

# Ontologías y multilingüalidad

**Dra. Guadalupe Aguado de Cea**

# Agradecimientos

**Mauricio Espinoza**

**Elena Montiel-Ponsoda**

**José Angel Ramos Gargantilla**

Por haberme facilitado algunas de las transparencias

# Indice

- Definición y finalidad de la multilingualidad
- Localización vs. internationalización
- De los sistemas monolingües a los sistemas multilingües
- Sistemas de PLN con multilingualidad
- Multilingualidad en los sistemas KB: ontologías
  - Interfaz
  - Datos
  - Representación
- Una nueva propuesta: Linguistic Information Repository- LIR
- Una herramienta: LabelTranslator

# Multilingualidad ¿para qué?

- Necesidad de multilingualidad en los sistemas de PLN
  - Sistemas de búsqueda de respuestas
  - Búsqueda de información multilingüe
  - Recuperación de información
  - Traducción automática
- Compartición de conocimientos → ontologías
- Ontologías → Web semántica

# ¿Cómo conseguir la multilingüalidad?

## **Localización** vs. internacionalización

**Localization** involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold (LISA)

En economía

Adaptar un producto a un entorno distinto del original (*a non-native environment*).

En sw y diseño web

Adaptar el contenido, la lengua y el diseño a la cultura y la lengua de llegada

**Ontology Localization**  
*involves the process of adapting an ontology to a particular language and culture.*

En ontologías

# Internacionalización

- ***Internationalization** is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design. Internationalization takes place at the level of program design and document development (LISA).*
- Es importante:
  - Separar el texto del código fuente – > evita que los traductores cambien el código fuente
  - No se limita al software: *online help, documentation and web sites* pasan por este proceso
  - Para los escritores técnicos: “*writing for a global audience*”, “*web site globalization*”

# Internacionalización y localización en ontologías

Internacionalización	Localización
<b>Contenido Léxico</b> : caracteres y símbolos manejados por el ordenador (ASCII encoding, UNICODE, etc.)	<b>Contenido Léxico –terminológico</b> : términos o palabras utilizadas para denominar los elementos de la ontología
<b>Contenido gramatical</b> : caracteres, estructuras sintácticas & símbolos empleados en algunos lenguajes de ontologías (RDF(S), OWL)	<b>Contenido conceptual</b> : decisiones de conceptualización: granularidad, expresividad, perspectiva, etc., especialmente en las ontologías de dominio
<b>Paradigma de representación</b> : <i>frames</i> , <i>semantic networks</i> , DL, (Ontologías)	<b>Contenido pragmático</b> : resultado final del modelo (GUI, etc.)
<b>A nivel de diseño</b>	<b>En fase posterior</b>

# De la monolingualidad a la multilingualidad en las ontologías

- Pocas ontologías multilingües
  - <http://olp.dfki.de/ontoselect/>
- Poca información disponible sobre representación de multilingualidad
- Reciente interés en los grupos de investigación internacionales:
  - LISA (*Localization Industry Standards Association*)
  - OSCAR (*Open Standards for Container/Content Allowing Re-use*)
  - OASIS (*Organization for the Advancement of Structured Information Standards*)
  - W3C
  - ISO *International Standards Organization*



# Algunos sistemas que incorporan multilingüalidad: EWN

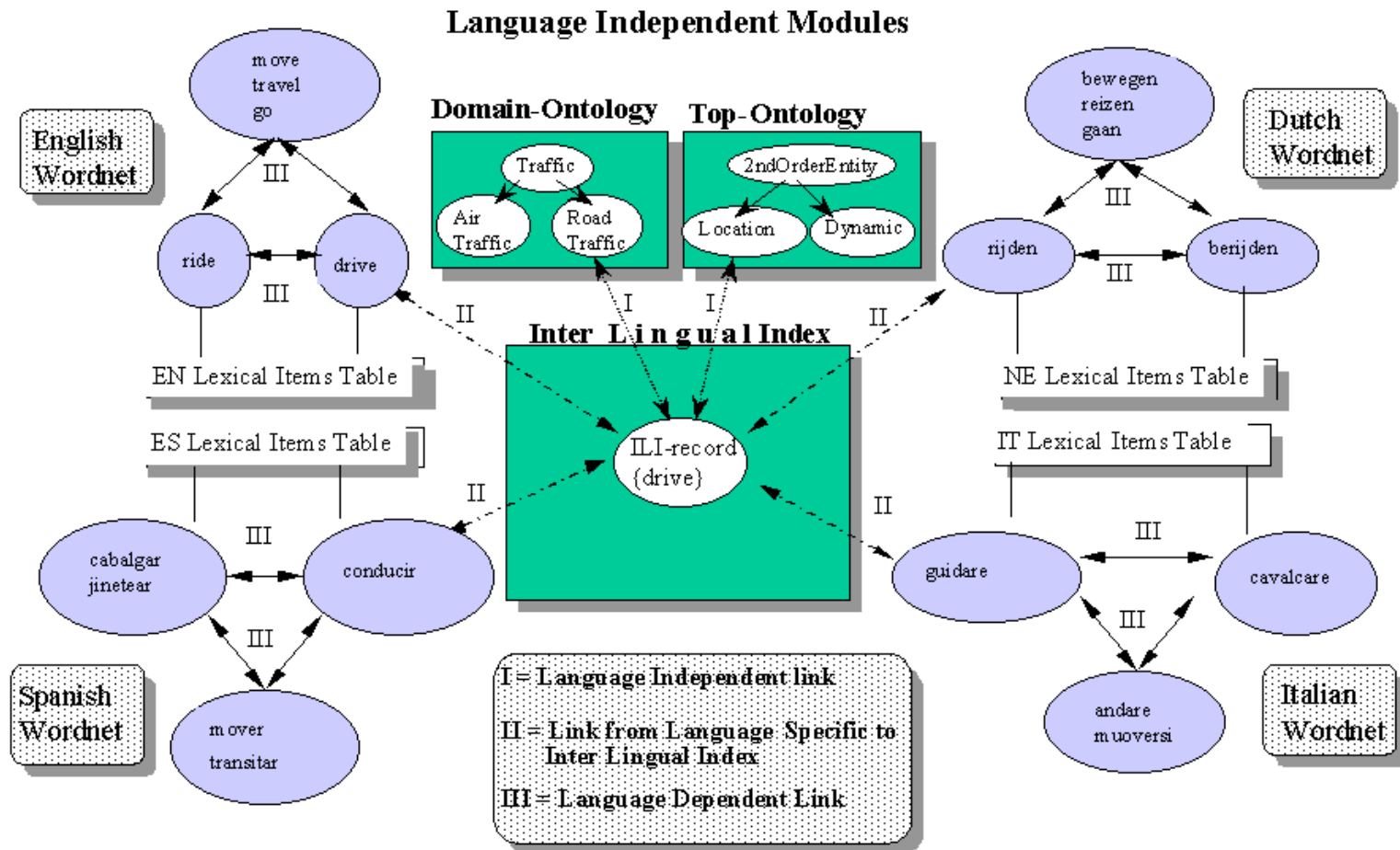
## ➤ EuroWordNet

- Basado en Wordnet, <http://wordnet.princeton.edu/perl/webwn>

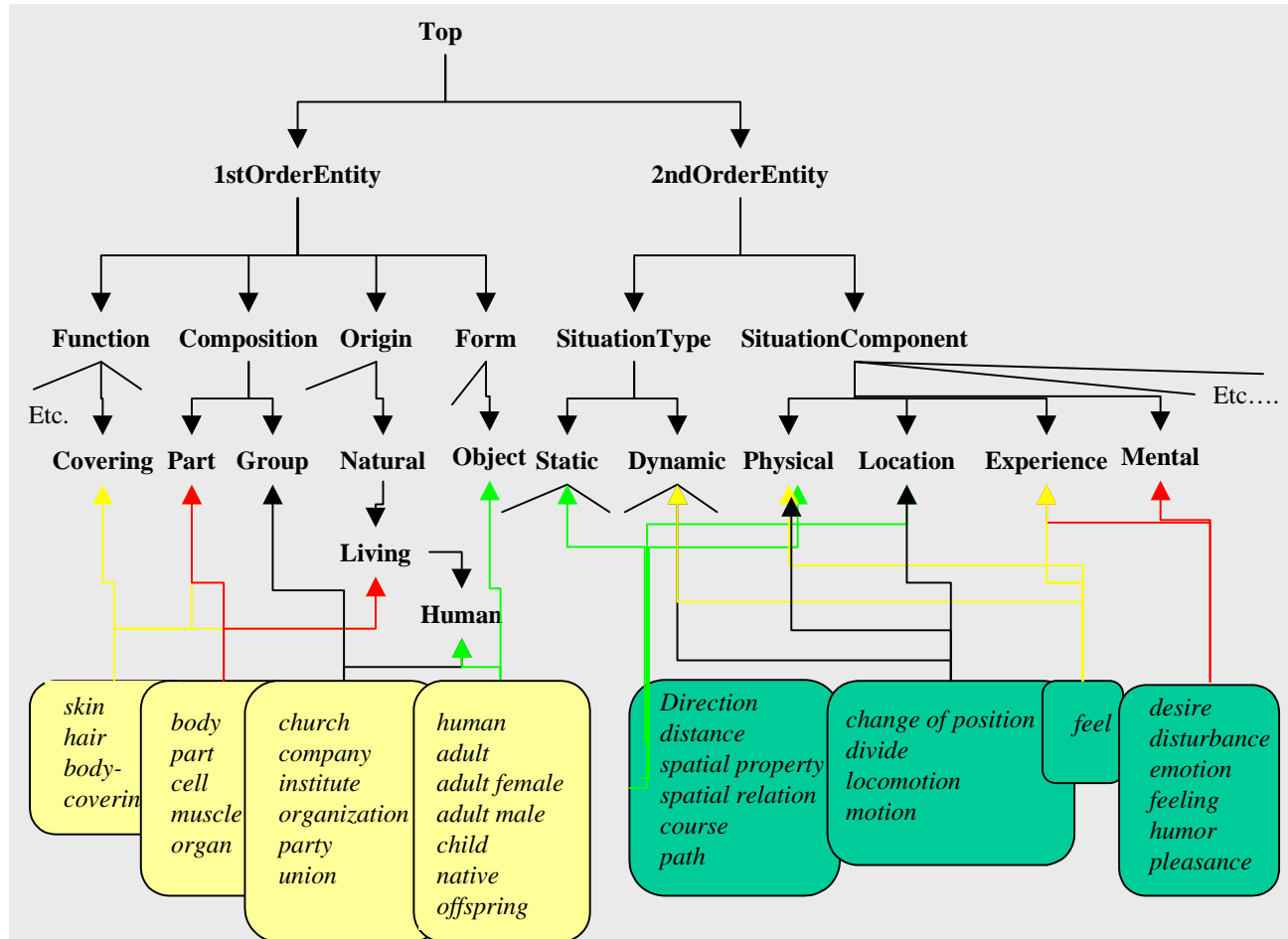
## ➤ Objetivos:

- Crear una BD léxica **multilingüe** para holandés, italiano, español, inglés, alemán, francés, estonio y checo
  - Mantener las relaciones específicas de las lenguas en sus redes
  - Lograr la máxima compatibilidad entre los distintos recursos
  - Construir las redes de forma independiente, reutilizando recursos propios de cada lengua
  - Redes de palabras: *wordnets* -> ontologías autónomas monolingües, conectadas mediante un ILI
  - *Synsets* (sinónimos)
- 
- **Instituciones: 8 universidades (UNED, UPC), 3 empresas.**
  - **Financiado por la U.E.**

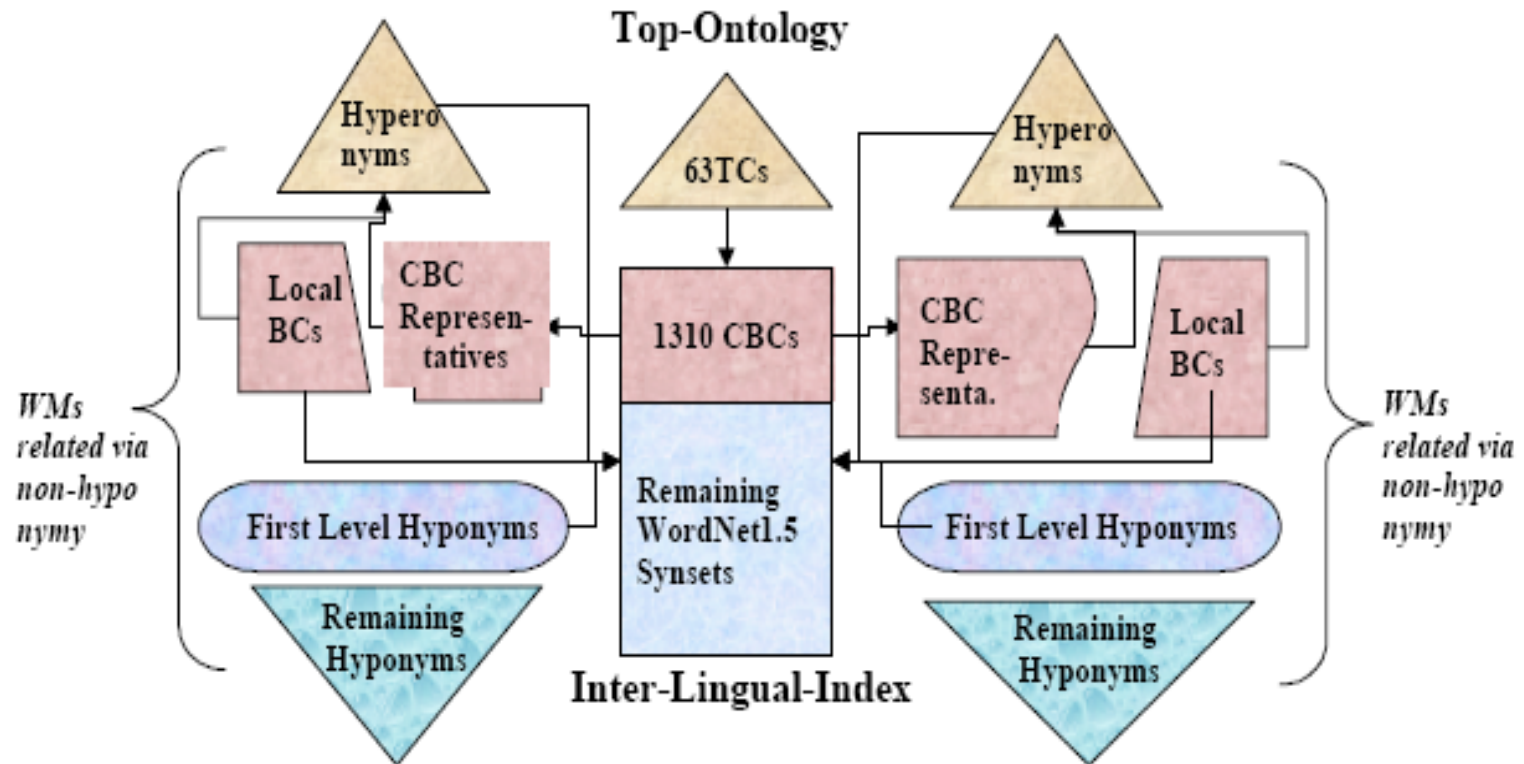
# EuroWordNet. Arquitectura



# EuroWordNet: Top Ontology



# Esquema general de dos *wordnets* mapeados al ILI Eurowordnet



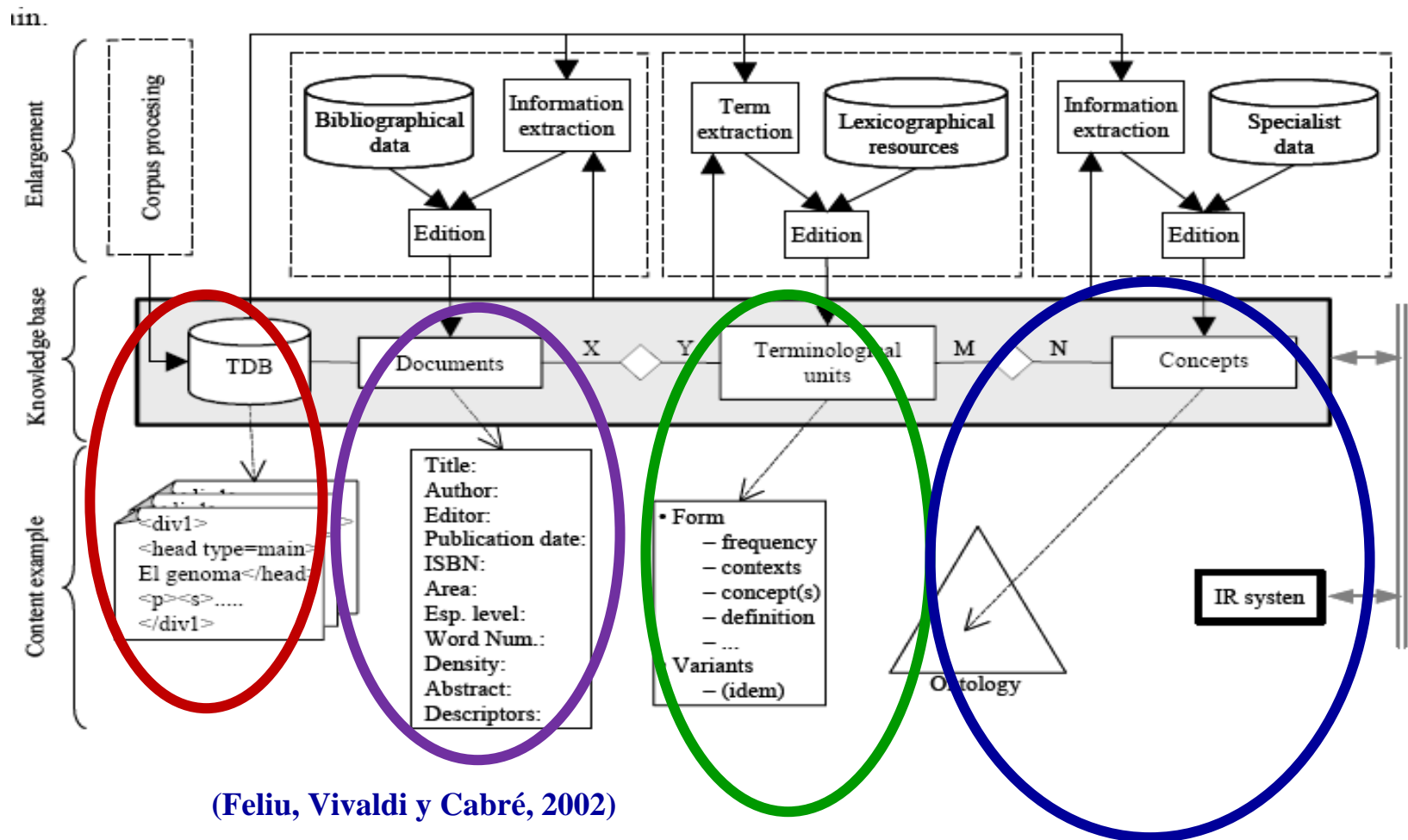
(Vossen, 2002)

TC: Top concepts  
CBC: Common Base Concepts  
BC: Base Concepts

# Algunos sistemas que incorporan multilingualidad: Genoma-KB

- **Módulo ontológico:** (MikroKosmos) 21 conceptos básicos: ALL, OBJECT (physical, mental, social), EVENT (physical, mental, social), PROPERTY (attribute, relation), etc.
  - Las relaciones descritas son: (Feliu 2004)
    - Similaridad, Hiponimia, Secuencialidad (lugar y tiempo)
    - Causalidad, Instrumentalidad, Meronimia, Asociación
- **Módulo terminológico**
  - Multilingualidad, POS, contexto, fuentes, lema, información administrativa
- **Módulo de corpus:** textos multilingües
- **Módulo de entidades:**
  - Módulo bibliográfico: refer. completas de textos y términos
  - Módulo factográfico: centros de invest. , personas, instituciones,

