



DH@Madrid Summer School 2016

La nube de datos lingüísticos (multilingües) enlazados y *ontolex-lemon*

Elena Montiel Ponsoda

emontiel@fi.upm.es

Agradecimientos a Jorge Gracia

Madrid, 29 junio de 2016

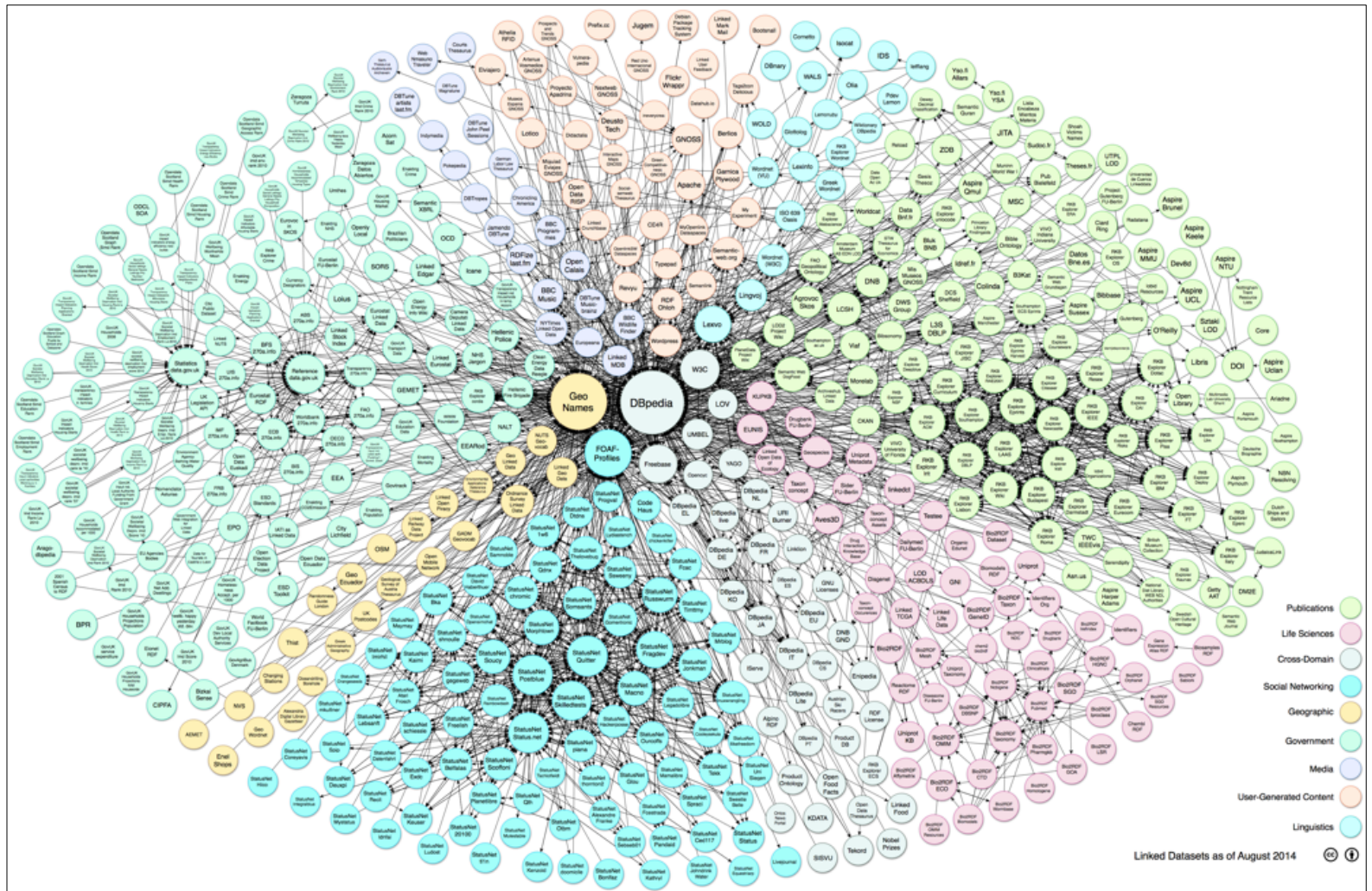


European
Research
Council



LLOD - Linguistic Linked Open Data: la nube de datos lingüísticos enlazados

... LOD cloud



Banque de datos lingüísticos enlazados



Motivación



<http://es.wiktionary.org>



"Red"
(computer network)



