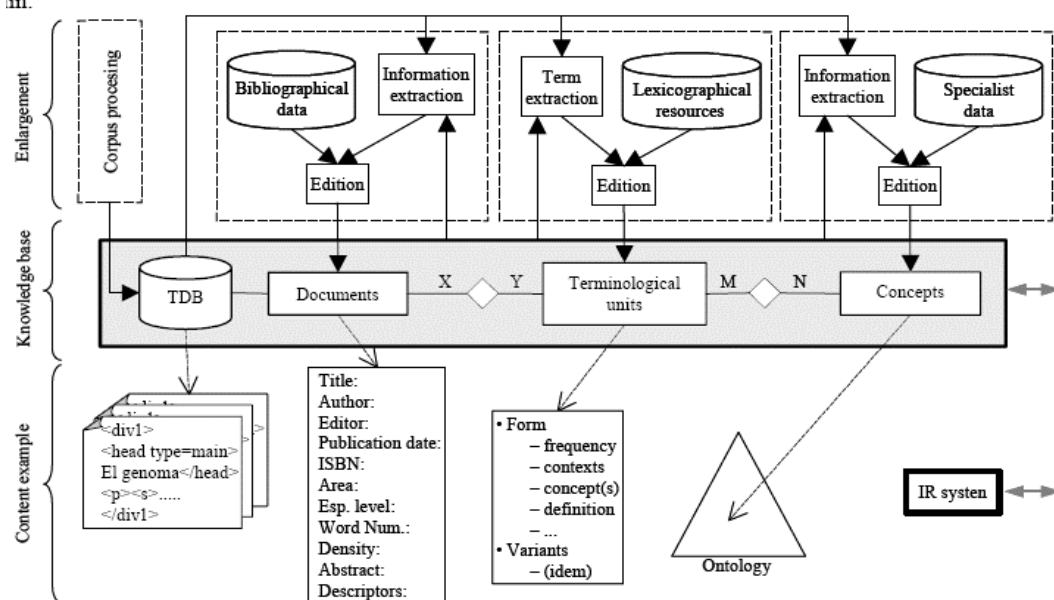


Figura 2: Arquitectura de la base de conocimiento GENOMA-KB (Feliu, Vivaldi y Cabré, 2002)

### 3.3 AGROVOC

La Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) desarrolló el Tesoro AGROVOC (Soergel et al., 2006) a principios de los años 80 con la finalidad de estandarizar el proceso de indexación y recuperación de datos en el dominio de la agricultura, la pesca, la alimentación y otros dominios relacionados como el medio ambiente. AGROVOC se desarrolló por expertos del dominio en inglés, y posteriormente se tradujo al español y al francés. Actualmente está disponible online<sup>6</sup> en 10 lenguas y en proceso de traducción para otras tantas.

La estructura del tesoro actual se centra en una representación conceptual basada en la concepción semántica del inglés, con traducciones al resto de lenguas, con lo que la información lingüística estaría integrada en el modelo de la ontología. Todas y cada una de las lenguas comparten un mismo concepto, y los términos se consideran equivalentes entre sí.

### 4 Niveles de multilingüidad

Como ya se ha apuntado, la multilingüidad, en una aplicación sobre un sistema basado en el conocimiento, puede darse en tres niveles diferentes:

- Interfaz
- Representación del conocimiento
- Datos

Una vez identificados los niveles en los que se manifiesta la multilingüidad, en los apartados siguientes se analizan cada uno de ellos y se presentan diversas propuestas sobre cómo incorporar la multilingüidad en las ontologías, destacando las ventajas y desventajas de las soluciones propuestas.

### 5 Interfaz

Aunque el análisis de la interfaz ya haya sido tratado por otras áreas más cercanas a la interacción hombre-máquina, queremos dejar patente la idea de que la multilingüidad de una interfaz puede o no conllevar la multilingüidad del resto de la aplicación (el sistema basado en el conocimiento subyacente).

<sup>6</sup> <http://www.fao.org/aims/>

Por ello creemos que es conveniente mencionarlo en este estudio para dar una idea más global del alcance de la multilingüidad en todas sus facetas.

La multilingüidad en la interfaz puede darse en dos visualizaciones diferentes: visualización de mensajes y visualización de contenido.

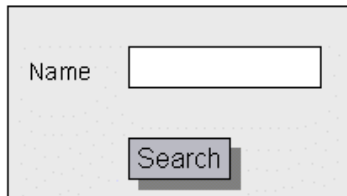


Figura 3: Ejemplo de mensajes monolingües

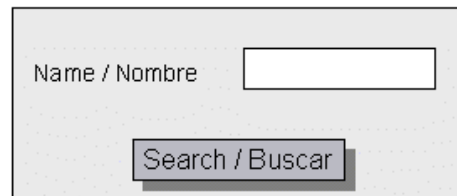


Figura 4: Ejemplo de mensajes multilingües simultáneos



Figura 5: Ejemplo de mensajes multilingües no simultáneos

En la Figura 3 aparece un ejemplo de interfaz con mensaje monolingüe en inglés. En la Figura 4 se muestra un ejemplo de interfaz con mensajes multilingües simultáneos, es decir, con más de una lengua a la vez (inglés y español en el ejemplo).

Finalmente, en la Figura 5 se encuentra un ejemplo de interfaz con mensajes multilingües no simultáneos: se puede ver en varios idiomas pero sólo uno cada vez.

## 5.2 Contenido

La visualización de contenido de forma multilingüe tiene dos posibles vías de resolución: (a) la base de conocimientos es multilingüe, o (b) la base de conocimientos es monolingüe y se usa un sistema de traducción.

En el primer caso, la aplicación sólo realizará una consulta a la base de conocimientos para obtener el contenido que mostrará la interfaz en el idioma requerido.

## 5.1 Mensajes

Los mensajes de una interfaz pueden ser monolingües o multilingües, y, dentro de los segundos, pueden ser multilingües simultáneamente o no.

En el segundo caso, la aplicación consultará la base de conocimientos y luego utilizará un recurso multilingüe (diccionario, glosario, base de datos, etc.) para traducir el contenido y mostrarlo a través de la interfaz.

En ambos casos, la interfaz resultante sería similar y podría seguir la misma política de simultaneidad o no simultaneidad que se ha mencionado en la visualización de mensajes. Sin embargo, es importante tener en cuenta el que la base de conocimientos sea multilingüe o no para los tiempos de respuesta.

## 5.3 Ventajas y desventajas

Cada una de las soluciones mencionadas presenta ventajas e inconvenientes. La visualización de mensajes de forma simultánea implica que la incorporación de nuevos idiomas requiere la modificación del código de visualización existente. Ahora bien, la visualización no simultánea no implica la modificación de todo el código existente, sino

la ampliación del número de interfaces y la consiguiente modificación de la parte de interfaz donde se codifican las opciones (banderas en el ejemplo de la Figura 5).

En cuanto a los tiempos de respuesta en las consultas, podemos encontrarnos con dos casos: (a) si se dispone de una base de conocimientos multilingüe, el tiempo de obtención de los contenidos que se muestran es el tiempo de respuesta de la base de conocimientos. Esto se debe a que se ha conferido multilingüidad a la base de conocimientos en tiempo de diseño. Sin embargo, para el caso (b) cuando la base de conocimientos es monolingüe, la traducción se realiza en tiempo de ejecución, y el tiempo de obtención de los contenidos es el tiempo de respuesta de la base de conocimientos más el tiempo de respuesta del recurso multilingüe que proporciona la traducción. En este segundo caso, pueden surgir ciertos problemas adicionales como, por ejemplo, la desambiguación, que, en el primer caso, se habrá resuelto en tiempo de diseño. Esto supone que la aplicación tendrá unos tiempos de respuesta diferentes en cada solución.

## 6 Representación del conocimiento

La multilingüidad en una representación del conocimiento se tiene que abordar a tres niveles: **Información**, **Modelización** y **Realización**. En cuanto al primero, la **Información**, se refiere al dato de que la representación del conocimiento se encuentra expresada en varios idiomas. La **Modelización** es la representación de los componentes que darán multilingüidad a la representación del conocimiento. Finalmente, la **Realización** es la expresión lingüística real de la multilingüidad en la representación del conocimiento.

