



“Datos lingüísticos enlazados: introducción a LIDER y a lemon-Ontolex”

Jorge Gracia, Elena Montiel-Ponsoda

Ontology Engineering Group (OEG)

Universidad Politécnica de Madrid (UPM)

{jgracia, emontiel}@fi.upm.es

Taller sobre “Datos Enlazados y Multilingüismo para la
Lingüística y las Humanidades Digitales” Madrid, 20/10/2015



Recursos Lingüísticos



Recursos Lingüísticos Electrónicos

- corpus
- lexicones
- diccionarios
- Recurss enciclopédicos
- “Wordnets”
- terminologías
- etc.

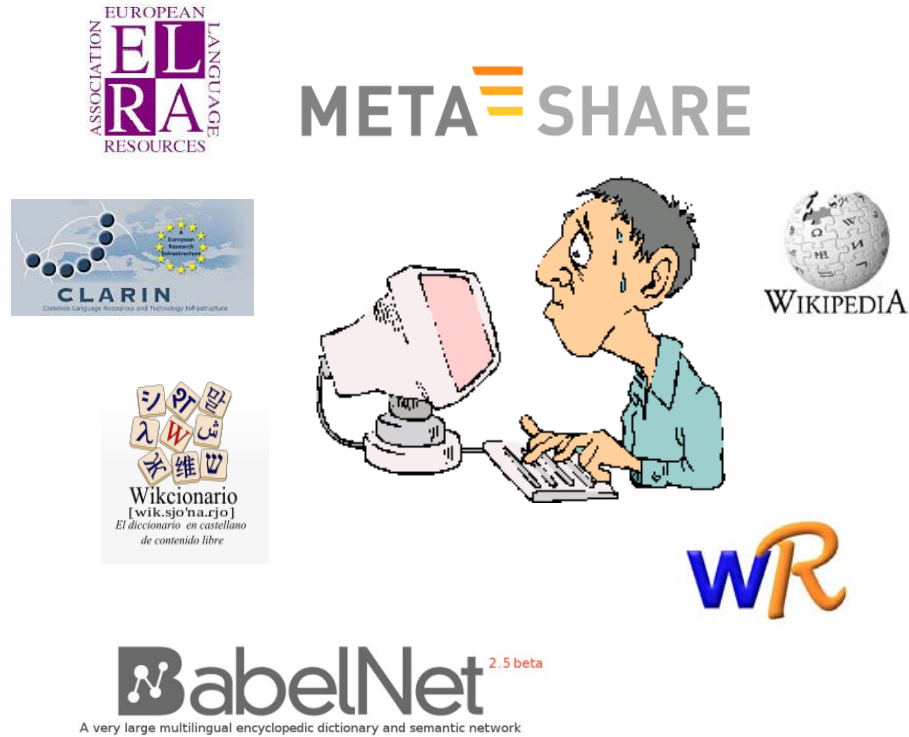


Actualmente...

- En formatos heterogéneos
- APIs no estándar
- Aislados entre sí
- Con diferentes niveles de acceso (desde “mándame un email” a servicios web)
- Múltiples catálogos de RLs con diferentes metadatos y esquemas de representación



Por tanto... El descubrimiento y reutilización de RLs resulta costoso en tiempo y esfuerzo





Wikcionario
[wik.sjo'na.rjo]
El diccionario en castellano
de contenido libre

<http://es.wiktionary.org>



REAL ACADEMIA
ESPAÑOLA



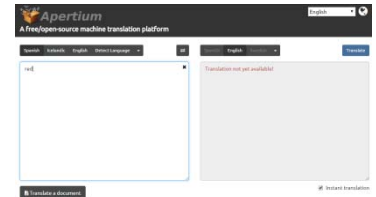
<http://rae.es>



“Red”
(computer
network)

Wikilengua del español

<http://www.wikilengua.org/index.php/Terminesp:red>



<http://www.apertium.org>



WIKIPEDIA

<http://es.wikipedia.org>





REAL ACADEMIA ESPAÑOLA

La institución Obras académicas Biblioteca y Archivo Consultas lingüísticas

Inicio » Recursos » Diccionarios » Diccionario de la lengua española

Diccionario de la lengua española

El *Diccionario de la lengua española (DRAE)* es la obra de referencia de la Academia. La edición actual —la 22.ª, publicada en 2001— incluye más de 88 000 entradas.

red

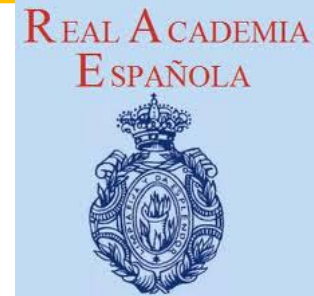
á é í ó ú ü ñ

■ Ayuda

red. Artículo enmendado

(Del lat. *rete*).

1. f. Aparajo hecho con hilos, cuerdas o alambres trabados en forma de mallas, y convenientemente dispuesto para pescar, cazar, cercar, sujetar, etc.
2. f. Labor o tejido de mallas.
3. f. [redecilla](#) (ll prenda de malla para el pelo).
4. f. Lugar donde se vende pan u otras cosas que se dan por entre verjas.





Wikcionario
[wik.sjo'na.rjo]
El diccionario en castellano
de contenido libre

<http://es.wiktionary.org>



REAL ACADEMIA
ESPAÑOLA



red

- **Pronunciación:** [red] (AF1)
- **Etimología:** del latín *rete*

Sustantivo femenino [editar]

Singular	Plural
red	redes

1

Malla cosa que fue creada para unir o sostener algo

2

Sistema interconectado

- **Ejemplos:** (...)

Véase también [editar]

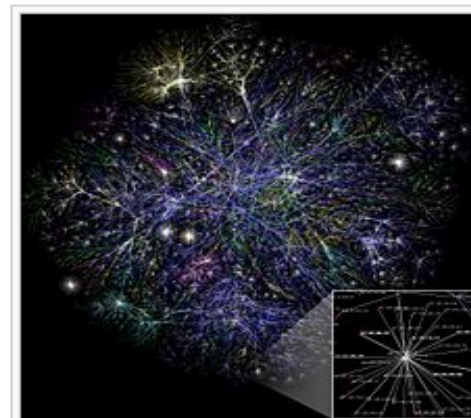
- [Wikipedia](#) tiene un artículo sobre **red**.

Traducción [editar]

- **Alemán:** [Netz](#) ^(de)



[1] malla Raschel



[2] la red en 2005





<http://es.wiktionary.org>



Terminesp:red

Wikilengua

Entrada de Terminesp

Este artículo es una entrada de la base terminológica [Terminesp](#). No se puede modificar, pero se pueden hacer comentarios y puede servir de base para un artículo en la *Wikilengua*.

Búsqueda en Terminesp

Norma UNE

[Norma UNE \(Terminesp\)::UNE 21302-131](#)

Título de la norma

Vocabulario electrotécnico: parte 131. Teoría de circuitos

Término en español:

red ([artículo](#))

Término en alemán:

Netzwerk (in der Netzwerktopologie)

Término en inglés:

network

Término en francés:

réseau

Definición

En topología de redes, conjunto de elementos ideales de circuito y de sus interconexiones, considerado como un todo.

<http://es.wiktionary.org>



Apertium

A free/open-source machine translation platform

English



Spanish

Icelandic

English

Detect Language



Spanish

English

Swedish



Translate

red



Network

<http://www.apertium.org>





<http://es.wiktionary.org>



Red de computadoras

(Redirigido desde «Redes de computadoras»)

Una **red de computadoras**, también llamada **red de ordenadores**, **red de comunicaciones de datos** o **red informática**, es un conjunto de **equipos informáticos** y **software** conectados entre sí por medio de **dispositivos físicos** que envían y reciben **impulsos eléctricos**, **ondas electromagnéticas** o cualquier otro medio para el transporte de **datos**, con la finalidad de compartir información, recursos y ofrecer **servicios**.¹

Como en todo proceso de **comunicación** se requiere de un **emisor**, un **mensaje**, un **medio** y un **receptor**. La finalidad principal para la creación de una red de computadoras es compartir los recursos y la información en la distancia, asegurar la **confiabilidad** y la disponibilidad de la información, aumentar la **velocidad de transmisión de los datos** y reducir el costo general de estas acciones.² Un ejemplo es **Internet**, la cual es una gran red de millones de computadoras ubicadas en distintos puntos del planeta interconectadas básicamente para compartir información y recursos.