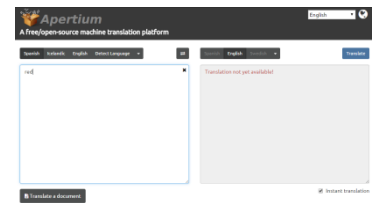
Wikilengua del español

http://www.wikilengua.org/index.php/Terminos_p:red



<http://es.wikipedia.org>

<http://rae.es>



<http://www.apertium.org>



"Red"

Pronunciation: [red]

Grammar category: sustantivo femenino

Singular: "red"

Plural: "redes"

Wikilengua del español

"Red"

Norm: UNE 21302-131

English: network

German: Netzwerk

**Información
complementaria
pero no
conectada**

"Red"

**Etimology Del latin
"rete"**

Gender: "f"

**Definition.: "Conjunto de
ordenadores o de equipos
informáticos conectados
entre sí...."**



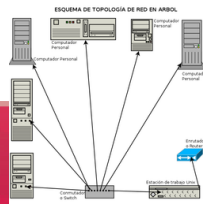
"Red"

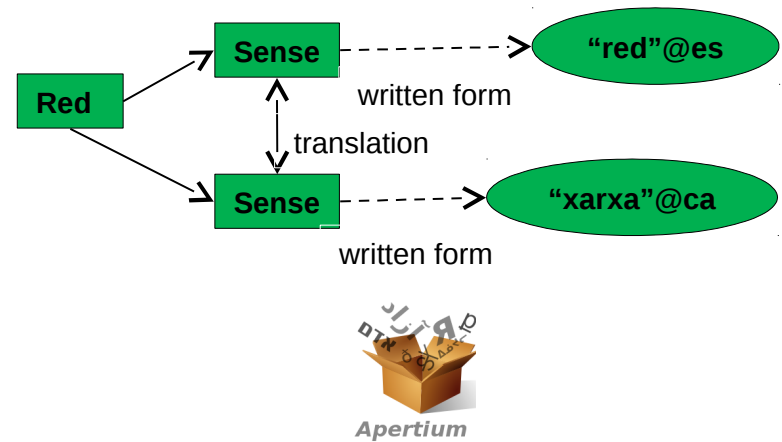
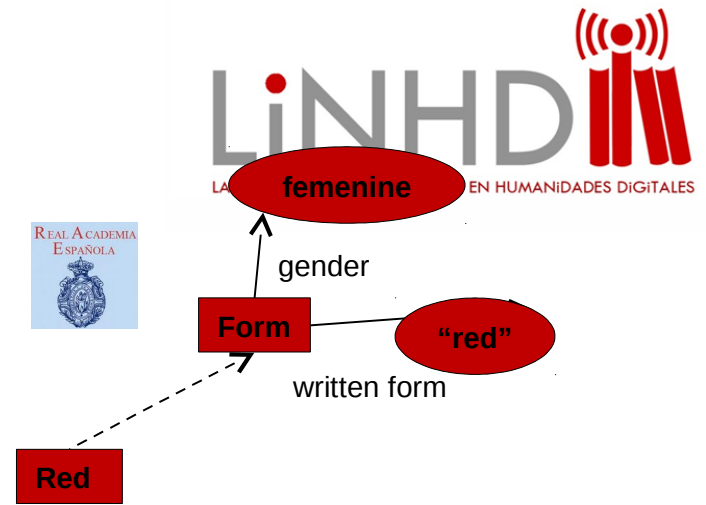
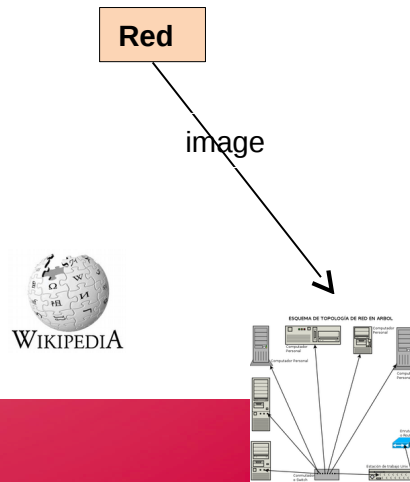
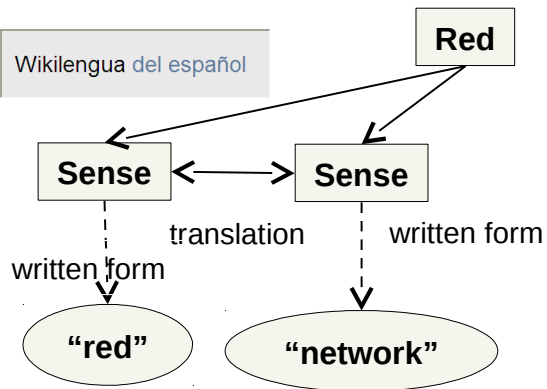
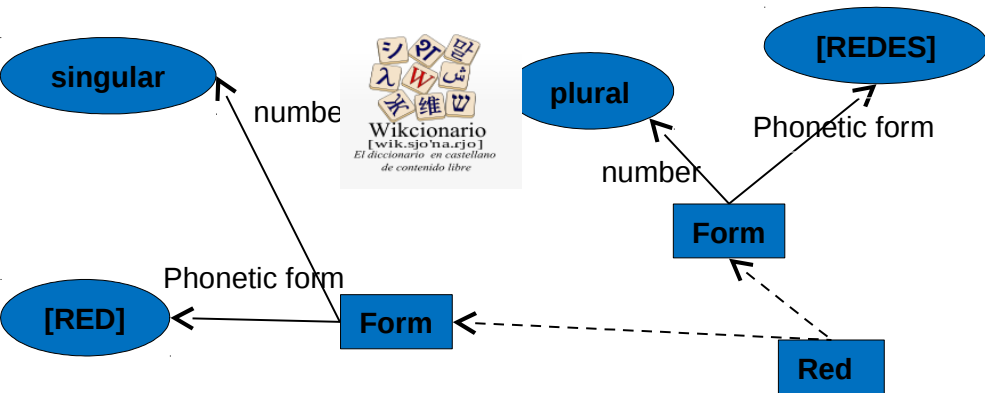
translations: "xarxa"(ca), "rede"(ga), ...