Tomemos un ejemplo sencillo para la mejor identificación de los niveles: un libro. La información serían los datos del libro (autor, título, año de publicación, país), la modelización sería similar a “páginas impresas consecutivas entre dos pastas y unidas por un lado común llamado lomo” y la realización sería el tomo físico. Igualmente, para el caso de las ontologías, la información está compuesta por los metadatos de la ontología (autor, fecha de creación, etc.), la modelización es la metaontología que describe los componentes de la ontología y sus relaciones (conceptos que tienen atributos y se unen por relaciones...) y la

realización sería una ontología propiamente dicha (ontología de coches, de libros, de profesiones, etc.).

Veamos cada una de ellas y las implicaciones de cada uno de los aspectos de la multilingüidad ejemplificando con el caso concreto de las ontologías.

### 6.1 Información

La información sobre la multilingüidad de una representación del conocimiento forma parte de los metadatos de dicha representación. Así, junto con metadatos como el autor o la fecha de creación de la representación del conocimiento, aparecerá la información sobre multilingüidad.

En principio, tan sólo sería necesario saber las lenguas en las que la representación (en este caso en particular, la ontología) está disponible. Y decimos en principio porque, como se verá en la parte de modelización, también será necesario expresar los componentes de la representación (de la ontología) que tienen esta multilingüidad.

Para ejemplificar esto, tomemos como base una representación de ontología en la que se almacenan como metadatos el autor, la fecha de creación y la dirección URI donde se encuentra. Con la finalidad de ofrecer una realización concreta es recomendable introducir esta información en un estándar o similar, como es el vocabulario de metadatos sobre ontologías OMV (Hartman y Palma, 2006).

Para almacenar la información sobre multilingüidad, existen dos posibles soluciones:

- 1) **Concepto lingüístico:** Se dispone del concepto `Language`, con su correspondiente caracterización, y se asocia con el concepto `Ontology`. Esta relación lleva asociada la información relativa a los componentes que están en dicha lengua. En la Figura 6 se han incluido unas etiquetas bajo el concepto `Language` a modo de ejemplo, teniendo en cuenta que una lengua tiene diversos componentes (alfabético, con caracteres cirílicos o latinos, por ejemplo, morfo-sintáctico, léxico y semántico). La multilingüidad pasaría por tener varias relaciones para las diferentes lenguas.

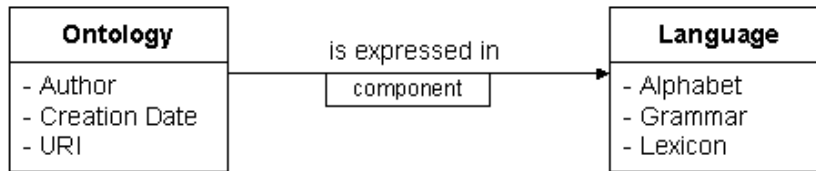


Figura 6: Información de multilingüedad con concepto que representa la lengua (Language)

2) **Modificación del concepto Ontology:**

Se dispone de un metadato con valor múltiple y ese metadato se representa por medio de tuplas de la lengua y del componente que está en dicha lengua.

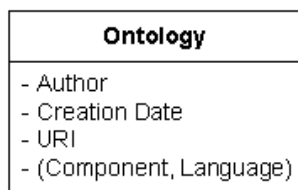


Figura 7: Información de multilingüedad modificando el concepto Ontology

En la Figura 6 se puede ver un ejemplo correspondiente a la primera solución y en la Figura 7 se presenta un ejemplo correspondiente a la segunda solución.

Cualquiera de las dos soluciones resolvería la expresión de información sobre la multilingüedad. Además, la representación de la información de multilingüedad es independiente del enfoque que se siga para la representación del contenido.

**6.1.1 Ventajas y desventajas**

Veamos ahora las implicaciones de cada una de las dos soluciones. La primera opción (Figura 6) requiere la instanciación del concepto Language con información del dominio lingüístico, lo que no es un trabajo trivial. Además, sistemas que traten con relaciones que lleven asociadas información semántica (en este caso *is expressed in* lleva asociado *component*) son difíciles de encontrar. Para superar este handicap se puede optar por una reificación, con lo que toma entidad de concepto Component y se establecen dos relaciones: *Composed by* entre Ontology y Component y *is expressed in* entre Component y Language.

Por otra parte, la segunda solución (Figura 7) supone una pérdida de información lingüística sobre la lengua. Por ello, resulta imprescindible realizar una evaluación previa de la información lingüística necesaria para el sistema en concreto. Sin embargo, si se quiere prescindir de este tipo de información, resulta la solución más sencilla, porque tiene menos información lingüística y es más fácil de implementar.

Ahora bien, en el caso en que se hiciera una modificación como la propuesta en la opción 2, pero sólo con un atributo en que se indicase la lengua (Language) en vez de la tupla (Component, Language), se perdería la posibilidad de que un componente de la conceptualización pudiera ser multilingüe o monolingüe independientemente del resto de componentes.

**6.2 Modelización**

La segunda de las perspectivas que hemos mencionado es la **modelización**. En este ámbito, la modelización de la multilingüedad en ontologías admite tres posibles soluciones: (a) ampliación con información lingüística del metamodelo de ontologías, (b) agregación de un modelo de información lingüística que se relacionará con el metamodelo de ontologías para representar la multilingüedad, y (c) utilización de un modelo de *mappings* para relacionar ontologías monolingües y obtener así multilingüedad.

En los tres subapartados siguientes se muestran diversos ejemplos de modelizaciones que ilustran las soluciones propuestas. En ningún caso se ha pretendido, por razones de espacio, una exhaustividad total en las posibles variantes que permita cada solución, sino que el objetivo es, más bien, facilitar al lector la comprensión de las teorías expuestas, teniendo en cuenta que los ejemplos no son reales pero sí extrapolables.

## 6.2.1 Metamodelo de ontologías modificado

En este caso, dentro del metamodelo de la ontología se incluye la modelización de la información de la multilingüedad. Hay que tener en cuenta que esta información se puede precisar a nivel de componentes (clase, propiedad, axioma, etc.) dentro de la ontología. Así, las modificaciones en el metamodelo de la ontología pueden ser más o menos profundas, según el número de componentes al que afecte.

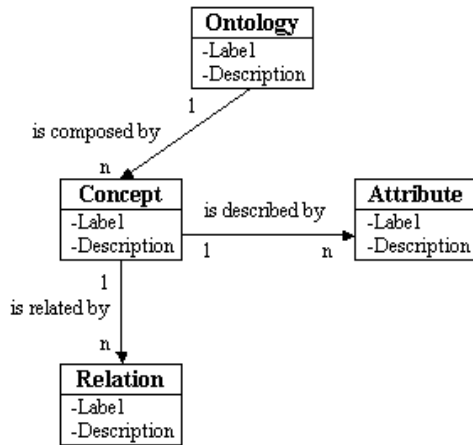


Figura 8: Ejemplo de metamodelo de ontologías

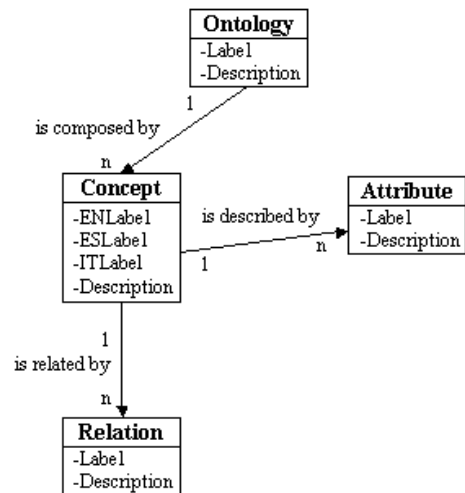


Figura 9: Ejemplo de ampliación con multilingüedad de etiquetas en conceptos, con idiomas previamente fijados

En la Figura 8 se puede ver un ejemplo simplificado de un metamodelo de ontologías siguiendo el paradigma de marcos.