La estructura y el modo de funcionamiento de las redes informáticas actuales están definidos en varios **estándares**, siendo el más importante y extendido de todos ellos el modelo **TCP/IP** basado en el modelo de referencia **OSI**. Este último, estructura cada red en siete capas con funciones concretas pero relacionadas entre sí; en TCP/IP se reducen a cuatro capas. Existen multitud de protocolos repartidos por cada capa, los cuales también están regidos por sus respectivos estándares.³

Índice [ocultar]

- 1 Historia
- 2 Descripción básica
- 3 Componentes básicos de las redes
 - 3.1 Software
 - 3.2 Hardware
 - 3.2.1 Tarjeta de red
 - 3.3 Dispositivos de usuario final
 - 3.3.1 Servidores

<http://www.apertium.org>



<http://es.wikipedia.org>





“Red”
Pronunciation: [red]
Grammar category: sustantivo femenino
Singular: “red”
Plural: “redes”



“Red”
Etimology: Del latín “rete”
Gender: “f”
Definition: “Conjunto de ordenadores o de equipos informáticos conectados entre sí...”

Información complementaria pero no conectada

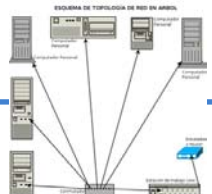
Wikilengua del español

“Red”
Norm: UNE 21302-131
English: network
German: Netzwerk



“Red”
translations: “xarxa”(ca), “rede”(ga), ...

“Red_de_computadores”
Category: redes informáticas
Image



Datos Enlazados

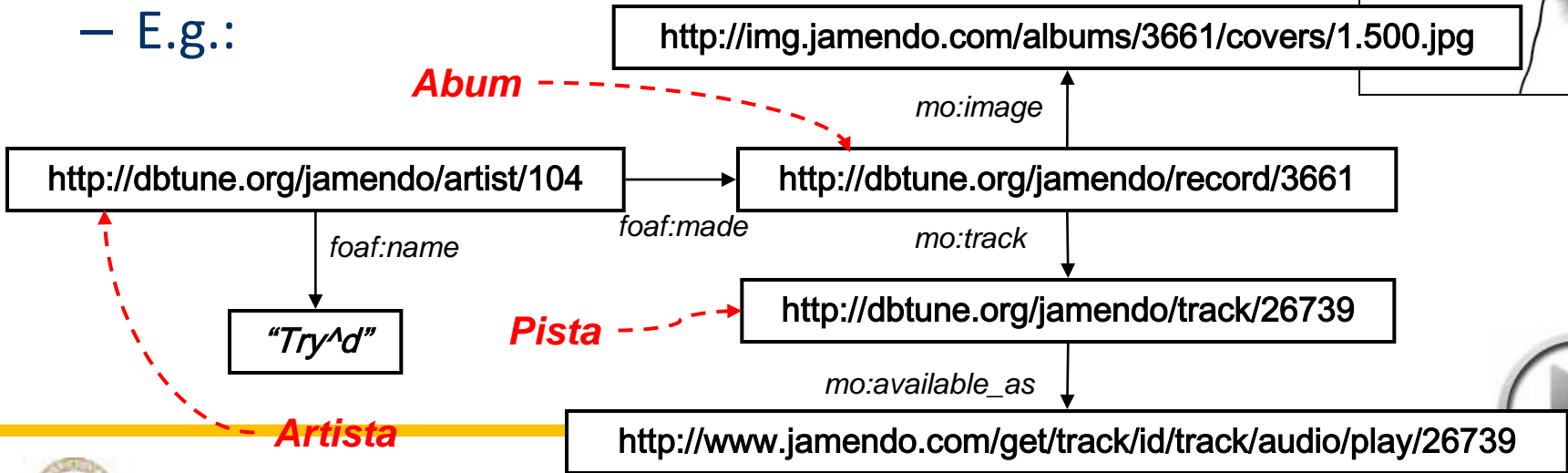


URIs (Uniform Resource Identifiers)

- Identifican recursos en la Web, tanto recursos de “información” como recursos del “mundo real”

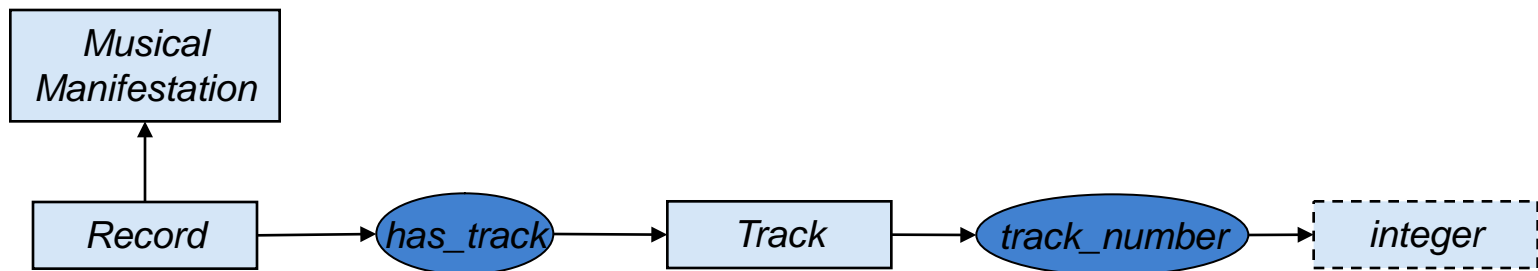
RDF (Resource Description Framework)

- Expresa relaciones entre recursos
- *subject -predicate -object*
- E.g.:



Ontologías Web (OWL , RDF-S)

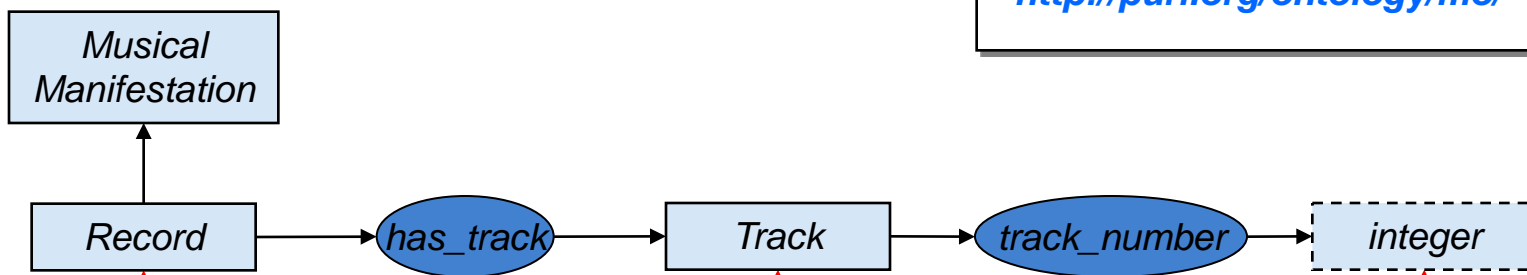
- Clases de objetos
- Propiedades, o relaciones ente clases
- Individuos, o instancias de esas clases
- Permite usar **razonadores** semánticos
- E.g.: from MUSIC ontology <http://purl.org/ontology/mo/>



Encajando las piezas

MUSIC ontology (OWL)