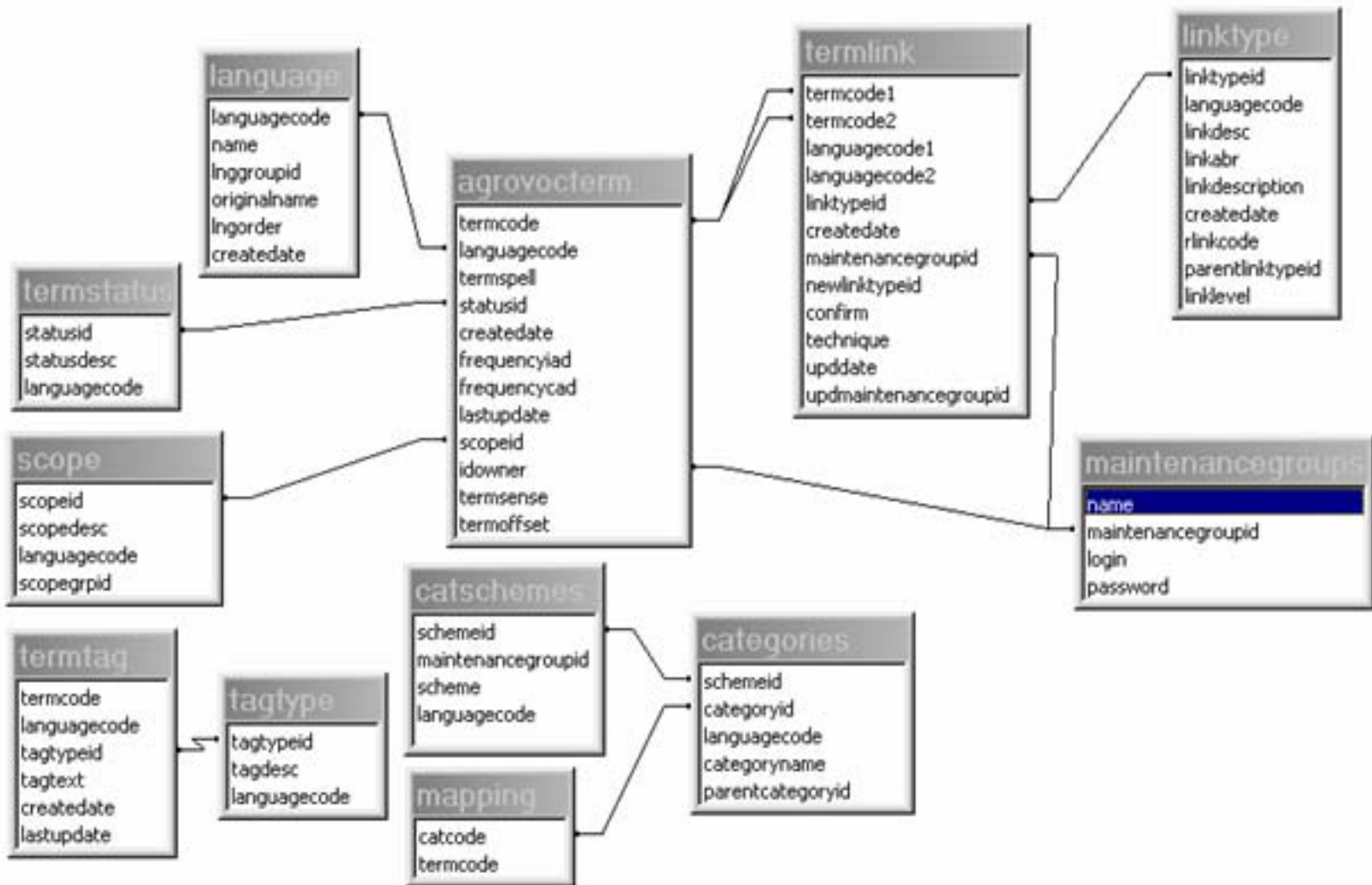
# Arquitectura de la base de conocimiento GENOMA-KB



# Algunos sistemas que incorporan multilingüalidad: AGROVOC

- *AGROVOC Thesaurus* desarrollado por la FAO (*Food and Agriculture Organization*) y la UE en 1980/1982.
- 3 lenguas iniciales. Actualmente 10 . Se incorporarán algunas más.
- Se define como “*a multilingual structured and controlled vocabulary*”.
- Utilizado para indexar y recuperar datos sobre pesca, agricultura y alimentación
- Muestra el nº de términos en tiempo real (41,580 términos en español)
- Relaciones:
  - Scope Note Reference, Is Referenced in, Scope Note, Used For, Used For+ , Broader Term , Narrower Term, Use, Related Term
- URL [http://www.fao.org/aims/ag\\_figures.jsp](http://www.fao.org/aims/ag_figures.jsp)

# AGROVOC: representación de la información multilingüe





# Cómo incorporar multilingüalidad en los SBC: ontologías (1)

La multilingüalidad puede darse en tres niveles:

## 1. Interfaz

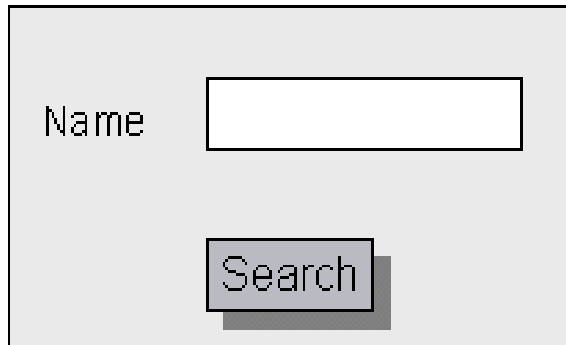
- Mensajes
- Contenido

## 2. Datos

## 3. Representación de conocimiento

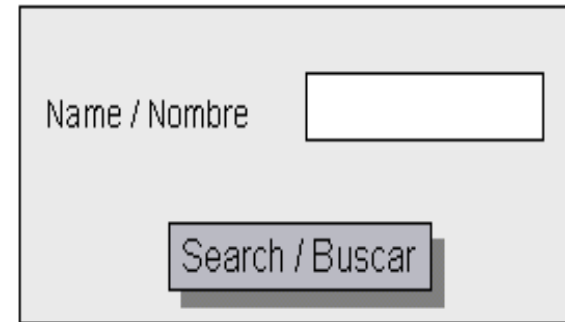
- *Aguado de Cea, G., Montiel Ponsoda, E., Ramos Gargantilla, J.A. “Multilingüalidad en una aplicación basada en el conocimiento”, Procesamiento del lenguaje natural, nº 38, Abril 2007*

# 1. Interfaz: Visualización de mensajes



A rectangular interface box with a light gray background. It contains the label "Name" on the left, followed by a white rectangular input field. Below the input field is a gray button with the text "Search".

**1. Mensajes monolingües**



A rectangular interface box with a light gray background. It contains the label "Name / Nombre" on the left, followed by a white rectangular input field. Below the input field is a gray button with the text "Search / Buscar".

**2. Mensajes multilingües  
simultáneos**



A rectangular interface box with a light gray background. It contains the label "Name" on the left, followed by a white rectangular input field. Below the input field are two flags: the United Kingdom flag and the Spanish flag. To the right of the flags is a gray button with the text "Search".

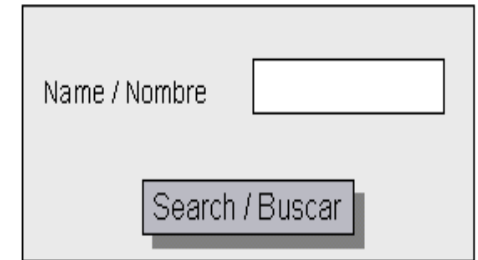


A rectangular interface box with a light gray background. It contains the label "Nombre" on the left, followed by a white rectangular input field. Below the input field are two flags: the United Kingdom flag and the Spanish flag. To the right of the flags is a gray button with the text "Buscar".

**3. Mensajes multilingües no simultáneos**

# Ventajas y desventajas

- En la **visualización simultánea**, la incorporación de otras lenguas requiere **modificar** código de visualización.



A screenshot of a web interface for simultaneous visualization. It features a light gray background. At the top left, the text "Name / Nombre" is displayed next to a white rectangular input field. Below this, centered, is a gray button with the text "Search / Buscar" in white.

- En la **visualización no simultánea**, no implica modificar todo el código, sino:
  - ampliar el nº de interfaces
  - modificar las opciones de selección.



A screenshot of a web interface for non-simultaneous visualization. It has a light gray background. The text "Name" is followed by a white input field. Below the input field are two small flags: the Union Jack (UK) and the Spanish flag. To the right of the flags is a gray button labeled "Search".



A screenshot of a web interface for non-simultaneous visualization, similar to the previous one but in Spanish. It has a light gray background. The text "Nombre" is followed by a white input field. Below the input field are two small flags: the Union Jack (UK) and the Spanish flag. To the right of the flags is a gray button labeled "Buscar".