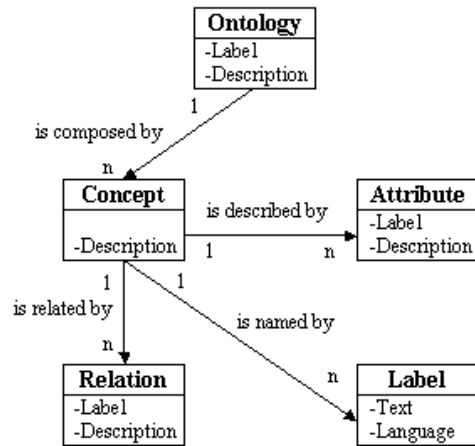


Figura 10: Ejemplo de ampliación con multilingüedad en atributos de los conceptos

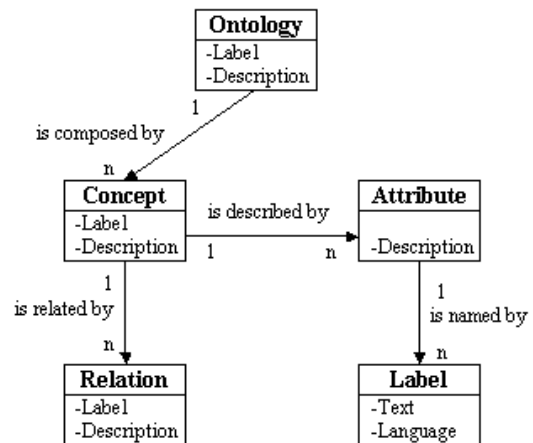


Figura 11: Ejemplo de ampliación con multilingüedad en etiquetas relacionadas con los atributos

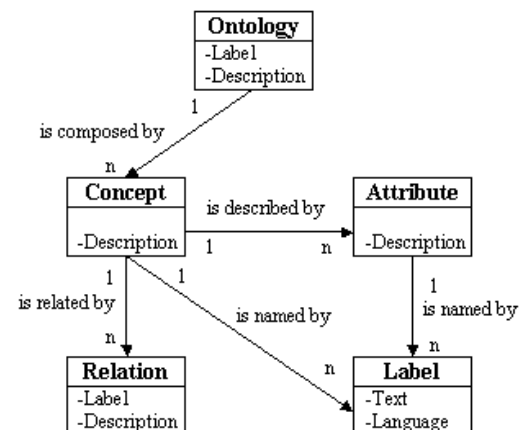


Figura 12: Ejemplo de ampliación con multilingüedad en etiquetas relacionadas con los conceptos y atributos

En las siguientes figuras presentamos varias modificaciones de ese metamodelo: con multilingüidad en etiquetas de conceptos (Figura 9 y Figura 10); con multilingüidad en etiquetas de atributos (Figura 11); con multilingüidad en etiquetas tanto de conceptos como de atributos (Figura 12) y, finalmente, incluyéndola en etiquetas de conceptos, de atributos y de relaciones (Figura 13). En realidad, puede haber más combinaciones, pero como se ha mencionado antes, no se pretende recoger aquí una relación exhaustiva de todas las posibles.

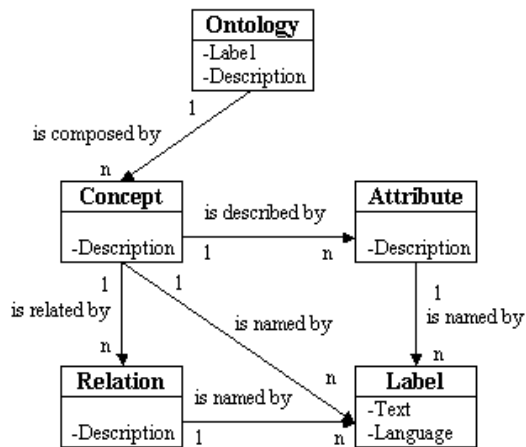


Figura 13: Ejemplo de ampliación con multilingüidad en etiquetas relacionadas con tres componentes

## 6.2.2 Metamodelo de ontología y Modelo de recurso lingüístico

La creación de un modelo para la información multilingüe utilizando un modelo de recurso lingüístico permite un aumento de las posibilidades de inclusión de información acerca del idioma de expresión de los componentes de las ontologías. El metamodelo de multilingüidad constará, pues, del metamodelo de la ontología, el modelo del recurso lingüístico y los enlaces entre el metamodelo de ontologías y el modelo de recurso lingüístico.

La Figura 14 ilustra una posible solución para un modelo de etiquetas (modelo del recurso lingüístico), que no modifica el metamodelo de ontologías que aparece en la Figura 8. En este modelo de recurso lingüístico la etiqueta viene caracterizada por un texto (Text), por ejemplo 'vino'.

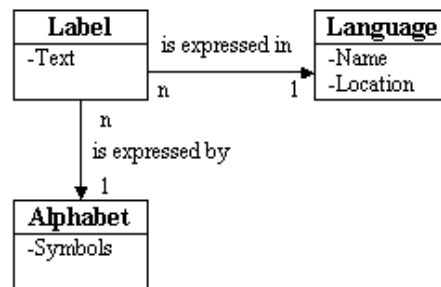


Figura 14: Ejemplo de modelo de etiquetas

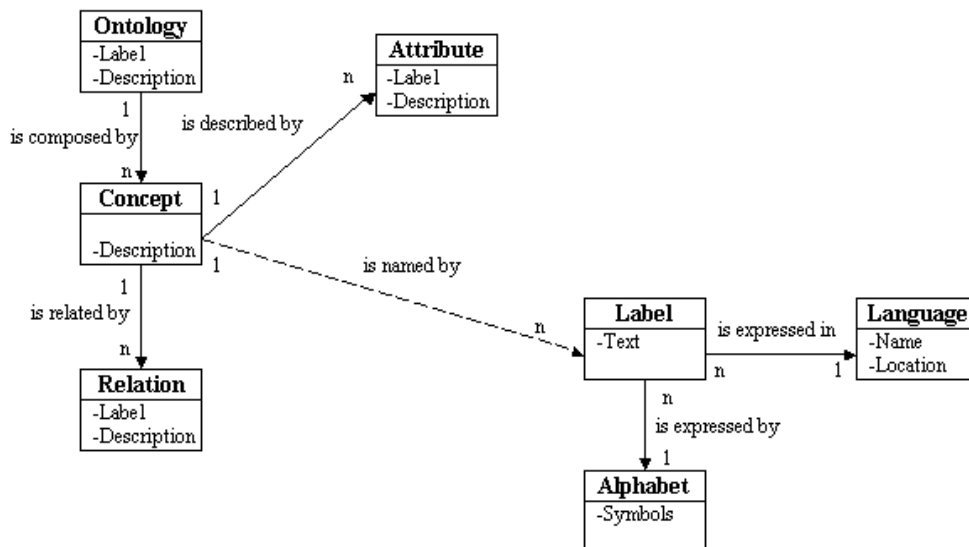


Figura 15: Ejemplo de metamodelo de multilingüidad para etiquetas de conceptos

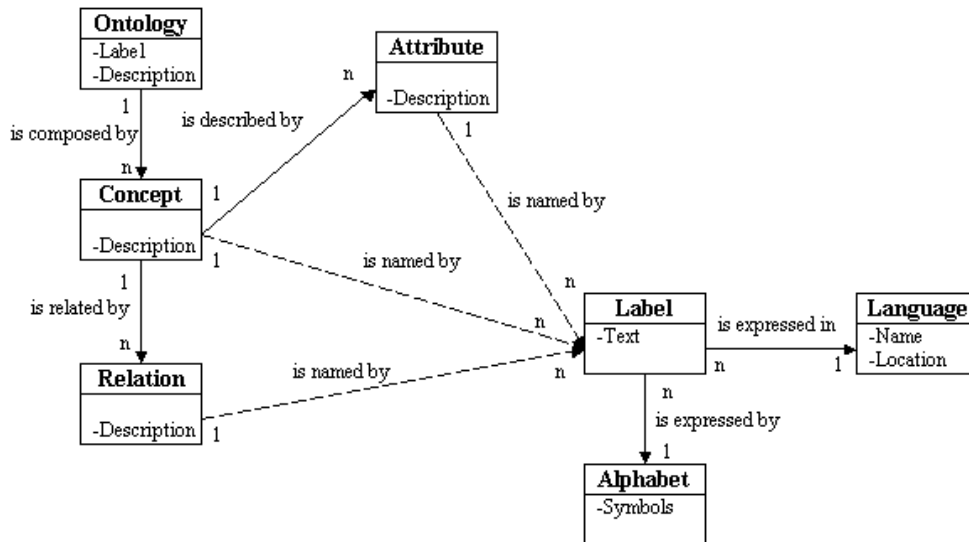


Figura 16: Ejemplo de metamodelo de multilingüidad para etiquetas de conceptos, de atributos y de relaciones

Se puede utilizar un modelo de recurso lingüístico con información específica para etiquetas o utilizar uno que modele más información aunque no sea necesaria para este uso.