<http://purl.org/ontology/mo/>



<http://dbtune.org/jamendo/record/3661>

<http://dbtune.org/jamendo/track/26739>

4

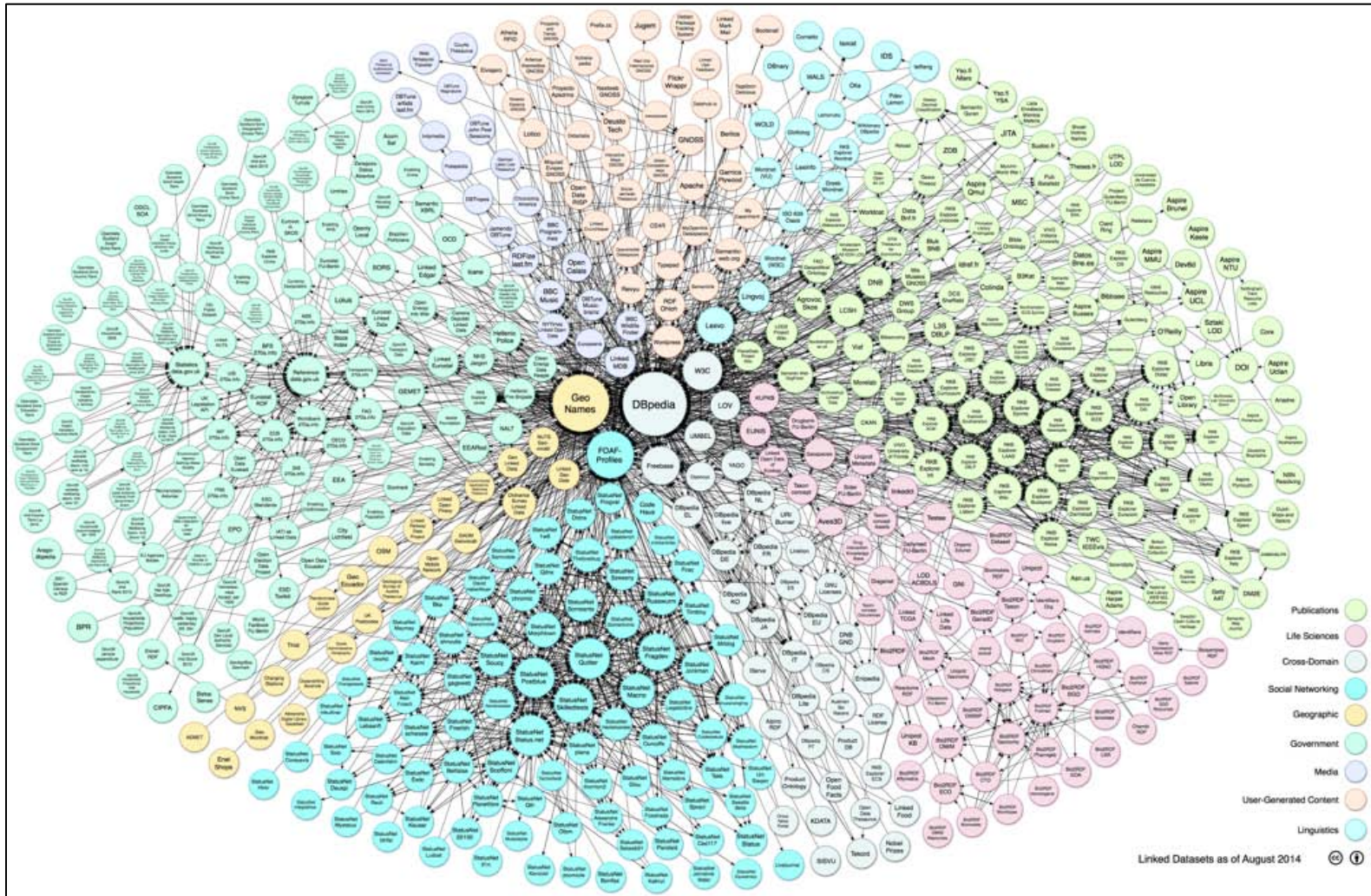
mo:has_track

mo:track_number

JAMENDO data (RDF)

<http://dbtune.org/jamendo>





Data accessible in a standardised way (SPARQL):

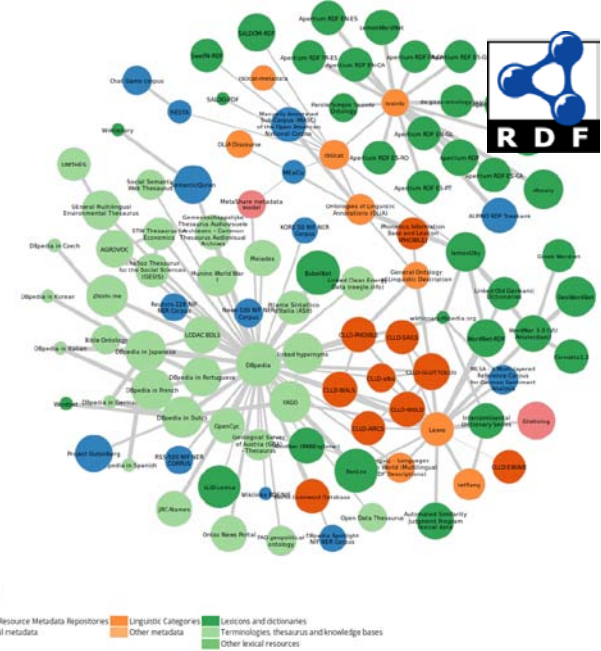
- *“Give me the title of the albums made by Tryad along with their publication date”*

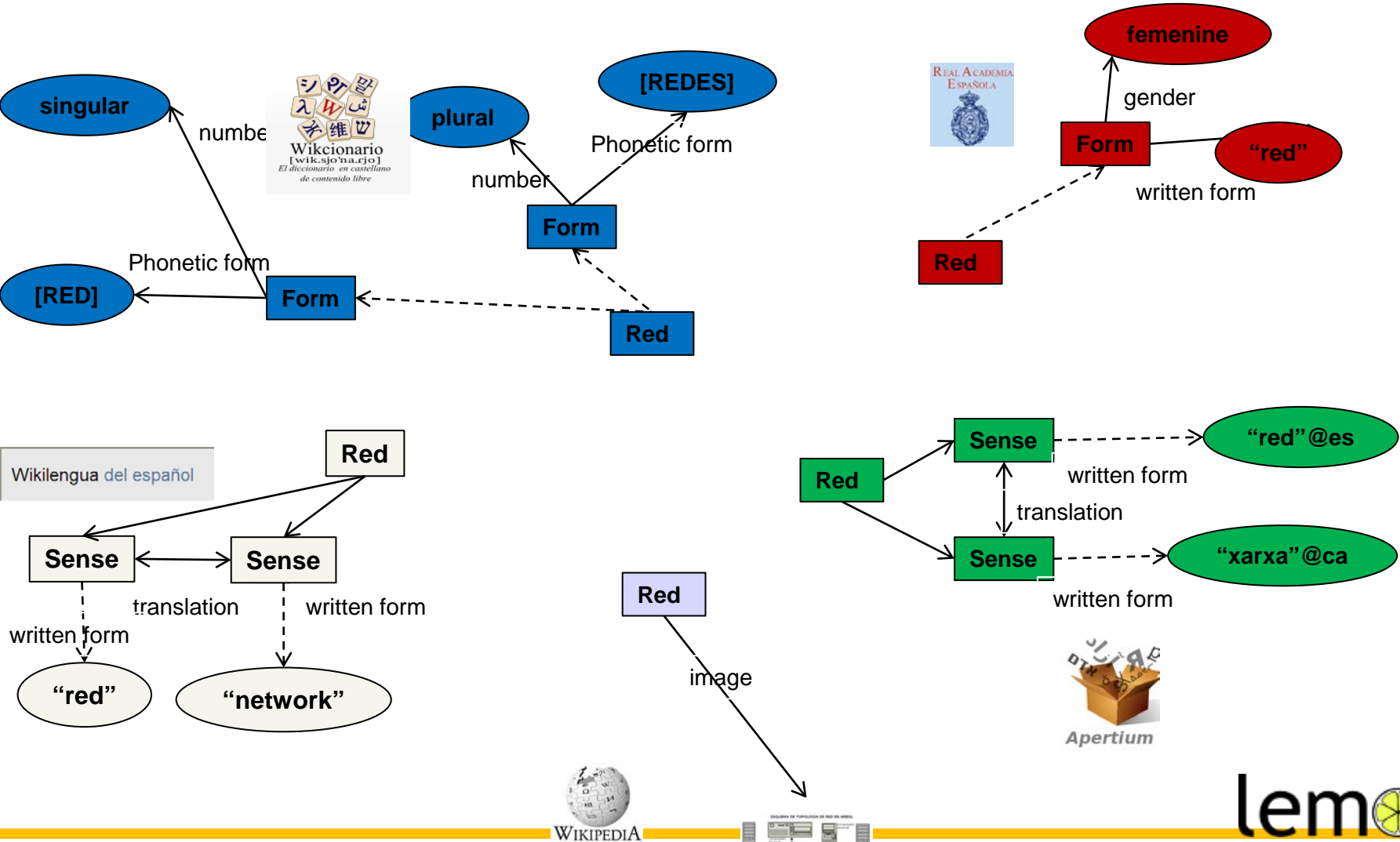
```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT DISTINCT ?title ?date
WHERE
{ ?artist rdf:type mo:MusicArtist;
  foaf:made ?album;
  foaf:name "t r y ^ d";
  ?album dc:title ?title;
  dc:date ?date }
    
```