# Interfaz: Visualización del **contenido** de forma multilingüe

- BC es **multilingüe**
  - La aplicación consulta a la BC
  - La interfaz muestra el contenido en el idioma seleccionado
- BC es **monolingüe**
  - La aplicación consulta a la BC
  - Se utiliza un sistema de traducción (recurso multilingüe)
  - La interfaz muestra la traducción
- Interfaz similar en ambos casos a visualización de mensajes

# Ventajas y desventajas

BC es multilingüe	BC es monolingüe
Tiempo de obtención de contenidos = tiempo de respuesta (TR) de la BC	Tiempo de obtención de contenidos = TR de BC + TR del recurso multilingüe
Razón: se ha conferido multilingualidad a la BC en tiempo de diseño	La traducción se realiza en tiempo de ejecución
Desambiguación: en tiempo de diseño	Desambiguación: puede alargar el TR

# Multilingüidad en los SBC (2)

La multilingüidad puede darse en tres niveles:

## 1. Interfaz

- Mensajes
- Contenido

## 2. Datos

## 3. Representación de conocimiento

- *Aguado de Cea, G., Montiel Ponsoda, E., Ramos Gargantilla, J.A. “Multilingüidad en una aplicación basada en el conocimiento”, Procesamiento del lenguaje natural, nº 38, Abril 2007*

# Multilingualidad en los **datos** de los SBC

Knowledge Representation

Article
- Title
- Authors
- Date
- Journal
- Language
- PDF File

Instances

Article01
- WebODE in a Nutshell
- Gómez-Pérez et al.
- 2003
- AI Magazine
- English
- WebODE.pdf

Article02
- Estudio y formalización...
- Fernández-López et al.
- 2006
- RIIA
- Español
- Estudio.pdf

- Representación monolingüe
- Información sobre los individuos es multilingüe

Knowledge Representation

Man
- First Name
- City
- Language

Instances

Man01
- Peter
- London
- English

Man02
- Pedro
- Madrid
- Español

Man03
- Pietro
- Roma
- Italiano

- La multilingualidad se trata como otro aspecto a modelar

**Datos multilingües en una RC monolingüe que considera la característica Language**

# Multilingüidad en los SBC (3)

La multilingüidad puede darse en tres niveles:

## 1. Interfaz

- Mensajes
- Contenido

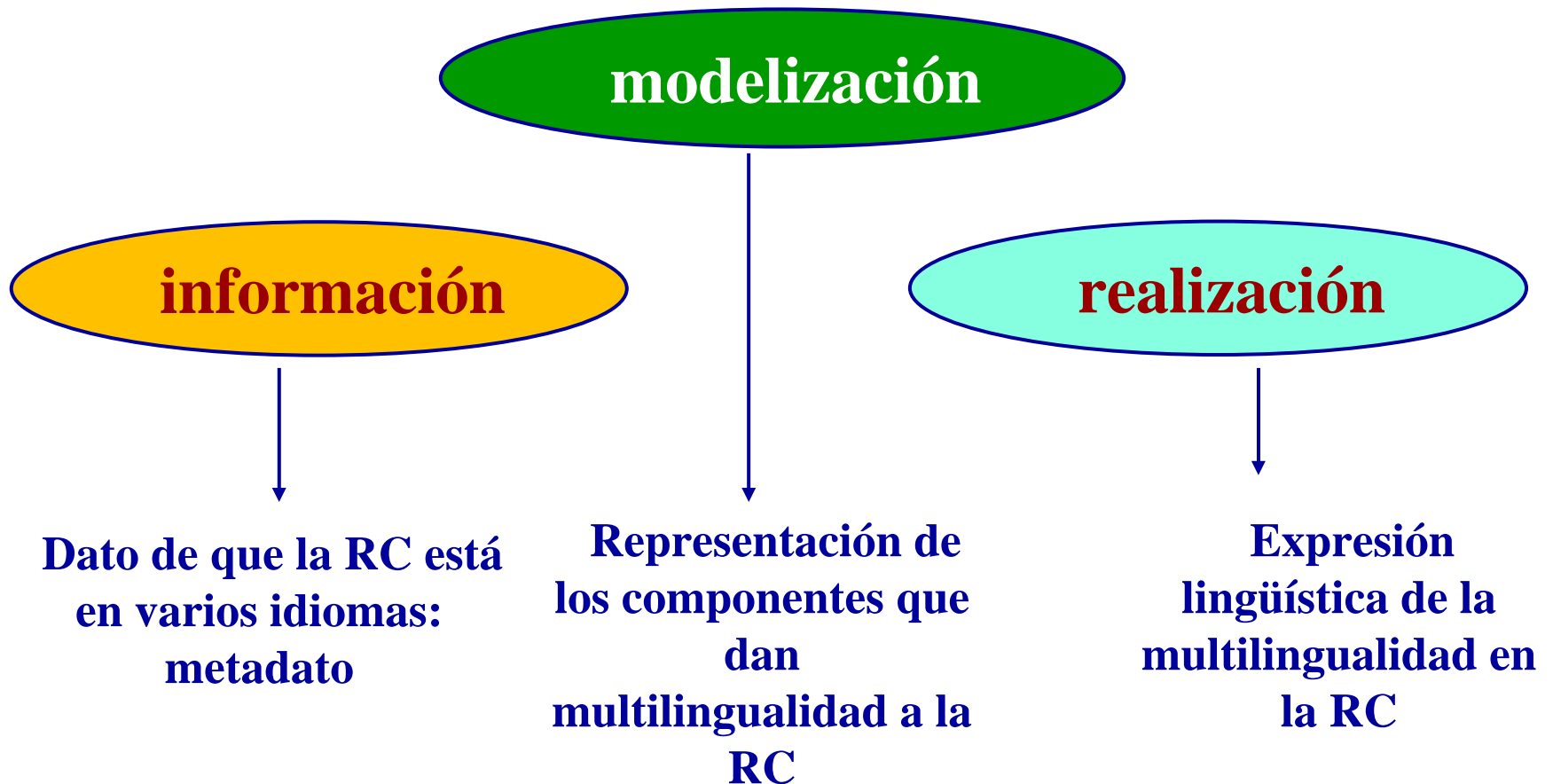
## 2. Datos

## 3. Representación de conocimiento

- *Aguado de Cea, G., Montiel Ponsoda, E., Ramos Gargantilla, J.A. “Multilingüidad en una aplicación basada en el conocimiento”, Procesamiento del lenguaje natural, nº 38, Abril 2007*

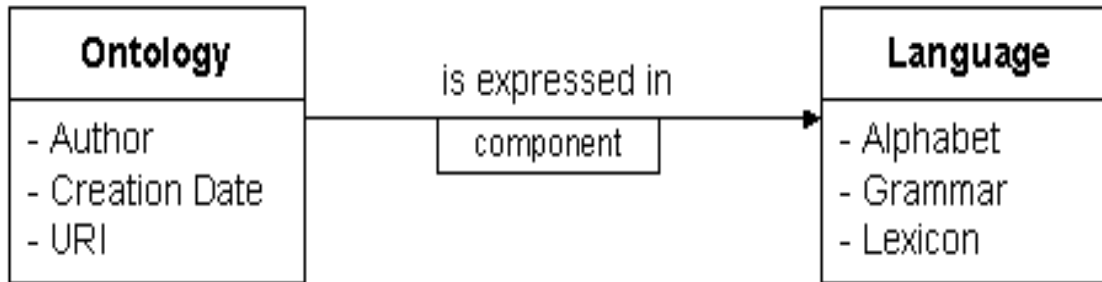


# Multilingualidad en ontologías: representación



# 1. Información. Ejemplo

## Estándar: OMV (Ontology Metadata Vocabulary)

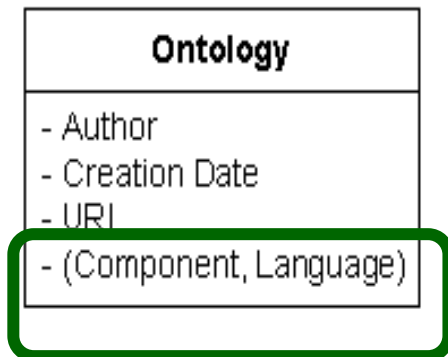


### ➤ **Desventaja:**

- Dificultad de instanciar el concepto **language** con toda la información
- Pocos sistemas tienen relaciones con información semántica asociada

### ➤ **Ventaja:** Riqueza de información lingüística

## Opción 1. Multilingualidad mediante **relación**



## Opción 2. Multilingualidad modificando el concepto **ontology**

### ➤ **Desventaja:** se pierde información lingüística

### ➤ **Ventaja:** es más sencilla, más fácil de implementar

# Modelización

Tres opciones:

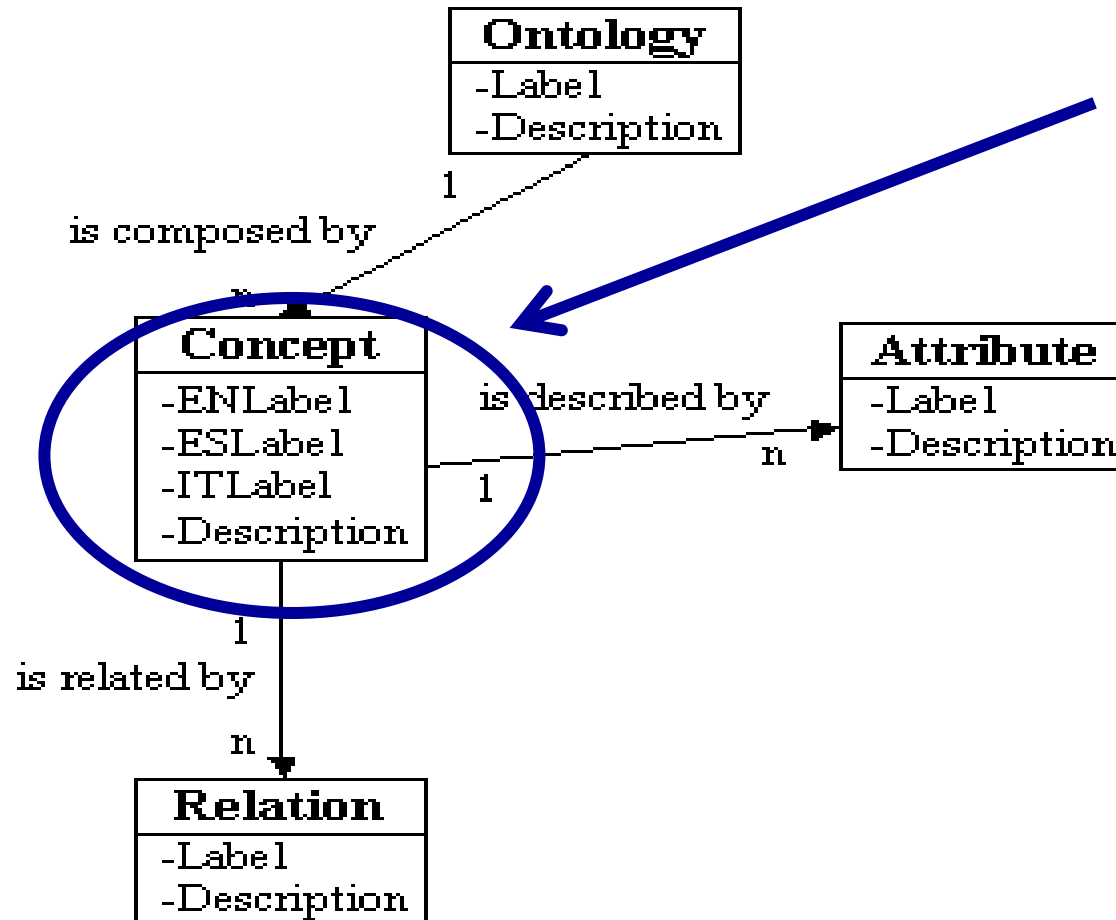
A. **Ampliar** el metamodelo de ontologías con información lingüística

B. Utilizar *mappings* para relacionar ontologías monolingües

C. **Agregar** un modelo de información lingüística y relacionarlo con metamodelo de ontologías

## 2. Modelización.

### A. Ampliación de metamodelo con info lingüística Opción A. (1)



- Multilingualidad en conceptos
- Mediante `rdfs:label` y `rdfs:comment` para definir etiquetas y descripciones de clases en LN.
- Localización en la capa terminológica.
- Es la **modalidad más utilizada**

# Ampliación de metamodelo con info lingüística

## ➤ Ventajas:

- Inclusión de otras lenguas es fácil, solo hay que añadir etiquetas a la ontología.
- Adecuado para ontologías de dominio muy especializado: el conocimiento está compartido y consensuado por expertos del dominio en diferentes lenguas.