Al igual que en la sección anterior, existen múltiples posibilidades a la hora de combinar las etiquetas y los componentes de la ontología para dotar a estos últimos de multilingüidad. En la Figura 15 se muestra un ejemplo de metamodelo de multilingüidad para etiquetas de conceptos. En la Figura 16 se muestra un metamodelo de multilingüidad para etiquetas de conceptos, de atributos y de relaciones.

### 6.2.3 Metamodelo de ontología y Modelo de mappings

Otra forma de conseguir que el modelo sea multilingüe es mediante la combinación de un modelo de *mappings* con el metamodelo de ontologías. En este caso, la multilingüidad se consigue con ontologías del mismo dominio pero en diferente lengua (todas monolingües) y *mappings* de equivalencia entre los términos de las diferentes lenguas.

Un ejemplo de este metamodelo se muestra en la Figura 17.

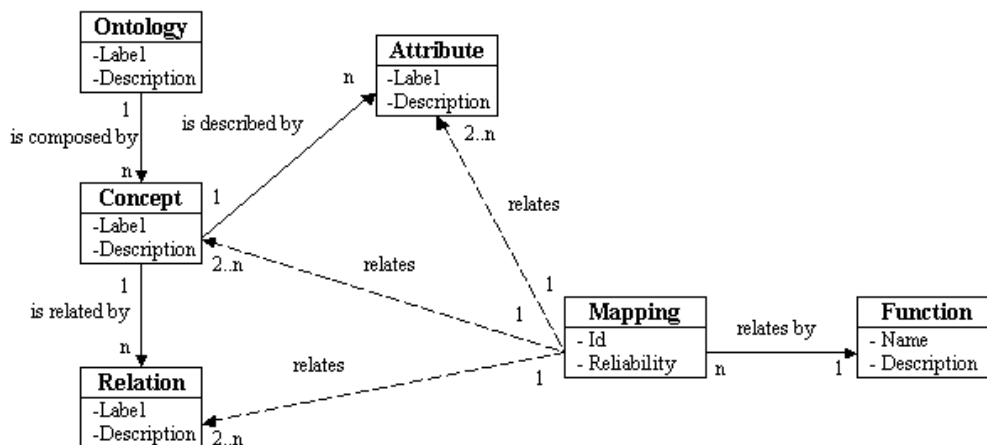


Figura 17: Ejemplo de metamodelo de multilingüidad con *mappings* para etiquetas de conceptos, de atributos y de relaciones

La finalidad de este metamodelo no es ser un metamodelo de multilingüidad, aunque con él se pueda representar la multilingüidad si atendemos sólo a los *mappings* que relacionan componentes a través de la función “translate” o “equivalence label”, por ejemplo.

Así pues, la multilingüidad puede presentarse de múltiples formas, dependiendo tanto del modelo de *mappings* que se utilice como del metamodelo de ontologías empleado, y de las relaciones que entre ambos se establezcan. Por ejemplo, si el *mapping* no es de aridad limitada a dos, es decir, relaciona más de dos elementos, se podría hacer un *mapping* múltiple entre todos los términos relacionados y de esta forma se evitaría un alto número de *mappings* binarios entre varias ontologías monolingües.

Este metamodelo es el menos intuitivo a este nivel de abstracción, pero quedará suficientemente claro al mostrar las diferentes realizaciones posibles que se derivan de él, y que se analizan en la sección 6.3.3.

### 6.3 Realización

La tercera de las perspectivas apuntadas más arriba es la realización de la multilingüidad, que está unida irremediamente a la modelización de la misma. Estas acciones son las dos únicas, de las que aparecen en este documento, que tienen una estrecha interrelación, de manera que la una depende de la otra, ya que la realización no es más que la instanciación de la modelización, es decir, el modelo será la instancia del metamodelo.

#### 6.3.1 Información lingüística dentro de la ontología

En este primer caso, la realización se corresponde con el enfoque en que se amplía el metamodelo de la ontología para introducir la información lingüística que permitirá la multilingüidad (5.2.1.).

Según sea la modificación del metamodelo que se haya seguido obtenemos diversas realizaciones. Así, la Figura 18 muestra un ejemplo de realización del metamodelo de ejemplo mostrado en la Figura 9.

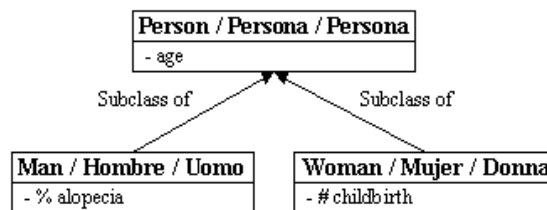


Figura 18: Ejemplo de modelo basado en un metamodelo de ontologías ampliado para etiquetas multilingües en conceptos

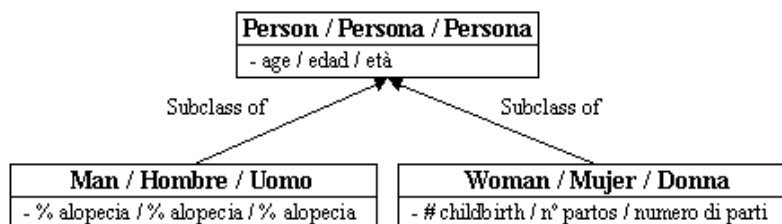


Figura 19: Ejemplo de modelo basado en metamodelo de ontologías ampliado para etiquetas multilingües en conceptos y atributos

En el ejemplo de la Figura 18 se observa que existen tres etiquetas que identifican a cada concepto. Sin embargo, tal como se puede ver en la Figura 9, los atributos no tienen multilingüidad al sólo disponer de una etiqueta.

Si la ampliación del metamodelo de ontologías afectase de igual forma a los atributos, dotándoles de las mismas tres etiquetas, se obtendría un modelo como el reflejado en la Figura 19.

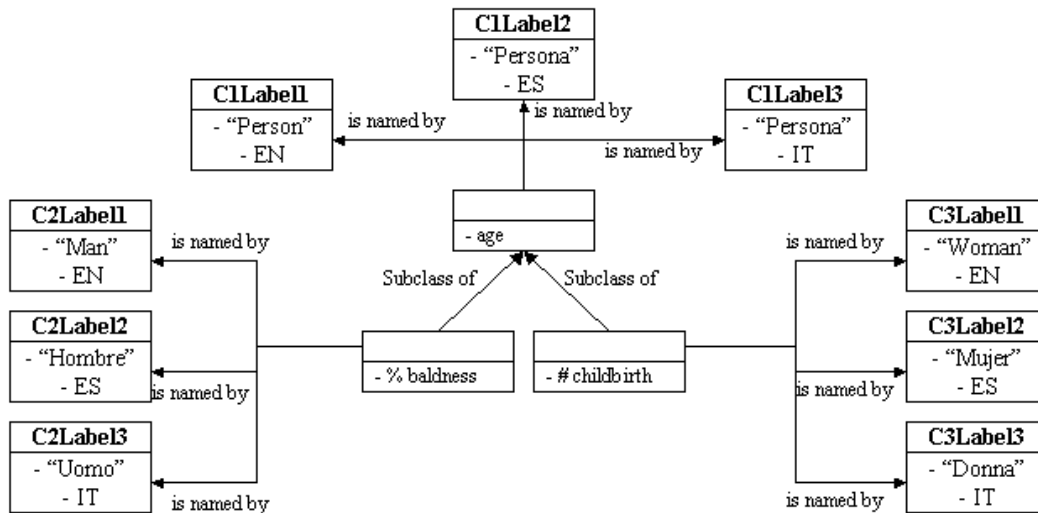


Figura 20: Ejemplo de modelo basado en metamodelo de ontologías ampliado con etiquetas asociadas a conceptos

Como se menciona en la sección anterior, se pueden realizar diferentes ampliaciones en el metamodelo de ontología para incluir la información lingüística que permite representar la multilingüidad. En la Figura 20 se puede ver un ejemplo de realización correspondiente al metamodelo que aparece en la Figura 10.