Title	Date
“Listen”	29-12-2006
“Public Domain”	20-01-2006
“Tryad Demo (Public Domain)”	14-04-2005

Linked Data permite la interoperación entre recursos lingüísticos

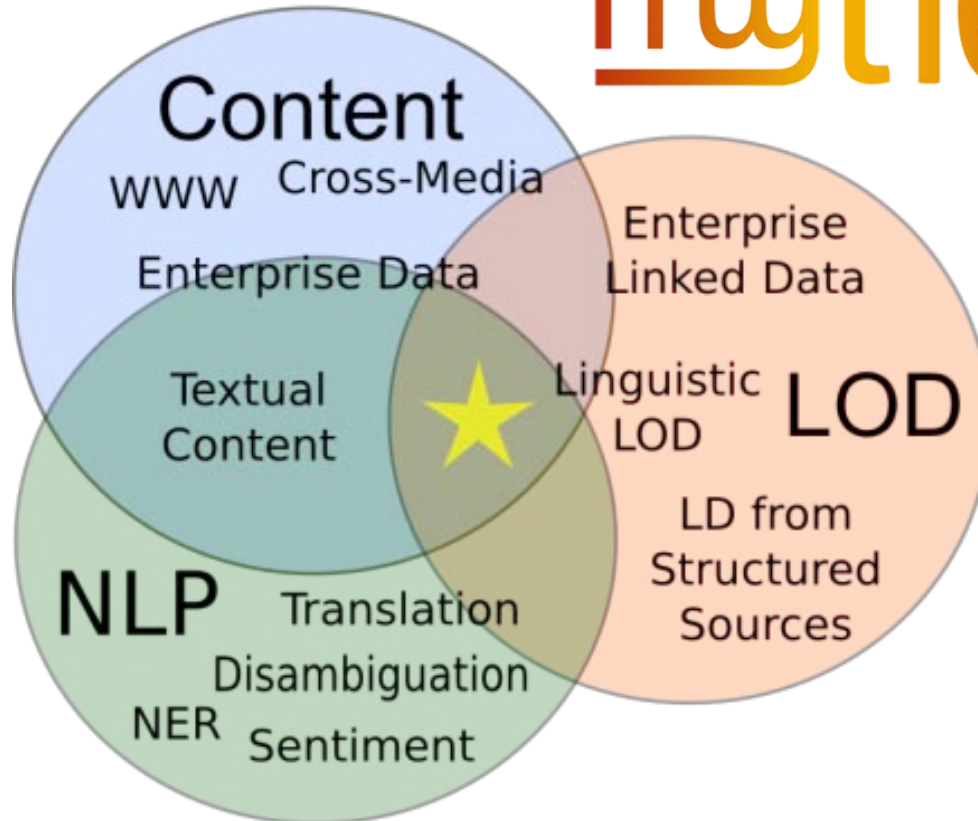




El proyecto LIDER



Linked Data as an enabler of cross-media and multilingual **content analytics** for enterprises across Europe

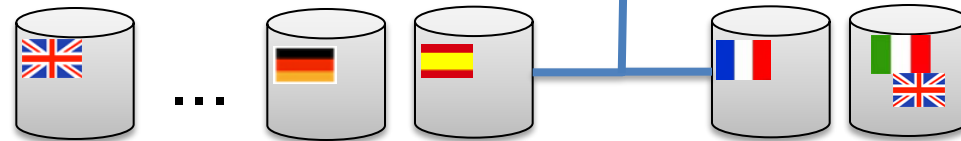
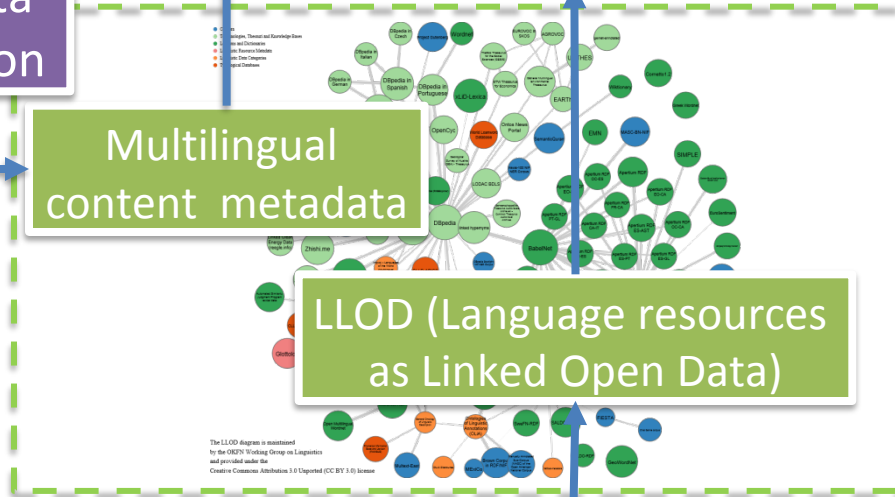




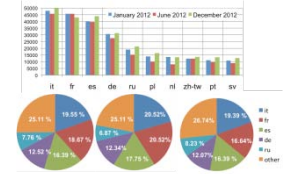
Content providers



Content Analytics consumers



Language resources (lexicon, corpora,...)





Certification

Benchmarking & Validation

Discovery

LLD Linking

Service Composition

LLD Publishing

LLD-aware Services

Vocabularies

Hosting

Scalability

Streaming

Interoperability

Metadata

Licensing

Provenance

Multilingual Data

Guidelines and Standardization



www.lider-project.eu



Join the community

“Linked Data for Language Resources”

www.w3c.org/community/ld4lt



twitter.com/multilingweb

Hashtag: #LiderEU

OntoLex-lemon

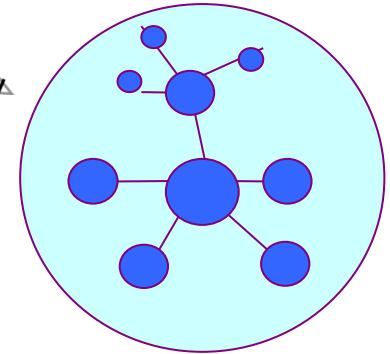


¿Qué es?

Modelo de representación de
información lingüística
multilingüe con respecto a
 una ontología



OntoLex

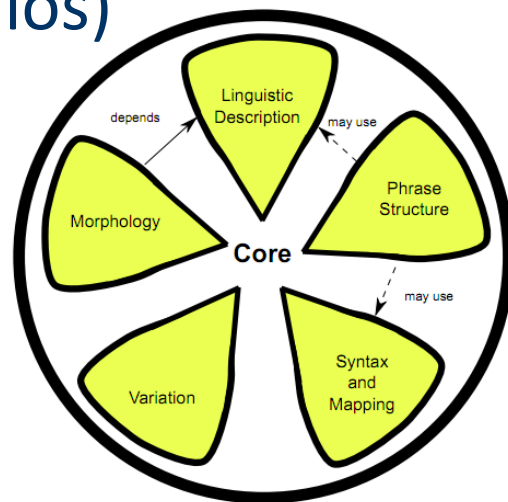


¿Qué es?