## ➤ Desventajas:

- La información lingüística que se incluye es limitada.
- Se asume una sinonimia completa, pero no es cierta

# Modelización

Tres opciones:

A. Ampliar el metamodelo de ontologías con información lingüística

B. Utilizar *mappings* para relacionar ontologías monolingües

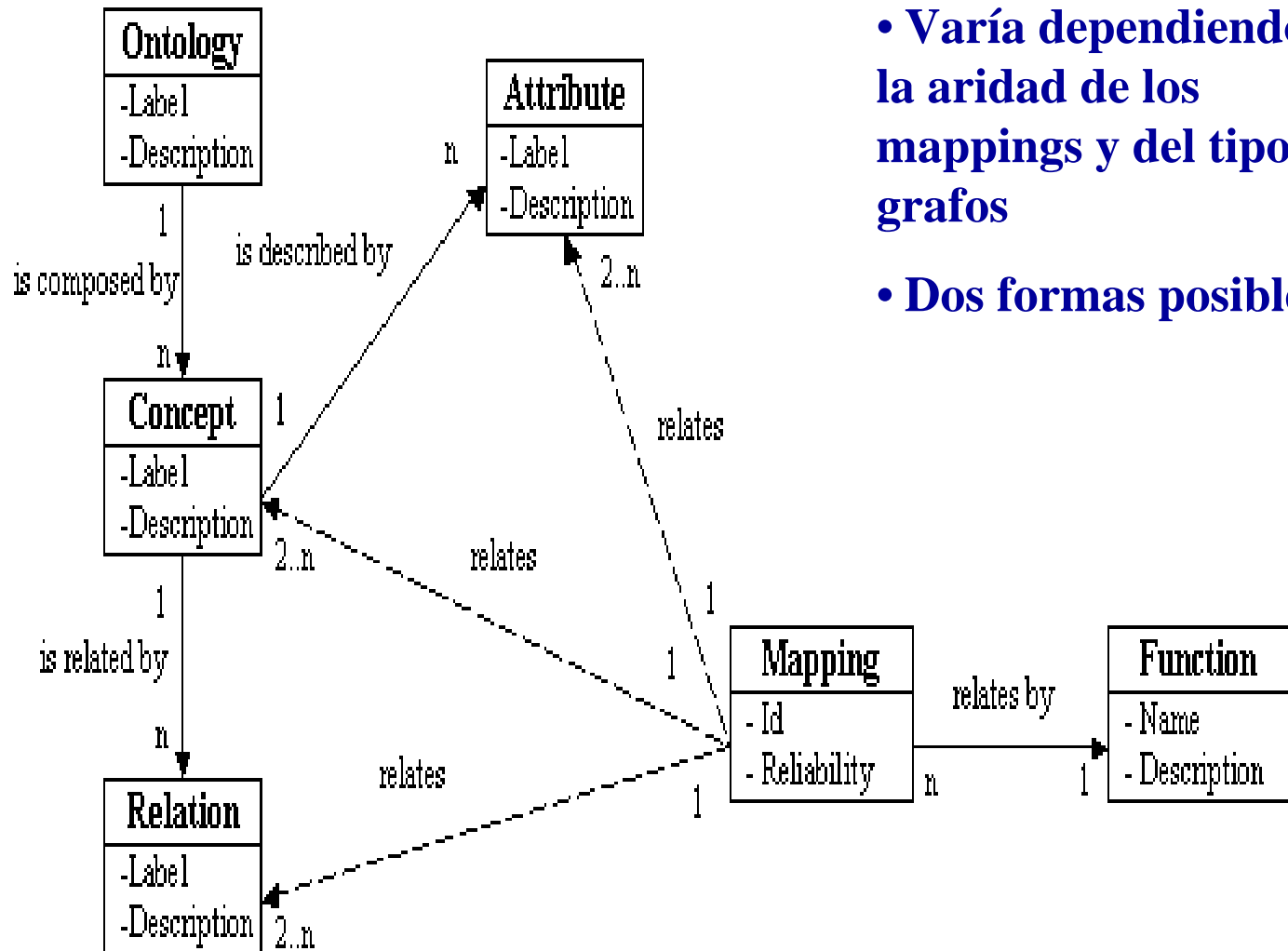
C. Asociar un modelo de información lingüística con metamodelo de ontologías

# Modelización.

## Metamodelo de ontología y modelo de *mappings*

- Las ontologías son del mismo dominio, pero diferente lengua
- Los *mappings* son de equivalencia entre los términos de las distintas lenguas
  - Translate
  - Equivalence label
- Esta opción no pretende ser metamodelo de multilingualidad
- Modelo menos intuitivo a nivel de abstracción
- Otros tipos de *mappings*, por ejemplo, en *C-OWL*:
  - Igualdad, contiene a, contenido en, solapamiento, y sus negaciones

## B. Metamodelo de Ontología y modelo de *mapping* : Ejemplo

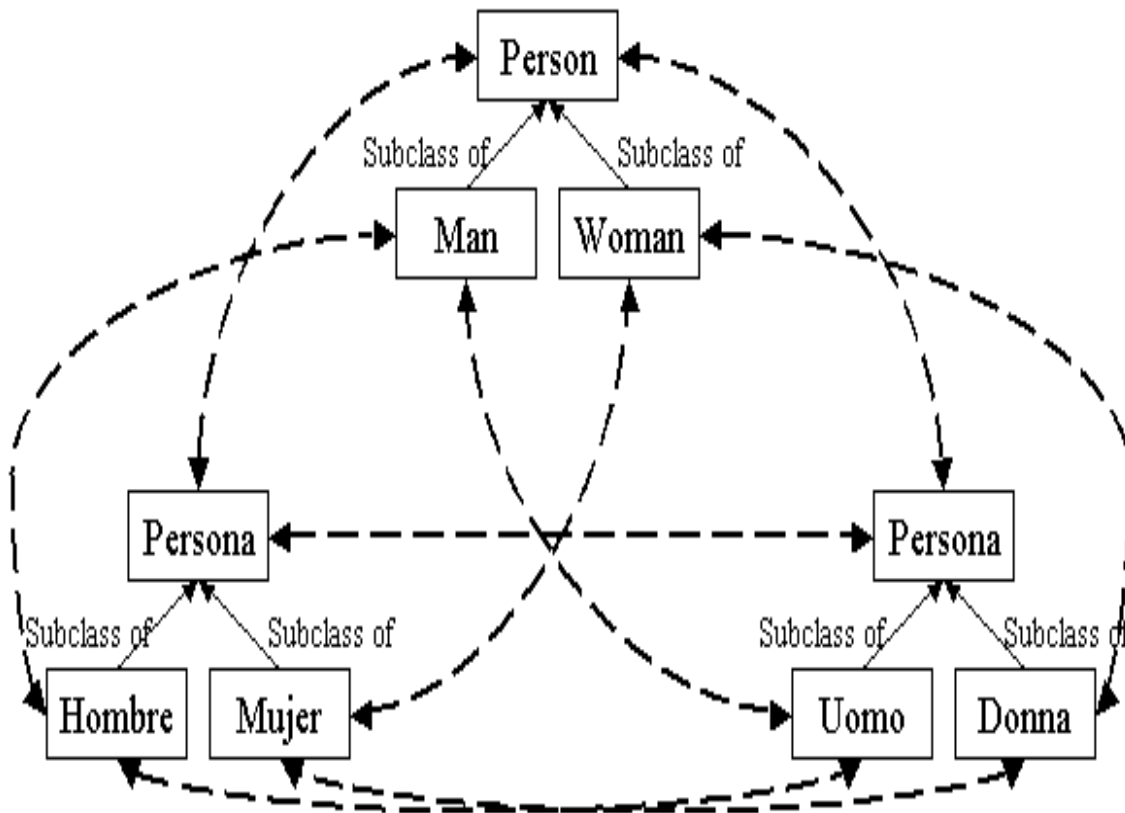


- Varía dependiendo de la aridad de los mappings y del tipo de grafos
- Dos formas posibles