### 6.3.2 Información lingüística fuera de la ontología

Cuando la información lingüística toma entidad y se convierte en modelo separado de la ontología en sí, estamos ante la posibilidad de utilizar metamodelos de multilingüidad que engloben el metamodelo de ontologías, el modelo del recurso lingüístico y las relaciones entre estos dos.

Ahora bien, puesto que el modelo del recurso lingüístico puede ser una base de datos

relacional, una base terminológica, un lexicón multilingüe, un tesoro multilingüe, etc., el número de emparejamientos ‘modelo de ontologías’ – ‘conjunto de instancias del recurso lingüístico’ se dispara. Además, puede darse el caso de que se utilice un metamodelo de ontologías que almacene información lingüística de una lengua en sus etiquetas o bien que las etiquetas de la ontología no pertenezcan a ninguna lengua.

Para ejemplificar estas posibilidades se muestran a continuación dos ejemplos reflejados en la Figura 21 y la Figura 22. En la primera de ellas la ontología no contiene información lingüística y toda la información que conforma la multilingüidad está almacenada en el recurso lingüístico, que, para facilitar la comprensión, se ha optado por representar de una manera simplificada.

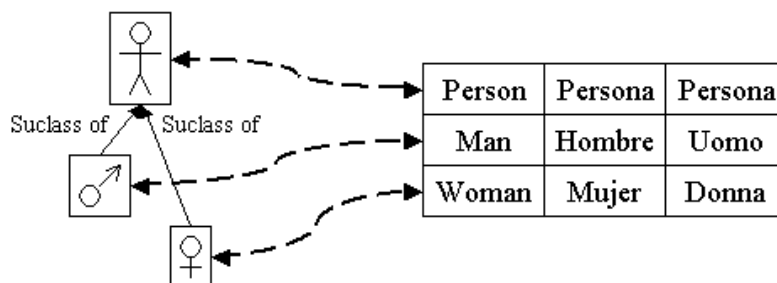


Figura 21: Ejemplo de modelo basado en un metamodelo de multilingüidad con metamodelo de ontologías “alingüe” y modelo de recurso lingüístico

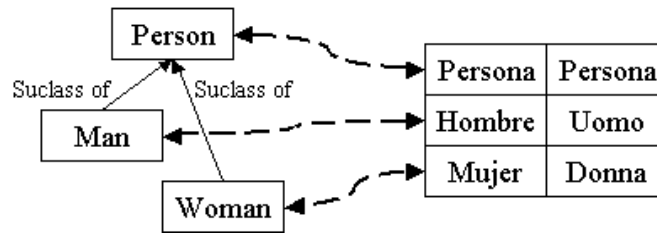


Figura 22: Ejemplo de modelo basado en un metamodelo de multilingüedad con metamodelo de ontologías monolingües y modelo de recurso lingüístico

En cambio, en la Figura 22, podemos ver que la información correspondiente a una de las lenguas está almacenada en forma de etiquetas en la ontología.

Para ilustrar mejor estos casos, cabe mencionar, como ejemplo de un sistema real que sigue uno de estos modelos, el sistema Genoma-KB (Cabré et al., 2004a y 2004b). Como se ha visto en el estado de la cuestión, este sistema consta de una ontología con etiquetas que no se atienen a ninguna lengua y una base terminológica en donde se encuentra la información lingüística que dota de multilingüedad a la base de conocimientos.

Aunque en los ejemplos mencionados en este apartado se toma, por razones de claridad, la multilingüedad de las etiquetas de los conceptos, no hay que olvidar que pueden existir también otros sistemas con multilingüedad para los diferentes componentes de las ontologías (atributos, relaciones, etc.) y todas sus combinaciones.

### 6.3.3 Uso de *mappings* con función lingüística

Como ya se ha apuntado en el epígrafe 6.2.3, los modelos basados en un metamodelo de multilingüedad que conste de un metamodelo de ontologías enlazado a un modelo de *mappings* pueden tener varias realizaciones diferentes según sea el modelo de *mappings*. Así, aunque se siga el mismo metamodelo de multilingüedad, se pueden obtener diferentes realizaciones siempre que las herramientas lo permitan.

Veamos alguna de estas posibilidades. Tomando como base el metamodelo del ejemplo que aparece en la Figura 17, se muestran a continuación tres realizaciones diferentes: con *mappings* binarios entre conceptualizaciones monolingües en forma de grafo ortogonal (Figura 23), con *mappings* n-arios entre conceptualizaciones monolingües (Figura 24) y con *mappings* binarios entre las conceptualizaciones monolingües y una conceptualización intermedia en forma de grafo radial (Figura 25).

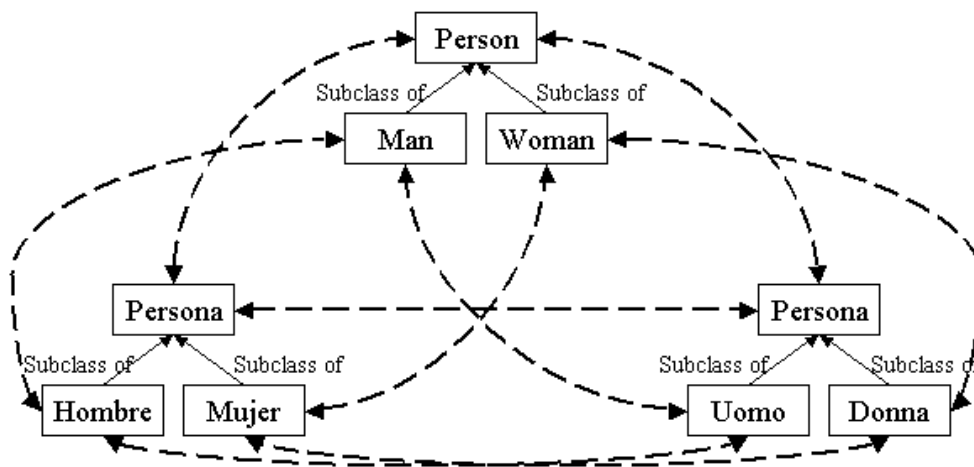


Figura 23: Ejemplo de modelo basado en metamodelo de multilingüedad con *mappings* binarios en grafo ortogonal

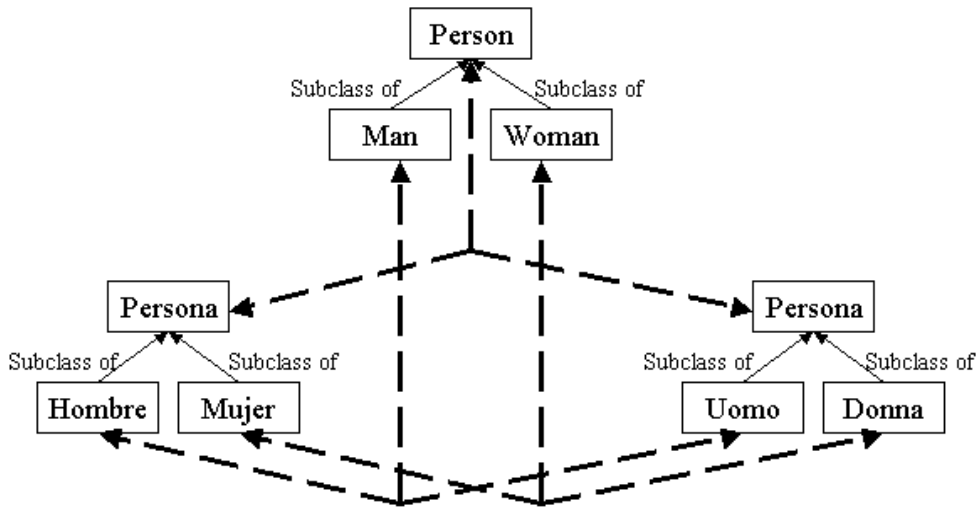


Figura 24: Ejemplo de modelo basado en metamodelo de multilingüidad con *mappings* n-arios