- Modelo de representación léxica en **RDF(s)** (la semántica viene definida en la ontología)

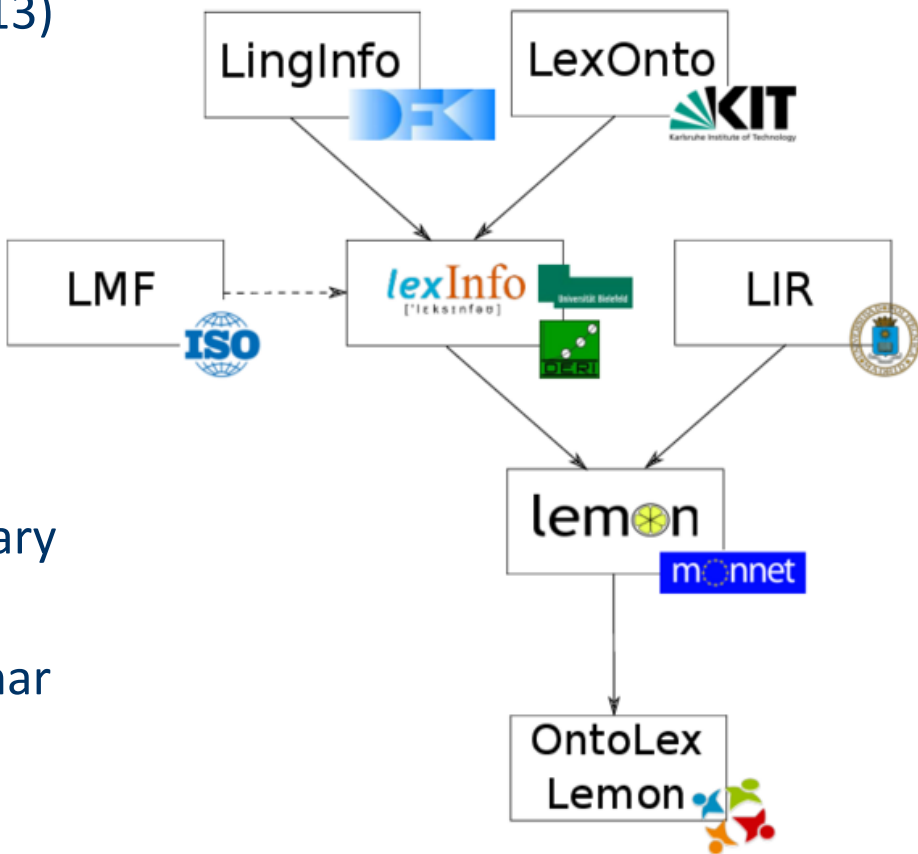
“An ontology-based semantic lexicon would leave the semantics to the ontology, focusing instead on providing domain-specific terms and object descriptions in the ontology.” (Buitelaar, 2010)

- Conciso y descriptivo (apunta a repositorios externos)
- Modular y extensible (consta de 5 módulos)



Orígenes

- Lexical Markup Framework (ISO 24613)
 - XML
- LexInfo, LIR
 - Represent lexical information relative to an ontology
 - OWL
- SKOS (W3C Standard)
 - Designed for Taxonomy/Vocabulary representation in RDF
- Modelo de referencia para transformar recursos en linked data
(Ver <http://linghub.lider-project.eu/>)



From McCrae, J. , “lemon: The Lexicon Model for Ontologies SD-LLOD-15

¿Qué necesidad intentamos cubrir?

Escasez de información lingüística en ontologías

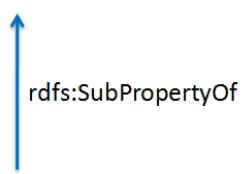
- Información morfosintáctica
- Variación terminológica
- Traducciones
- Metadatos

RDF(S), OWL

dct:creator →^{rdfs:label} “Creator”@en

SKOS

dct:creator →^{skos:prefLabel} “Creator”@en



Ejemplo 1: “author” – escasez de info léxica

En la ontología DBpedia, es una propiedad con este identificador:

<http://dbpedia.org/ontology/author>

Y existe en la Web con otros identificadores también:

<http://rdvocab.info/roles/author>

<http://d-nb.info/standards/elementset/gnd#author>

...



Gracias a OWL podemos

- describir dicha entidad con propiedades y relaciones
- relacionar dicha entidad con entidades en otras ontologías
- razonar para inferir equivalencias
- ...



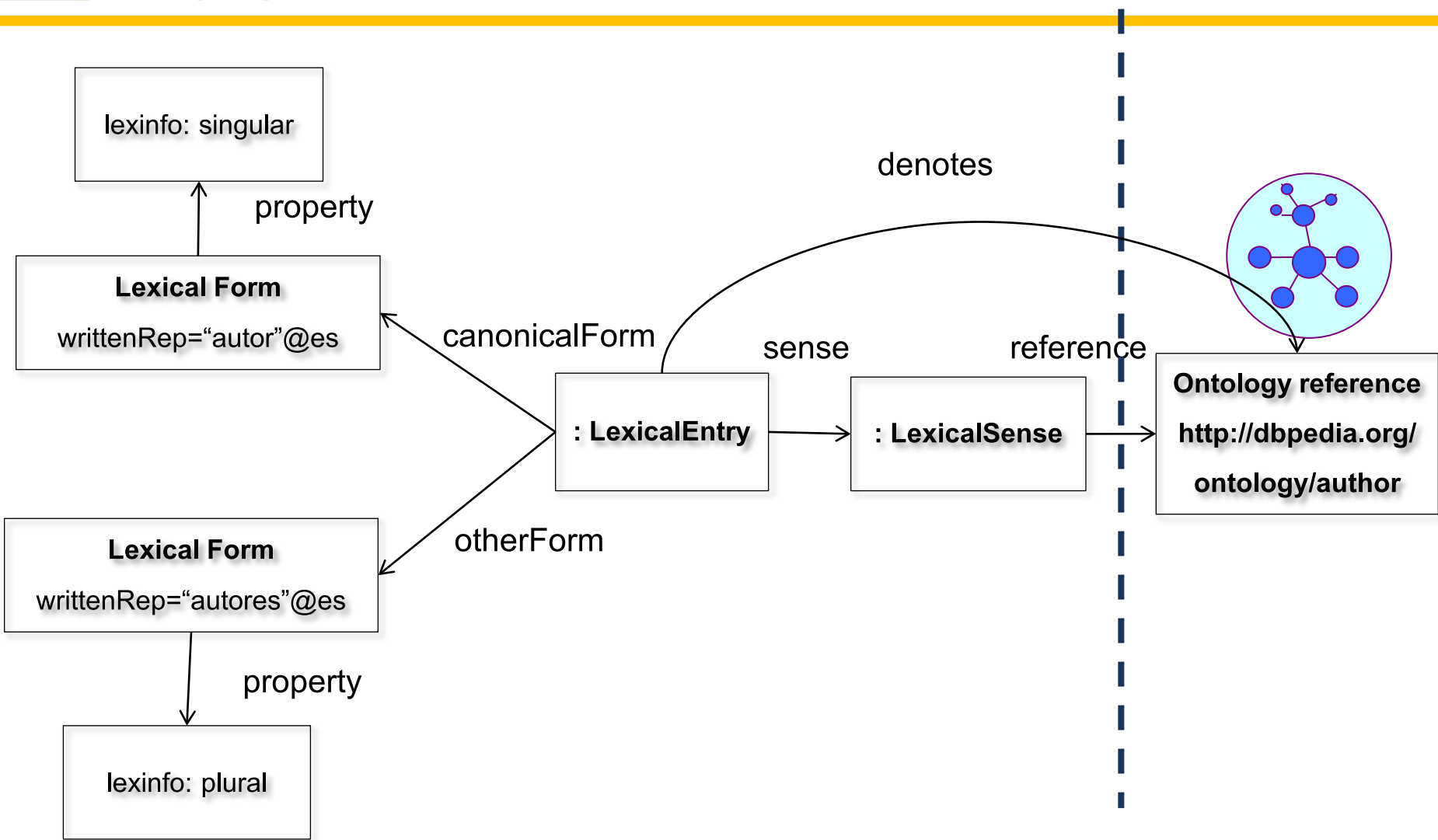
Pero... la misma entidad se puede designar de muchas maneras