## B. Combinación del *meta-modelo de ontología* con un modelo de *mapping*

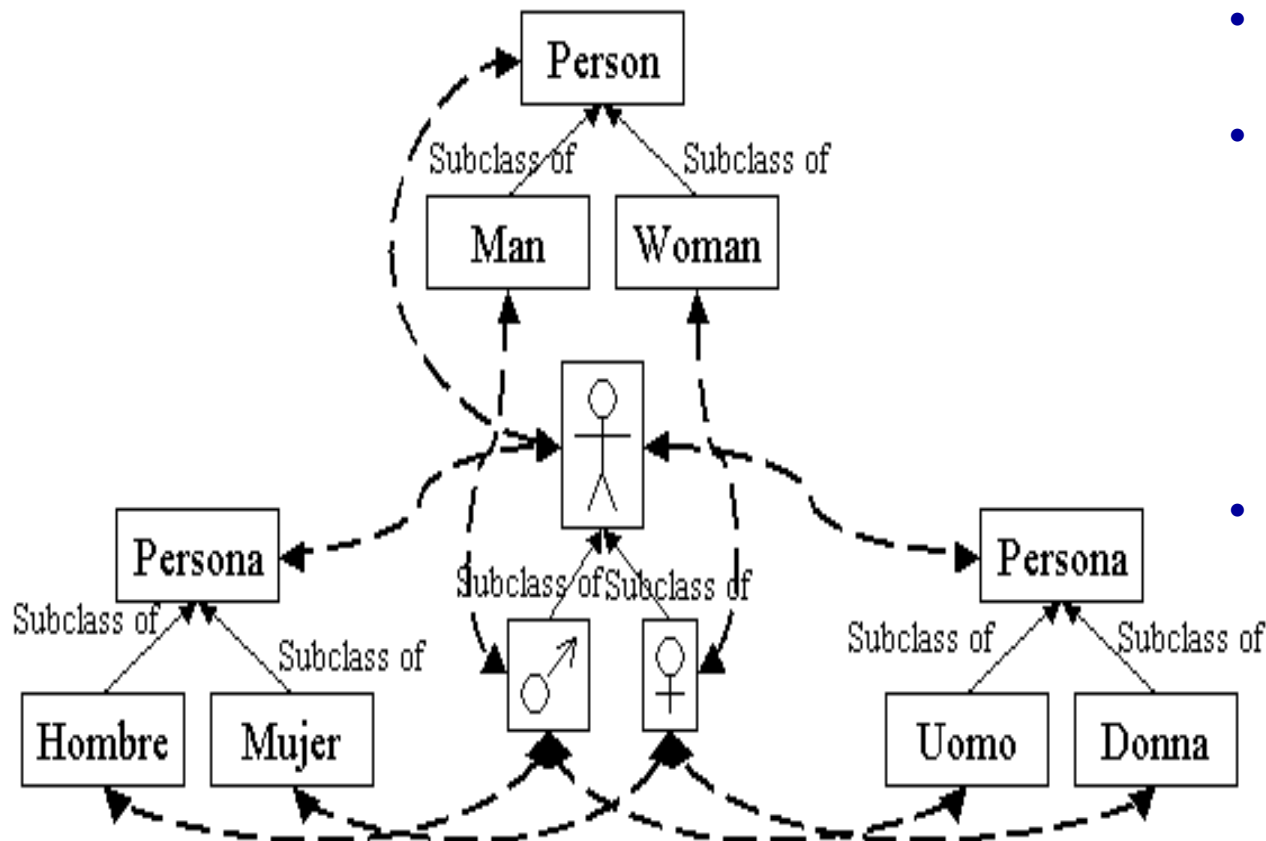
### 1. Mapping binario en grafo **ortogonal**.



- La localización se hace en la capa **conceptual**.
- *Each monolingual ontology structure knowledge is mapped to the rest of ontologies in a **pair-wise way***
- Es menos intuitivo desde un punto de vista abstracto

# B. Combinación del *meta-modelo de ontología* con un modelo de *mapping*

## 2. Mappings binarios con grafo **radial**



- La localización se hace en la capa **conceptual**.
- Las ontologías monolingües se mapean mediante una **interlingua**: conjunto de conceptos para establecer las equivalencias
- **EWN**

# Ventajas y desventajas de combinar el metamodelo de *ontología con mappings*

## ➤ Ventajas:

- Las conceptualizaciones se mantiene en cada lengua
- Adecuado para ontologías muy dependientes de una cultura determinada: sistema judicial.

## ➤ Desventajas:

- Mucho esfuerzo para conceptualizar el mismo dominio en lenguas diferentes.
- Se requieren tres tipos de expertos:
  - del dominio,
  - Lingüistas
  - en ingeniería ontológica.

# Modelización

Tres opciones:

A. Ampliar el metamodelo de ontologías con información lingüística

B. Utilizar *mappings* para relacionar ontologías monolingües

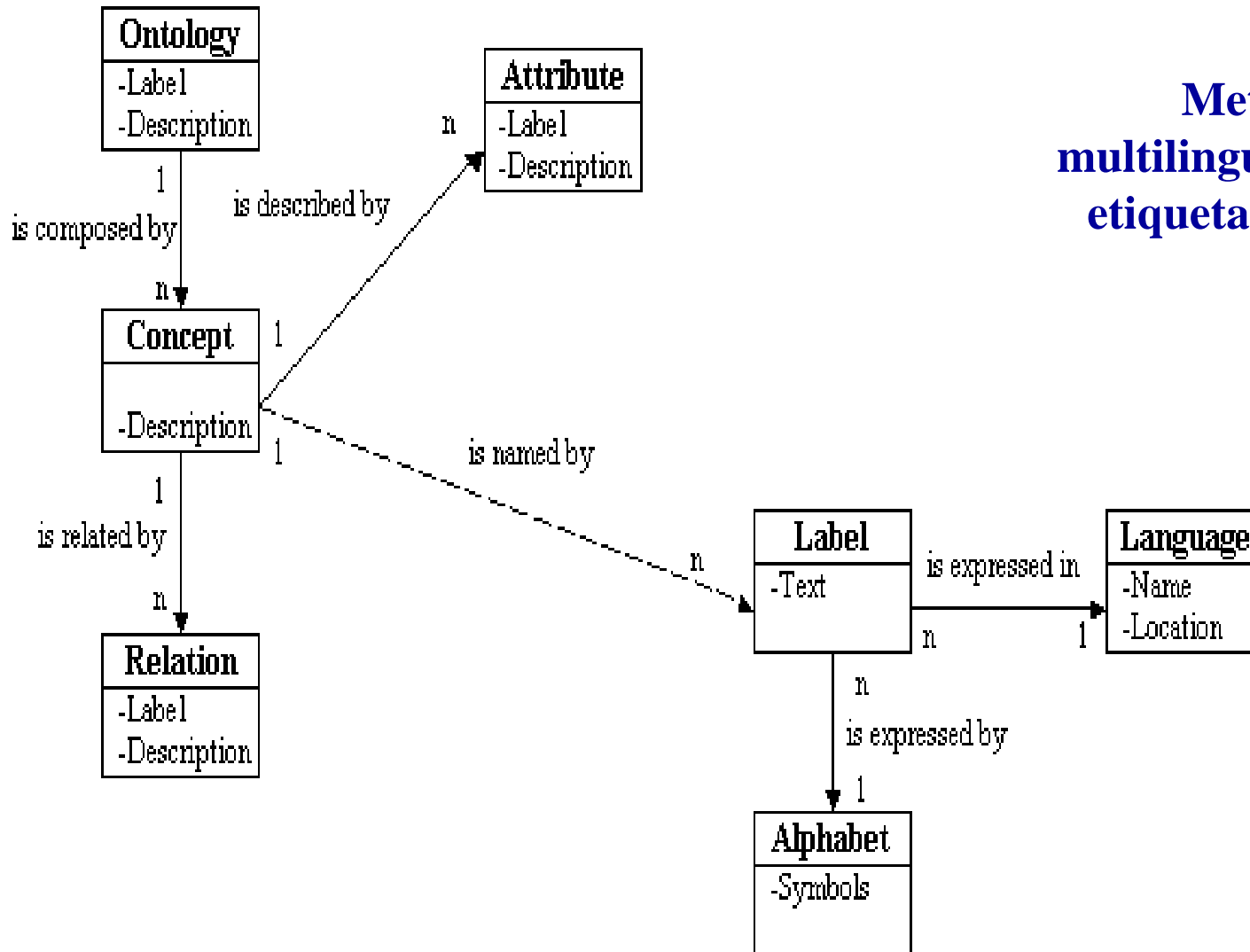
C. Asociar un modelo de información lingüística con metamodelo de ontologías

# Modelización

## C. **Asociar** un modelo de información lingüística con metamodelo de ontologías

- La localización se hace al nivel **terminológico** y **conceptual**
- Los elementos de la ontología se enlazan con los datos multilingües almacenados fuera de la ontología
- La información lingüística puede estar en una BD (GENOMA-KB, OncoTerm), o en una ontología, etc.
- La capa de conceptualización de la ontología puede modificarse para satisfacer las necesidades de localización, (creación de módulos específicos de la lengua, etc.)