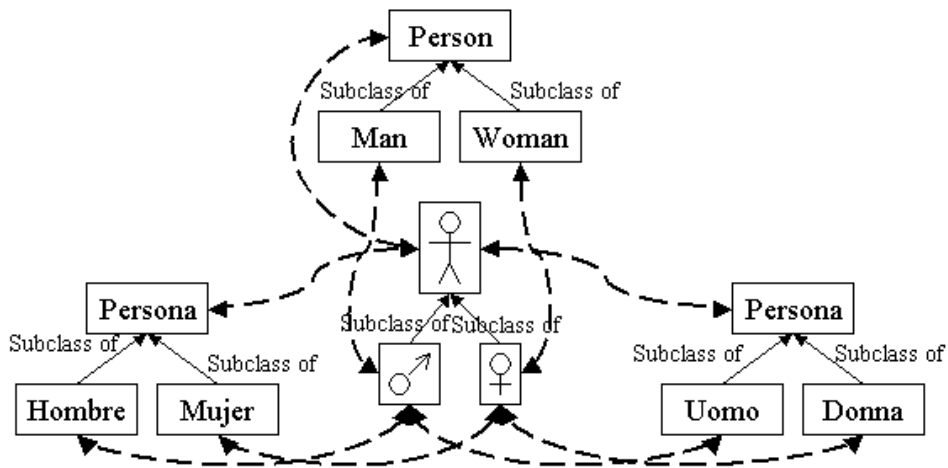


Figura 25: Ejemplo de modelo basado en metamodelo de multilingüidad con *mappings* binarios en grafo radial

Al igual que se ha explicado en el apartado anterior, todos los modelos de ejemplo correspondientes a esta sección se están tomando como multilingües a nivel de etiquetas de concepto, pero, siempre que lo admita el metamodelo de *mappings*, se puede representar multilingüidad del componente o conjunto de componentes que se desee.

En el caso del modelo de la Figura 24 se pueden ver *mappings* de aridad tres, pues la aridad ha de ser igual al número de ontologías monolingües del sistema.

La gran desventaja de esta solución es que ningún sistema contempla *mappings* de aridad superior a dos, salvo soluciones *ad hoc* para operaciones concretas. Sin embargo, para expresar esta misma información de otro modo

con *mappings* binarios, se puede recurrir a la construcción de una conceptualización sin información lingüística que haga las veces de solución reificada de la anterior.

La realización que sigue este último modelo, ejemplificado en la Figura 25, es similar a la utilizada por EuroWordNet (Vossen, 2002 y 2004), analizada en el estado de la cuestión, donde la conceptualización intermedia es una Interlingua (ILI) que no pertenece a ninguna de las lenguas que reconoce el sistema. En EuroWordNet las ontologías no son meras traducciones tal y como las tomamos para la realización, sino conceptualizaciones propias de cada lengua.

## 6.4 Comparativa de las soluciones

Para poder elegir entre una de las soluciones que aquí se han expuesto, es necesario disponer de unos criterios básicos sobre los que se apoya la decisión que se vaya a tomar, ponderando para cada caso particular, la conveniencia de cada solución.

Para facilitar esta tarea, hemos confeccionado una tabla en donde se reflejan estos datos.

Las características y los criterios que se han tomado como referencia son:

(a) **Número de metamodelos y modelos** del metamodelo de ontologías multilingües

(b) **Número de modelos y conjuntos de instancias** del modelo multilingüe

(c) **Número de razonadores (R)**: depende del número de modelos y conjuntos de instancias. Nos encontraremos 3 tipos de razonadores:

- Razonadores sobre ontologías (OR)
- Razonadores sobre *mappings* (MR)
- Razonadores sobre recurso lingüístico (LRR)

(d) **Complejidad de las consultas**: el nivel de complejidad de una consulta se infiere del número de modelos y conjuntos de instancias del modelo multilingüe y del número de componentes de ellos que han de ser consultados para obtener el resultado. Así, hemos identificado 5 niveles de complejidad diferentes para nuestro propósito, que estarán comprendidos en un rango de 1 (la más baja complejidad) a 5 (la más alta complejidad).

| Nivel de complejidad | 1 modelo o conjunto de instancias | 2 modelos o conjuntos de instancias |
|----------------------|-----------------------------------|-------------------------------------|
| 1 componente         | 1                                 | 3                                   |
| 2 componentes        | 2                                 | 4                                   |
| 3 componentes        | -                                 | 5                                   |

(e) **Complejidad de adición de una nueva lengua**: el grado de complejidad que nos encontraremos cuando queramos añadir una nueva lengua a nuestra base de conocimiento depende de los elementos que haya que modificar, pudiendo estar entre 1 (la menor) y 3 (la mayor):

| Objetivo de la modificación         | Nivel de complejidad |
|-------------------------------------|----------------------|
| Metamodelo                          | 3                    |
| n modelos o conjuntos de instancias | 2                    |
| 1 modelo                            | 1                    |

(f) **Complejidad del mantenimiento de la consistencia**: este mantenimiento depende del número de modelos y conjuntos de instancias del modelo multilingüe. Cuantos más modelos sean necesarios (más gestores de esos modelos), mayor dificultad habrá en mantener la consistencia.

| Complejidad                 | Nivel de complejidad |
|-----------------------------|----------------------|
| c (constante)               | 1                    |
| n (lineal)                  | 2                    |
| n <sup>2</sup> (cuadrática) | 3                    |

(g) **Disponibilidad real** de herramientas y sistemas existentes: este criterio se refiere al número de componentes (gestores, razonadores, etc.) con los que se puede contar realmente.

Una vez establecidos los criterios de evaluación, veamos cómo se aplican a los ejemplos de los 3 metamodelos multilingües presentados en la sección anterior.

### 1) Metamodelo de ontologías modificado.

En los ejemplos recogidos en la Figura 9, la Figura 10, la Figura 11, la Figura 12 y la Figura 13, pueden verse los modelos multilingües fruto de la inserción de la información lingüística dentro de una clase del metamodelo de las ontologías.

En estos modelos, el número de modelos (a) es 1, y, por tanto, el número de razonadores (c) es también 1. El nivel de complejidad de la consulta (d) reflejado en los modelos de la Figura 18 y la Figura 19 es también 1, porque para obtener el resultado se necesita consultar un modelo y un único componente, lo que representa una gran ventaja. Sin embargo, para los modelos multilingües provenientes de incorporar información lingüística como clases dentro del metamodelo de ontologías (Figura 20), la complejidad de la consulta aumenta a 2, ya que en las resoluciones se requiere de dos componentes (las clases con información conceptual y las clases con información lingüística).

Por otra parte, para añadir una nueva lengua (e), nos encontramos con algunas limitaciones, porque el número de lenguas en el metamodelo multilingüe de la Figura 10 se ha establecido en tiempo de diseño y, por lo tanto, es fijo. Aumentar el número de lenguas supondría una gran desventaja, porque sería necesario cambiar el metamodelo de toda la base de conocimiento. Por esto es por lo que la complejidad de esta tarea sería 3. En cambio, para el caso de añadir otras lenguas en los casos reflejados en la Figura 10, Figura 11, la Figura 12 y la Figura 13 no haría falta cambiar el metamodelo, con lo que la complejidad de esta tarea sería 1.