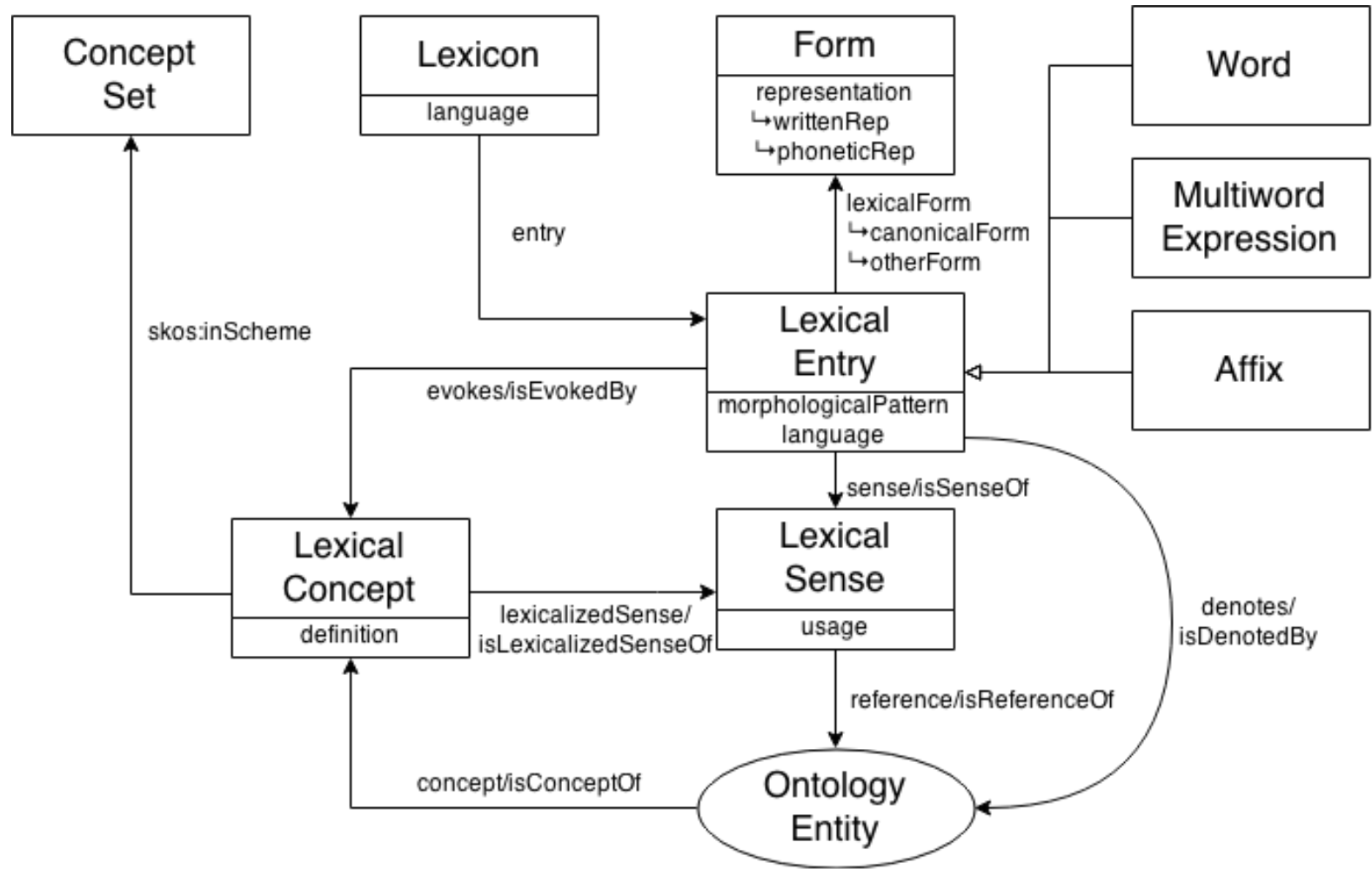
“autora”, “autores”, “authors”, “Verfasserin”

Y OWL/RDF(s) tiene ciertas limitaciones...

Así que necesitamos que las palabras sean entidades!!







Ejemplo 2: “escritor” – **variación denominativa**

En DBpedia es una clase:

<http://dbpedia.org/ontology/Writer>

Pero en la mayoría de vocabularios aparece como

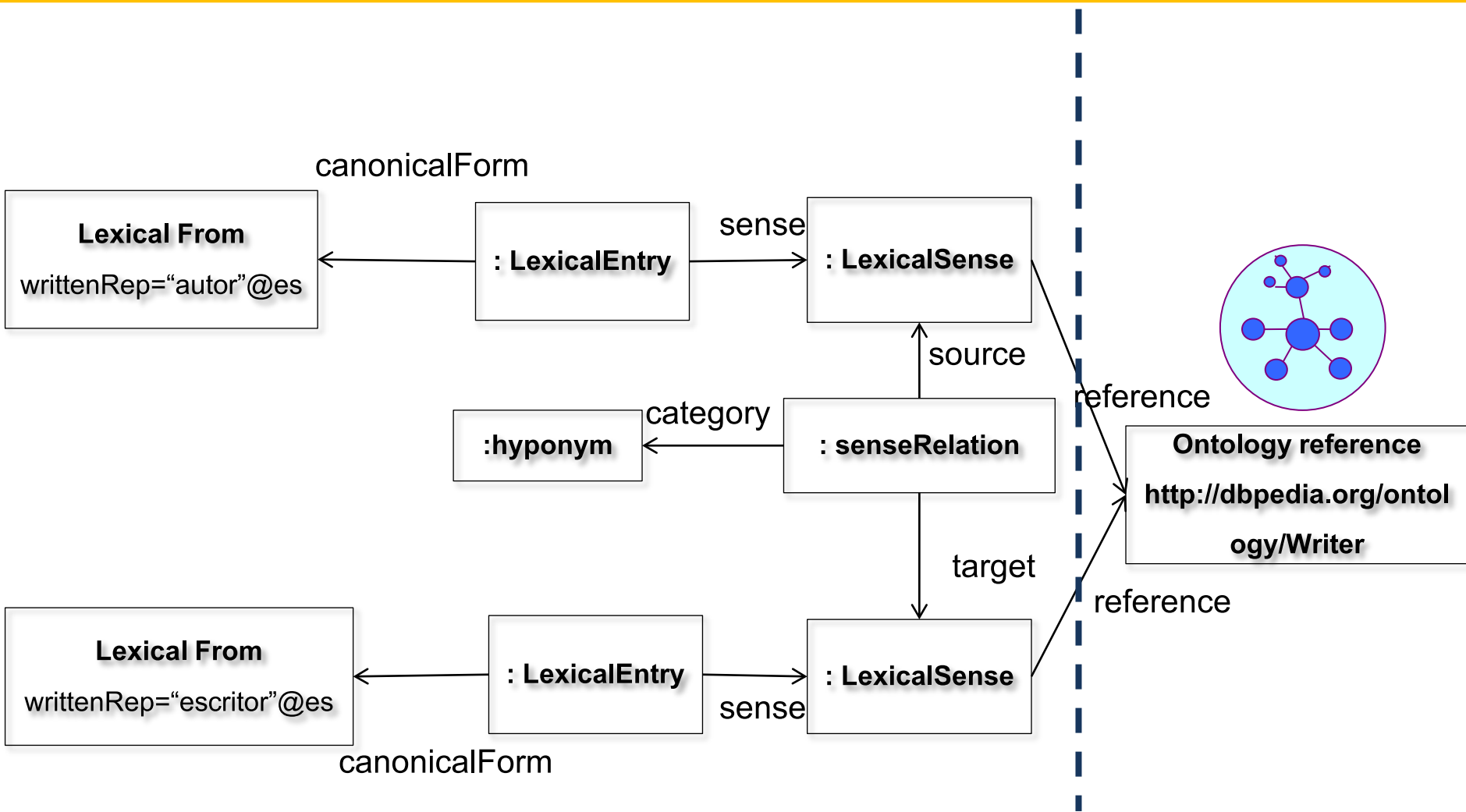
“autor” y es una propiedad:

<http://rdvocab.info/roles/authorWork>

<http://d-nb.info/standards/elementset/gnd#author>

Queremos dar cuenta del uso de autor y escritor como sinónimos (o incluso hipónimos) para facilitar enlazado con otros recursos

Según <http://lov.okfn.org/>: writer 59 vs. Author 8.871

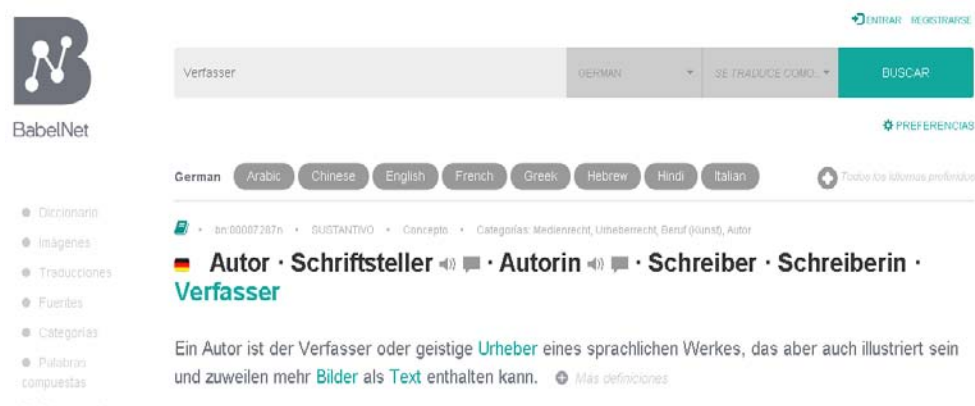


Ejemplo 3: “Verfasser” - traducciones

En DNB es una propiedad cuya etiqueta es “Verfasser”:

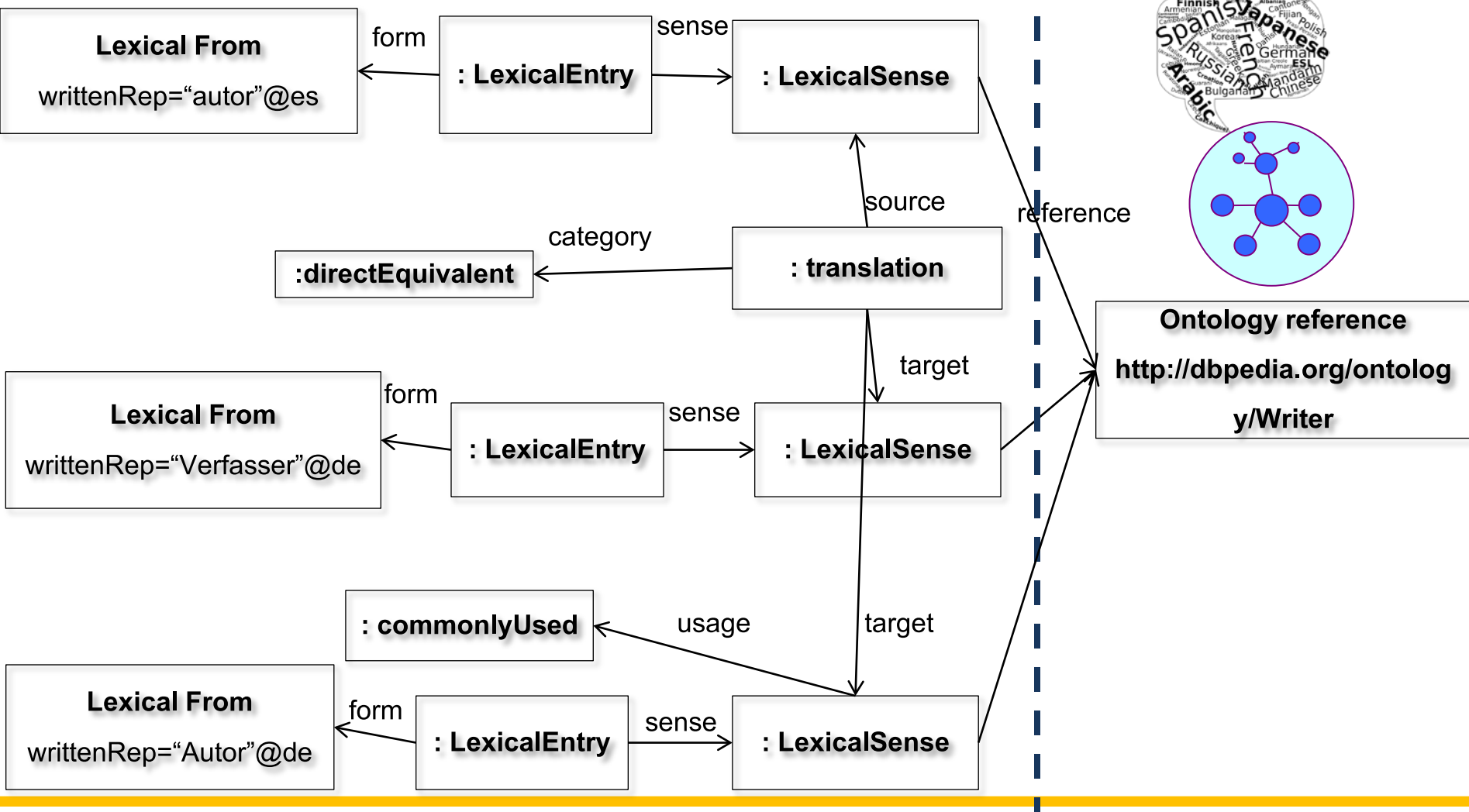
<http://d-nb.info/standards/elementset/gnd#author>

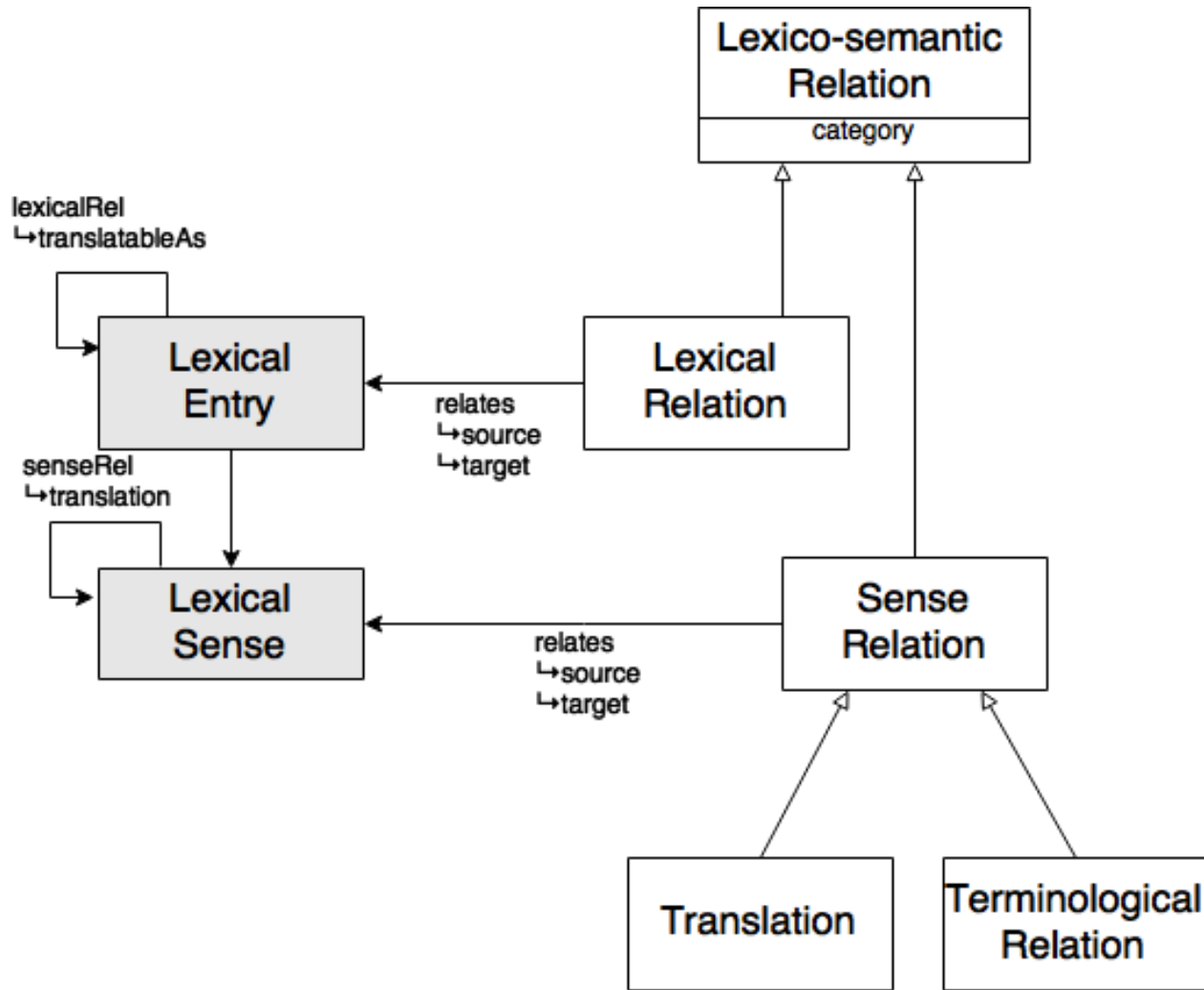
¿Cuál sería la mejor traducción?