## C. Meta-modelo de ontología enlazado con *modelo lingüístico multilingüe* : Ejemplo



**Metamodel de multilingualidad para las etiquetas de conceptos**

**Aumenta las posibilidades de incluir información sobre los componentes de la ontología y los lingüísticos**

# Ventajas y desventajas

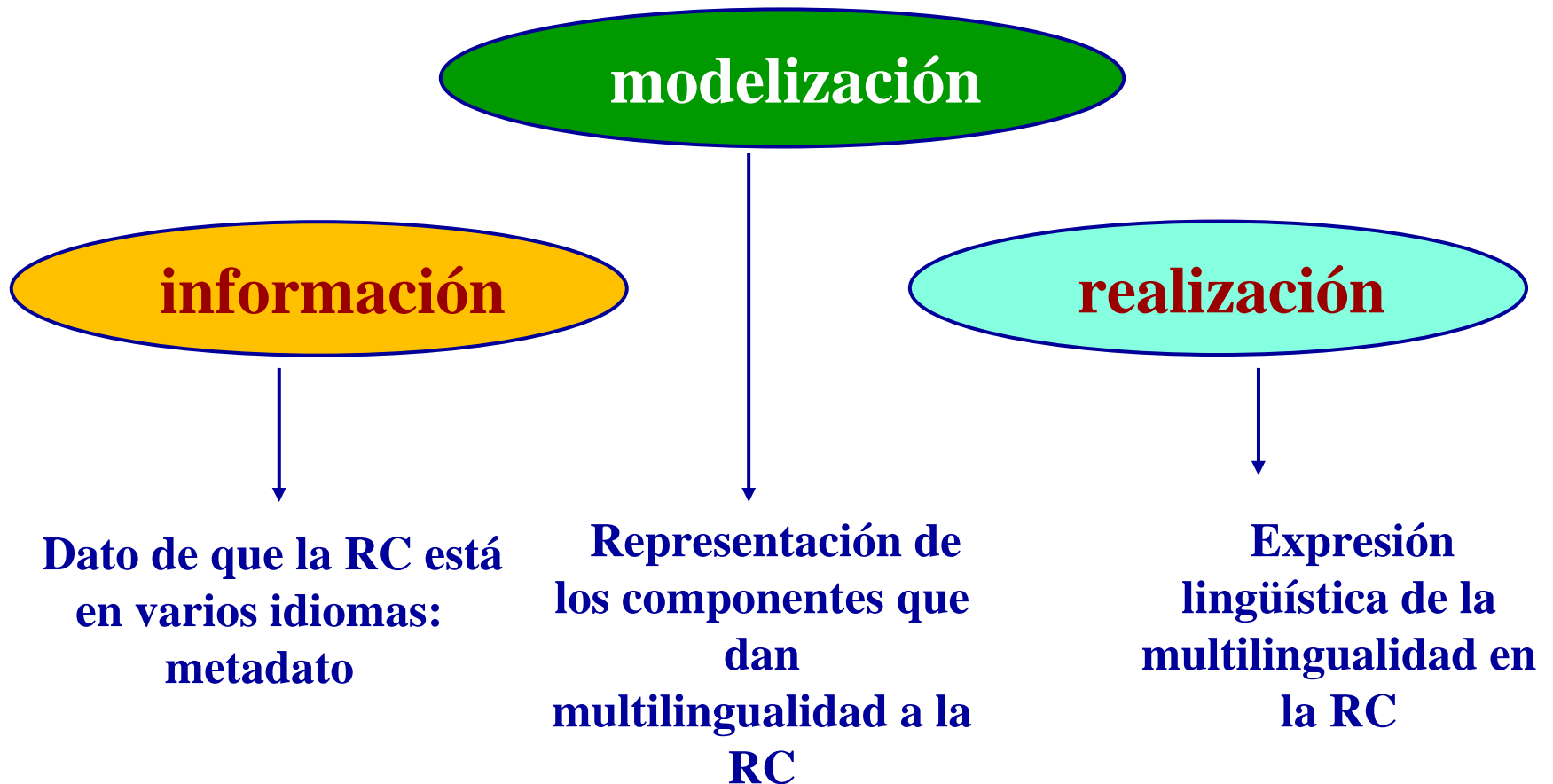
- **Ventajas:**

- Posibilidad de incluir tanta información lingüística como se desee
- Facilidad para enlazar elementos lingüísticos en las diferentes lenguas
- Respeto de las diferencias o matices de las lenguas al nivel terminológico
- Inclusión de información relevante, por ejemplo , el origen (*provenance of the linguistic elements*)
- Linguistas y expertos del dominio pueden acceder a la capa terminológica en un entorno distribuido, sin necesidad de experiencia en ingeniería ontológica.

- **Desventajas:**

- Posible pérdida de especificidades de cada lengua, salvo que se capturen en módulos específicos de la ontología, por ej. capa conceptual, o en el modelo lingüístico:, por ej. capa terminológica

### 3. Multilingualidad en ontologías: **realización**





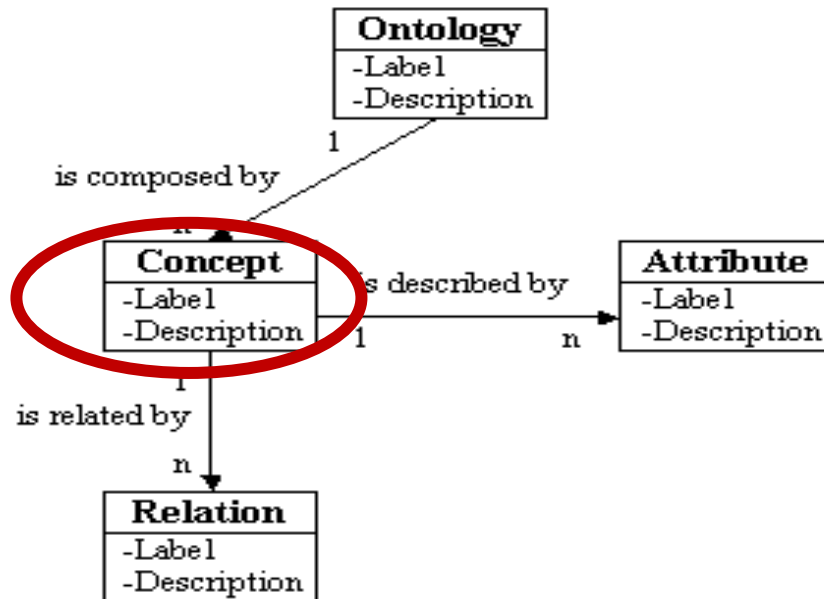
# Multilingualidad en ontologías

## 3. Realización

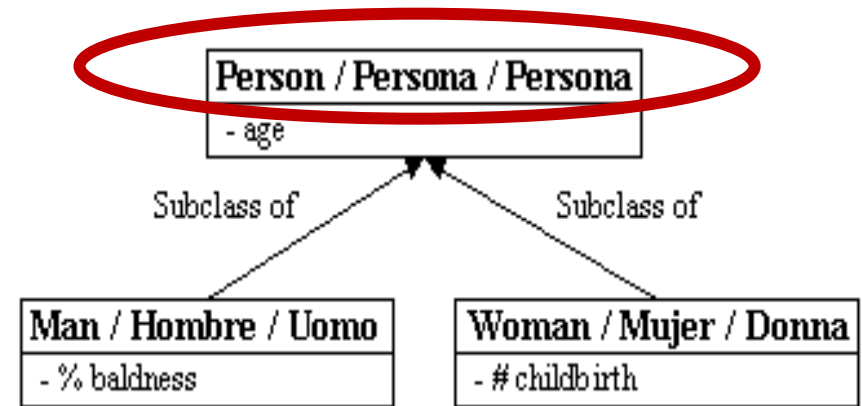
- Estrechamente relacionado con la modelización
- Se considera la instanciación del modelo
- Dos posibilidades:
  - Información lingüística **dentro** de la ontología
  - Información lingüística **fuera** de la ontología
    - BD relacional
    - BD terminológica
    - Lexicón multilingüe
    - Tesauro multilingüe

## 2. Realización

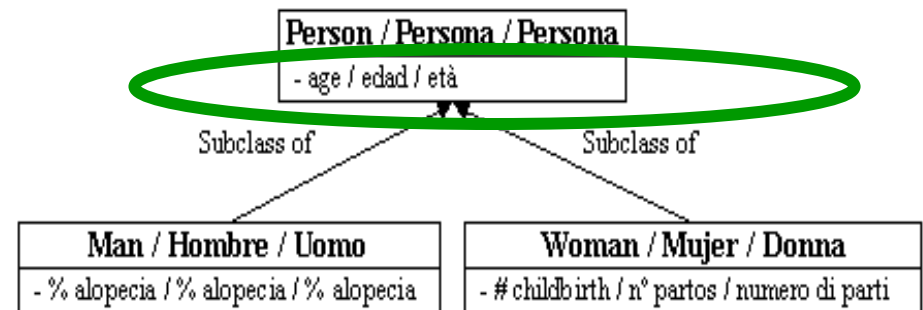
### Información lingüística **dentro** de la ontología (1)



Metamodelo de ontología



Multilingualidad en la ontología:  
**conceptos**, no atributos

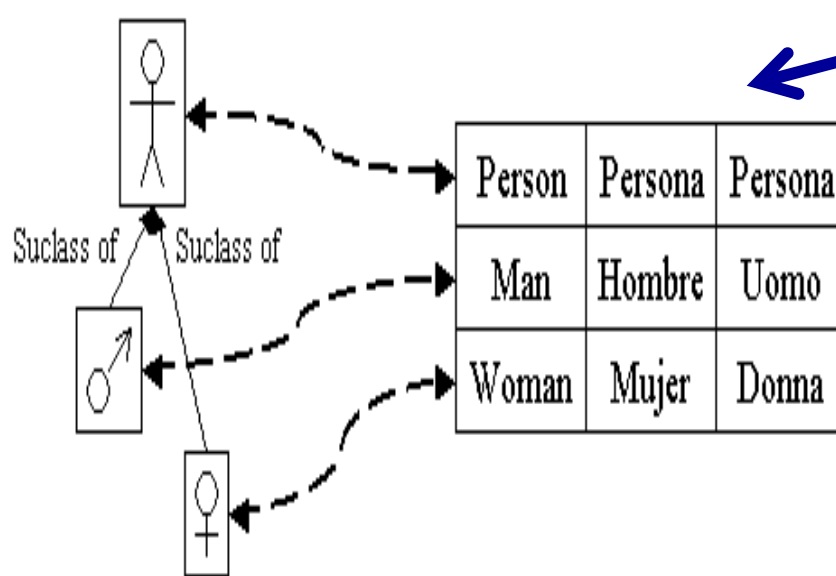


Mismo metamodelo de ontología

Multilingualidad en los **atributos**

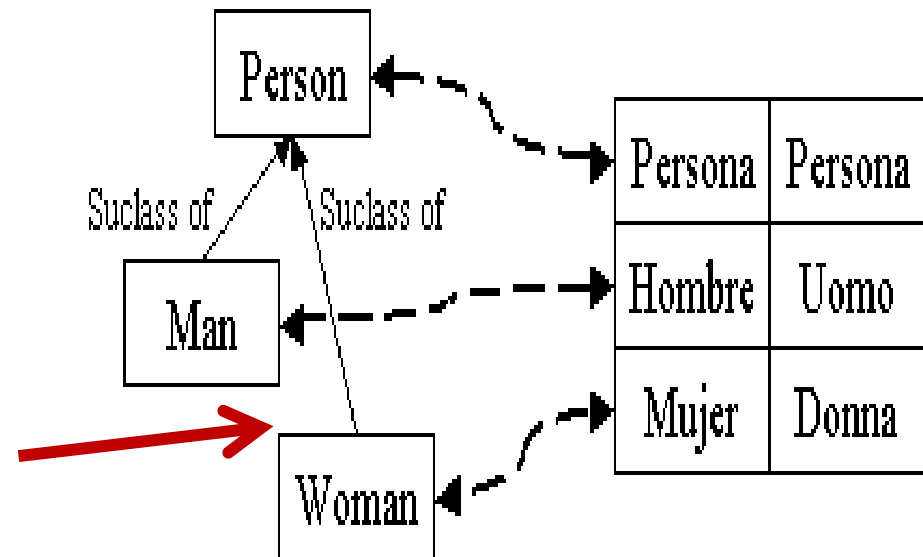
## 2. Realización.