**2) Metamodelo de ontologías + Modelo de recurso lingüístico.** Veamos ahora el caso de metamodelos de ontologías enlazados con modelos de recursos lingüísticos (Figura 15 y Figura 16), que se corresponden con los dos ejemplos de modelos multilingües, representados en la Figura 21 y la Figura 22. En estos casos, nos encontramos con 2 modelos o conjuntos de instancias (a), que implican 2 razonadores (c), y el nivel de complejidad de las consultas (d) es 3.

| Metamodelo de ontologías multilingüe   | Metamodelo ontologías modificando Concept Figura 9 | Metamodelo ontologías insertando conceptos Figura 10, Figura 11, Figura 12 y Figura 13 | Metamodelo ontologías + modelo RL Figura 15 y Figura 16 | Metamodelo ontologías + modelo mappings Figura 17 |   |                                |
|--|--|--|---|---|---|--------------------------------|
|  |  |  |   | Mappings en distribución ortogonal Figura 23      | Mappings en distribución radial Figura 25 | Mappings de aridad n Figura 24 |
| Modelo de ontologías multilingüe   | Figura 18 y Figura 19                              | Figura 20  | Figura 21 y Figura 22                                   | Mappings en distribución ortogonal Figura 23      | Mappings en distribución radial Figura 25 | Mappings de aridad n Figura 24 |
| Número de metamodelos y modelos (a)  | 1  | 1  | 2   | 2   | 2   | 2                              |
| Número de modelos y conjuntos de instancias: ontologías (O), mappings (M) y LR (b) | 1(O)   | 1(O)   | 1(O) + 1(LR)  | n(O) + n(n-1)/2 (M)                               | n(O) + 1(M)                               | n+1(O) + n(M)                  |
| Número de razonadores (c)  | 1 OR   | 1 OR   | 1 OR<br>1 LRR   | 1 OR<br>1 MR                                      | 1 OR<br>1 MR                              | 1 OR<br>1 MR                   |
| Complejidad de consultas (d)   | 1  | 2  | 3   | 4   | 5   | 4                              |
| Complejidad de adición de lengua (e)   | 3  | 1  | 1-3 (depende del recurso lingüístico)                   | 2   | 1   | 1                              |
| Complejidad mantenimiento consistencia (f)   | 1  | 1  | 2   | n + n(n+1)/2                                      | n+1                                       | 2n+1                           |
| Disponibilidad real (g)  | SÍ   | SÍ   | (depende del recurso lingüístico)                       | SÍ  | SÍ  | NO                             |

Tabla 1: Criterios de selección de soluciones

Por otra parte, el nivel de complejidad de añadir una nueva lengua (e) puede variar de 1 a 3, dependiendo del modelo del recurso

lingüístico. Si ello implica modificación del modelo en el recurso lingüístico la complejidad será 3 (en este caso representará una desventaja

importante), pero si sólo se trata de añadir nuevas instancias de la información lingüística, la complejidad será 1. Del mismo modo, el criterio de la disponibilidad de las herramientas y sistemas existentes está condicionado al modelo del recurso lingüístico.

**3) Metamodelo de ontologías + Modelo de mappings.** El metamodelo de multilingüidad formado por el metamodelo de ontologías y el modelo de *mappings* (Figura 17) tiene 2 metamodelos o modelos (a), pero el número de modelos de ontologías y conjuntos de instancias de mappings sobre los que mantener la consistencia (f) no es igual para cada instanciación. Así, se obtienen tres diferentes planteamientos, representados en la Figura 23, la Figura 24 y la Figura 25, respectivamente, donde este número varía:

- Si el sistema se forma con *mappings* binarios en una distribución ortogonal (Figura 23) el número es:  $n$  (ontologías) +  $n(n-1)/2$  (conjuntos de *mappings*).
- Si el sistema se forma con *mappings* binarios en una distribución radial (Figura 25) el número es:  $n + 1$  (ontologías) +  $n$  (conjuntos de *mappings*).
- Si el sistema se forma con *mappings* n-arios (Figura 24) el número es:  $n$  (ontologías) + 1 (conjuntos de *mappings*).

Estas soluciones tienen una gran desventaja al tener que repetir información conceptual en todas las ontologías (monolingües). La complejidad de las consultas (d) está entre 4 y 5 porque el número de modelos y los múltiples componentes que habría que consultar para resolverlas. Una limitación importante a tener en cuenta es el mantenimiento de la consistencia. El número de *mappings* puede ser muy alto, en consonancia con el número de modelos de ontologías, con lo que la consistencia puede verse amenazada. La adición de una nueva lengua (e) supone un nivel de complejidad bajo:

- Si el sistema se forma con *mappings* binarios en una distribución ortogonal (Figura 23) el nivel será el más alto de los tres casos, siendo de 2, al ser necesario tener que cambiar  $n$  conjuntos de instancias de *mappings*.
- Si el sistema se forma con *mappings* binarios en una distribución radial (Figura 25) el nivel de complejidad es 1, porque tan

sólo un conjunto de instancias de *mappings* ha de modificarse.

- Si el sistema se forma con *mappings* n-arios (Figura 24) el nivel de complejidad será el mismo que en el caso anterior, ya que sólo habrá que modificar un conjunto de instancias de *mappings*.

Sin embargo, el problema para el uso de la última solución (Figura 24) no es la complejidad asociada que lleve, sino que no existen sistemas y herramientas que sean capaces de manejar *mappings* de aridad mayor que dos.

Para hacer uso de la tabla y obtener la solución más adecuada a nuestras necesidades, habrá de tener en cuenta todas las características de la situación de partida a la hora de crear nuestra aplicación. Así, a los aspectos reflejados directamente en la tabla y que son propios del uso de la aplicación (frecuencia de modificaciones en el número de idiomas, frecuencia de consultas, etc.), se añaden los derivados de las situaciones particulares de cada caso. Por ejemplo, si partimos de cero y queremos crear un sistema basado en el conocimiento y una aplicación multilingüe sobre cualquier plataforma ya existente que gestione este sistema, no implica ninguna restricción adicional a las características intrínsecas del uso de la aplicación. Sin embargo, si partimos de una ontología creada y que está alojada en un determinado servidor que nos conviene mantener, las características de este servidor harán que el número de posibles soluciones a adoptar sea inferior (el servidor puede limitar el número máximo de modelos a usar, no dar soporte a *mappings*, etc.).

## 7 Datos

El último de los niveles en los que se manifiesta la multilingüidad es en los datos, es decir, la información sobre los individuos. En este caso, vendrá condicionada por la propia naturaleza de los individuos y estará modelada en la representación del conocimiento que se toma para el dominio.

Dependiendo de la naturaleza de los datos y del dominio de que se trate, la modelización de la multilingüidad dentro de la representación del conocimiento puede ser muy variada. La

multilingüidad se tratará como otro carácter del dominio que habrá que modelar.

En la Figura 26 se muestran dos ejemplos de datos multilingües modelados con una representación del conocimiento monolingüe. En ambos casos, los datos individualmente son monolingües, pero su conjunto es multilingüe y esa multilingüidad está contemplada dentro de

la representación de conocimiento por la característica Language, que es parte de la descripción del concepto (Article o Man).

En la Figura 27 se muestra otro ejemplo de datos multilingües, pero en este caso los individuos que aparecen son mono y multilingües.