The screenshot shows the BabelNet interface. At the top left is the BabelNet logo. Below it is a navigation menu with options: Diccionario, Imágenes, Traducciones, Fuentes, Categorías, and Palabras compuestas. The main search area has a search bar containing 'Verfasser', a language dropdown set to 'GERMAN', and a 'BUSCAR' button. Below the search bar are language tabs for German, Arabic, Chinese, English, French, Greek, Hebrew, Hindi, and Italian. The search results for 'Verfasser' are displayed, showing the term in German and its translations: Autor, Schriftsteller, Autorin, Schreiber, and Schreiberin. A definition in German is provided: 'Ein Autor ist der Verfasser oder geistige Urheber eines sprachlichen Werkes, das aber auch illustriert sein und zuweilen mehr Bilder als Text enthalten kann.' There are also links for 'Más definiciones'.

Queremos añadir info sobre frecuencia de uso en las traducciones, etc.





Ejemplo 4: “autor” – subcategorización de verbos, nombres y adjetivos

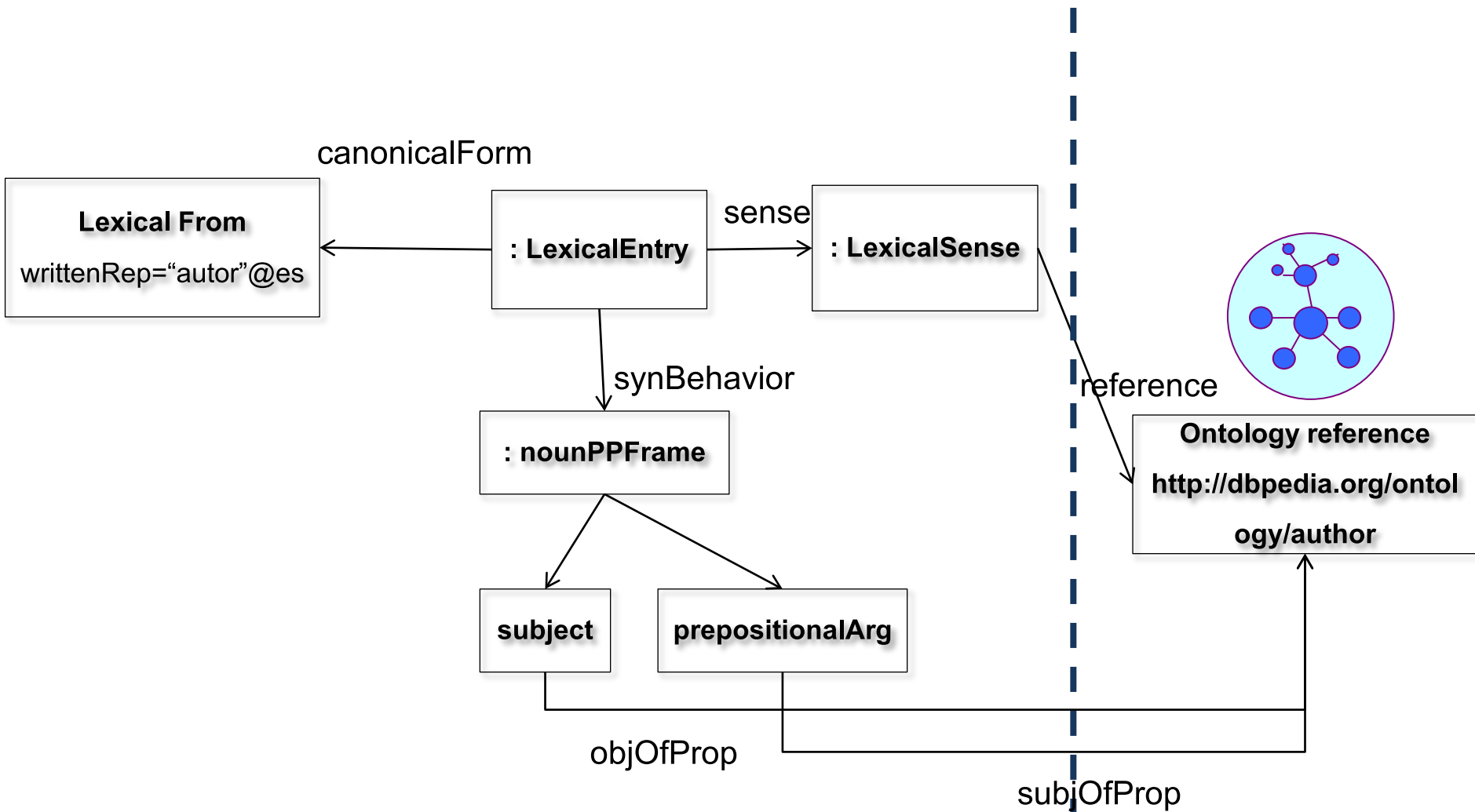
La propiedad de DNB <http://d-nb.info/standards/elementset/gnd#author> es un nombre relacional, que se comporta de la siguiente manera:

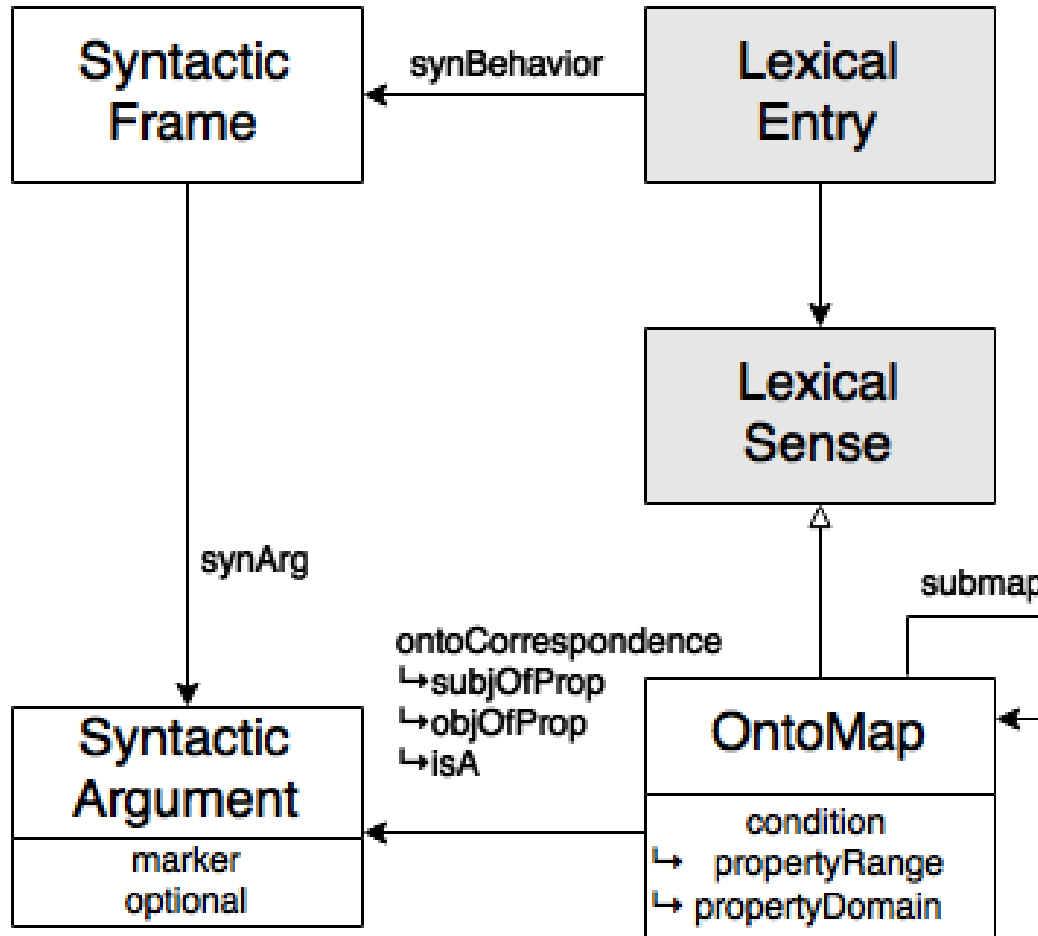
“Cervantes **es autor de** obras pertenecientes a...”

“Cervantes **es el autor de** Don Quijote”

Queremos dar cuenta del comportamiento sintáctico de dicho nombre, así como de los verbos que describen muchas propiedades o relaciones







Resumen limitaciones:

1. Describir morfosintácticamente entradas léxicas (las palabras adquieren la condición de entidad)
2. Capturar matices de significado, especificidades de una lengua o cultura
3. Dar cuenta de la variación denominativa y de las traducciones
4. Dar cuenta del comportamiento sintáctico (subcategorización) de las entradas léxicas