### Información lingüística **fuera** de la ontología



A. Ontología **‘alingüe’**,  
modelo de recurso  
lingüístico  
Genoma KB

B. Ontología **monolingüe**  
Modelo de recurso  
lingüístico



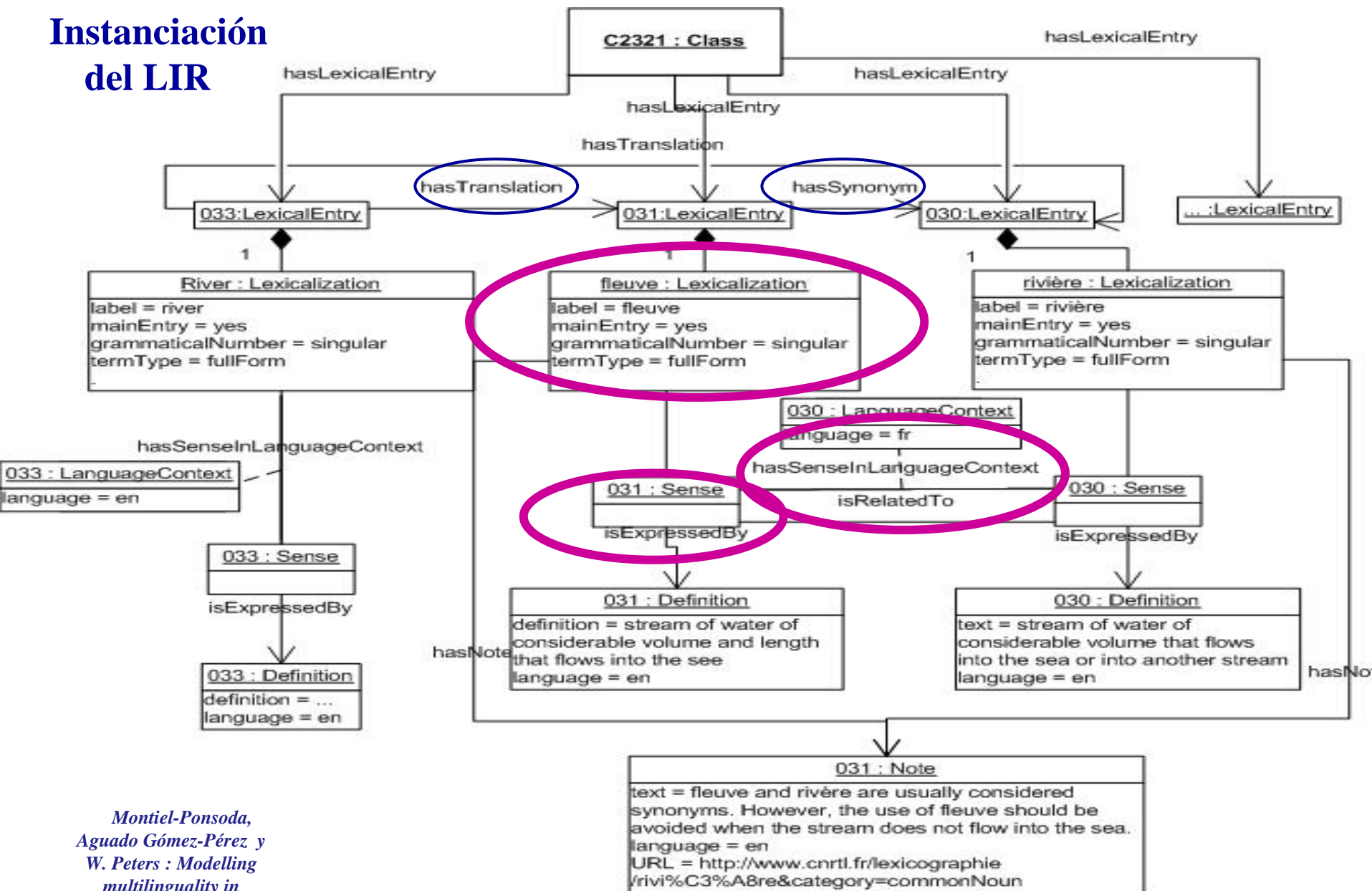
# Una nueva propuesta

## Linguistic Information Repository - LIR

- La información lingüística se organiza en torno a la `LexicalEntry` class.
- Una entrada léxica es una relación ternaria: `Lexicalization`, `Sense` and `LanguageContext`.
- Enlazando `Note` con `Sense` or `Definition`, se mantienen explícitas las diferencias de matices y significados en las lenguas
- `Note` se enlaza con `Lexicalization`, pero se podría enlazar con cualquier clase del modelo para aportar información complementaria.
- Modelado como una ontología.



# Instanciación del LIR



Montiel-Ponsoda,  
Aguado Gómez-Pérez y  
W. Peters : *Modelling  
multilinguality in  
ontologies: a holistic  
approach*, 2007

# Ventajas del LIR

- Preserva la **independencia** entre la ontología y la capa lingüística.
- **Enlaza** la información multilingüe con todos los elementos de la ontología.
- Facilita la recuperación e **integración** de información lingüística al adoptar estándares para describir rasgos lingüísticos.
- Con ello, facilita el **mantenimiento** de las especificidades lingüísticas
- Permite la **localización** al nivel conceptual y terminológico.
- Facilita la **interoperabilidad** y **extensibilidad**, si se requiere más información.
- Resuelve los desacuerdos o diferencias en la conceptualización
- **Permite el acceso** a recursos multilíngües gracias al **LabelTranslator**.

# Características del LabelTranslator

<http://droz.dia.fi.upm.es/lt/evaluation.htm>

- Plugin integrado en NeOn
- Mantiene la información lingüística
  - Modelo lingüístico (ML): *lexical entries, lexicalizations, senses, definitions, usage context, etc.*
  - Representado en OWL
- Acceso fácil al ML desde otros *plug-ins* y aplicaciones externas
  - *to search, add, remove or update.*
- Uso de un método sincronizado.
  - El usuario crea/importa una nueva ontología en OWL, FLogic, etc.
  - El LabelTranslator construye un modelo lingüístico (LIR) asociado a la ontología seleccionada
  - Mantiene la ontología y el conocimiento lingüístico *sincronizado*



# Guía para la localización

## TAREA 1.

- Memorias de traducción
- Glosarios
- Diccionarios
- Tesauros
- BD terminológicas, etc

## TAREA 2.

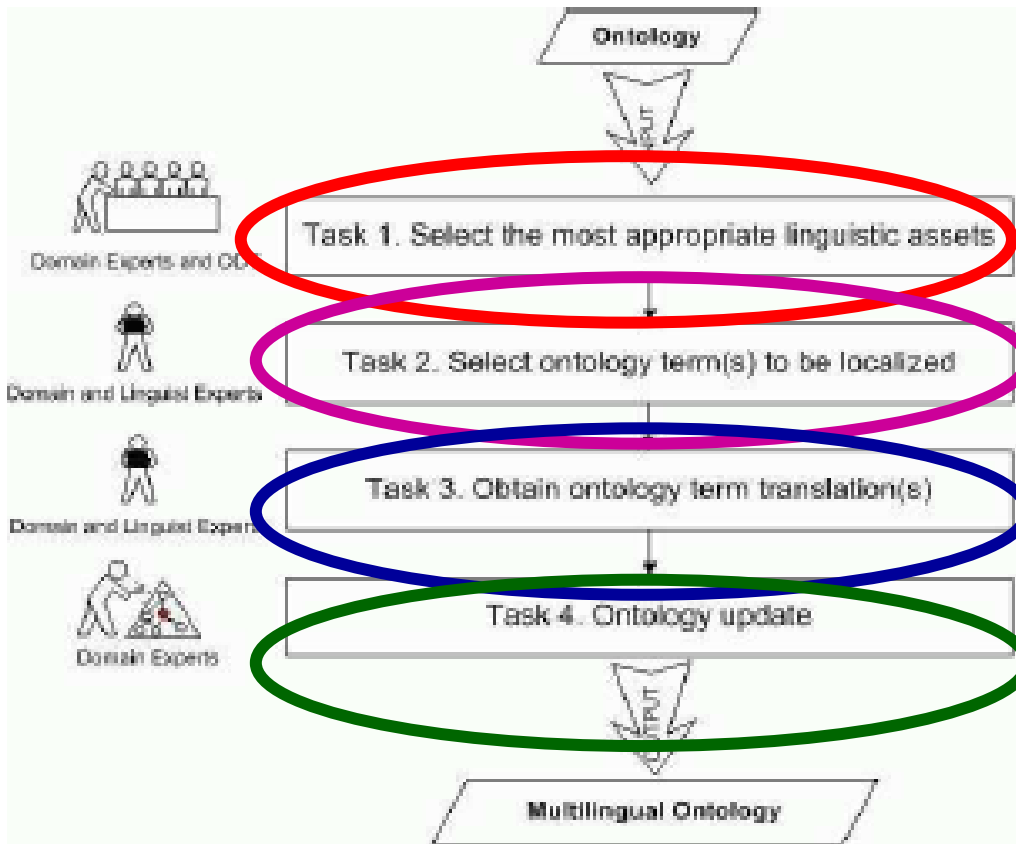
- Entrada: una ontología (toda o parte)
- Salida: términos de la ontología, que se van a localizar con información del contexto

## TAREA 3.

- Entrada: términos de la ontología, que se van a localizar.
- Salida: conjunto de traducciones para cada término de la ontología en la LM

## TAREA 4.

- Entrada: los términos traducidos.
- Salida: la ontología en la LM con información lingüística asociada a cada término.



# Objetivos de la práctica

- Objetivo general:
  - Evaluar la usabilidad y calidad de la traducción
- Objetivos específicos
  - Evaluar la calidad de la traducción de los términos de una ontología
  - Medir la satisfacción del usuario al usar la herramienta.
  - Evaluar la guía de traducción

<http://droz.dia.fi.upm.es/lt/evaluation.htm>

# Ontologías y multilingualidad

**Dra. Guadalupe Aguado de Cea**