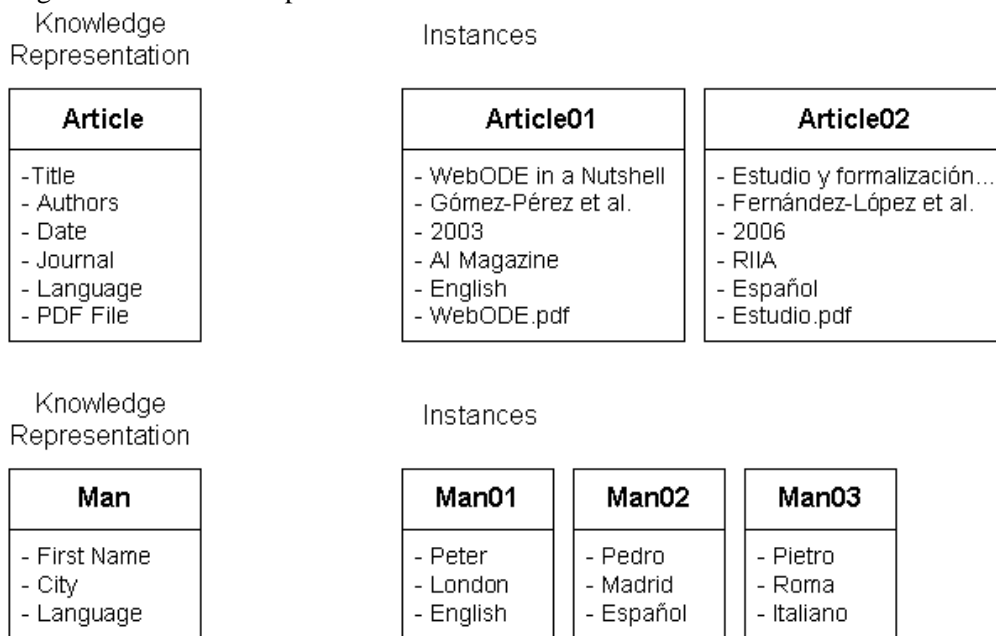


Figura 26: Ejemplos de datos multilingües (individuos monolingües)

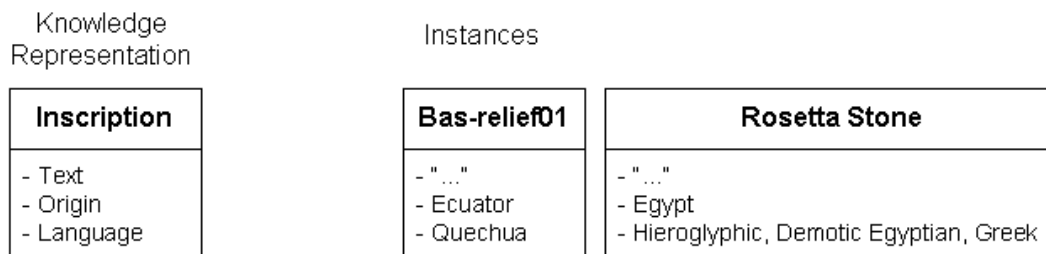


Figura 27: Ejemplo de datos multilingües (individuos multilingües)

## 8 Hibridaciones

Como se ha explicado a lo largo de este trabajo, la multilingüidad de la base de conocimiento se puede dar en uno o en varios componentes de la representación del conocimiento. Por tanto, pueden diseñarse sistemas en los que se haya optado por dotar de multilingüidad a cada componente, mediante una solución distinta. Por ejemplo, se puede tener la multilingüidad correspondiente a las etiquetas de los conceptos dentro del metamodelo de la ontología y, sin

embargo, aportar la multilingüidad para atributos y relaciones a través de la conexión con un metamodelo de un recurso lingüístico.

Se podría incluso optar por tener multilingüidad a nivel de aplicación global para el contenido, realizándose la traducción de las relaciones en tiempo de ejecución, mientras que los atributos se pueden traducir en tiempo de diseño.

## 9 Conclusiones

Como se ha explicado, el problema de dotar de multilingüidad a las aplicaciones sobre sistemas basados en el conocimiento, también denominado proceso de “localización”, ha adquirido gran importancia. Tras una revisión de diversos recursos multilingües, nos centramos en ofrecer una breve explicación de las nociones más relevantes de los tres niveles que pueden soportar multilingüidad en una de estas aplicaciones, es decir, la interfaz, la representación del conocimiento y los datos. En este trabajo se hace un examen exhaustivo de las diferentes posibilidades de representación y sus implicaciones a nivel técnico. Cabe destacar la importancia de la identificación de las ventajas e inconvenientes en todos y cada uno de los aspectos de la representación del conocimiento, principalmente en la elección de la modelización de la multilingüidad.

Los metamodelos de representación de multilingüidad que en este artículo se recogen son tres:

- (a) Metamodelo de ontología modificado
- (b) Metamodelo de ontología y Modelo de recurso lingüístico
- (c) Metamodelo de ontología y Modelo de *mappings*

Todos ellos cumplen con la función de proporcionar multilingüidad al sistema. Sin embargo, las implicaciones que se deriven de cada elección pueden tener distintos alcances, como se ha resumido en la Tabla 1, principalmente en cuanto a la complejidad de la consulta y la complejidad que implica la inclusión de una nueva lengua. Así pues, un metamodelo de ontología ampliado con información lingüística (a) presenta distintas modificaciones del metamodelo dependiendo de los componentes de la ontología que vayan a soportar multilingüidad, lo que a su vez implica cambios en la complejidad de la consulta. El caso (b) de agregación de un modelo de recurso lingüístico relacionado con el metamodelo de ontología permite una única conceptualización para todas las lenguas y la independencia de la información lingüística. En cuanto a la complejidad de la consulta, ésta aumenta por la existencia de dos modelos, mientras que la inclusión de una nueva lengua no requiere mayor complejidad que la que plantea el recurso lingüístico. Por último, el uso de un modelo de *mappings* para relacionar

ontologías monolingües (c) tiene dos implicaciones importantes: la primera se refiere a la existencia de ontologías en todas y cada una de las lenguas, y la segunda a que la disponibilidad de herramientas que traten con el modelo de *mappings* dependerá de la aridez de los mismos.

Finalmente, cabe destacar la posibilidad de combinar opciones de representación a los distintos niveles e incluso para los distintos componentes, creando de ese modo sistemas híbridos que se pueden adaptar a las necesidades del recurso y, lo que es más importante, dejan la puerta abierta a la reutilización de recursos existentes.

## 10 Agradecimientos

El trabajo aquí expuesto ha sido financiado por el proyecto europeo *NeOn: Networked Ontologies* (FP6-027595).

## Bibliografía

- Abascal, J. y R. Moriyón. 2002. Tendencias en interacción Persona-Computador. *Revista Iberoamericana de Inteligencia Artificial*, Vol. 6, nº 16.
- Esselink, B. 2000. *A practical guide to software localization*. John Benjamins Publishing.
- Hartmann, J. y R. Palma. 2006. OMV - Ontology Metadata Vocabulary for the Semantic Web, 2006. v. 2.0, available at <http://omv.ontoware.org/>
- Peñas, A. y J. Gonzalo. 2004. Acceso a información multilingüe. Número monográfico de la *Revista Iberoamericana de Inteligencia Artificial*, Vol. 8. nº 22.
- Cabré, M. T., C. Bach, R. Estopà, J. Feliu, G. Martínez y J. Vivaldi. 2004a. The GENOMA-KB project: towards the integration of concepts, terms, textual corpora and entities. *LREC 2004 Fourth International Conference on Language Resources and Evaluation*. Lisboa: European Languages Resources Association. pp. 87-90.
- Cabré, M. T., R. Estopà y J. Feliu. 2004b. A Specialized Knowledge Base: from Distributed Information to the Specialized Dictionary Construction. *11th EURALEX International Conference Proceedings* Lorient: Euralex. pp. 867-872.

- Feliu, J., J. Vivaldi y M.T. Cabré. 2002. Towards an Ontology for a Human Genome Knowledge Base. *LREC2002. Third International Conference on Language Resources and Evaluation. Proceedings*. Las Palmas de Gran Canaria, pp. 1885-1890. ISBN: 295-1740-808.
- Miller G., R. Beckwith, C. Fellbaum, D. Gross y K. J. Miller. 1990. Revised in 1993. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4), 235–244.
- Moreno, A. 2000. Diseño e implementación de un lexicón computacional para lexicografía y traducción automática. *Estudios de Lingüística Española*, Volumen 9.
- Soergel D., B. Lauser, A. Liang, F. Fisseha, J. Keizer y S. Katz . 2006. Reengineering Thesauri for New Applications: the AGROVOC Example. <http://journals.tdl.org/jodi/article/viewArticle/jodi-126/111>
- Vossen, P. 2002. EuroWordNet General Document. Disponible en <http://www.vossen.info/docs/2002/EWNGeneral.pdf>
- Vossen, P. 2004. EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index. *Semi-special issue on multilingual databases*, IJL 17/2.