**"Red_de_computador
es"**

Category: redes informáticas

Image





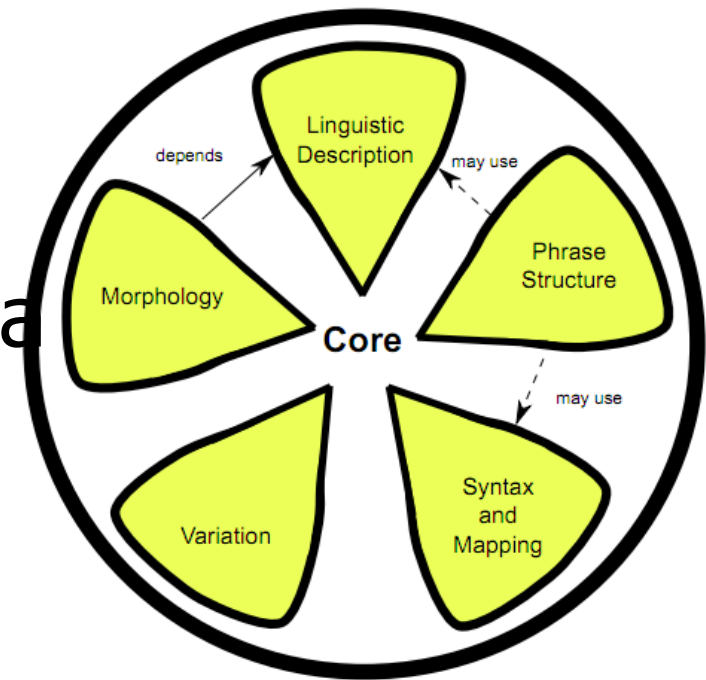
lemon

lemon-ontolex



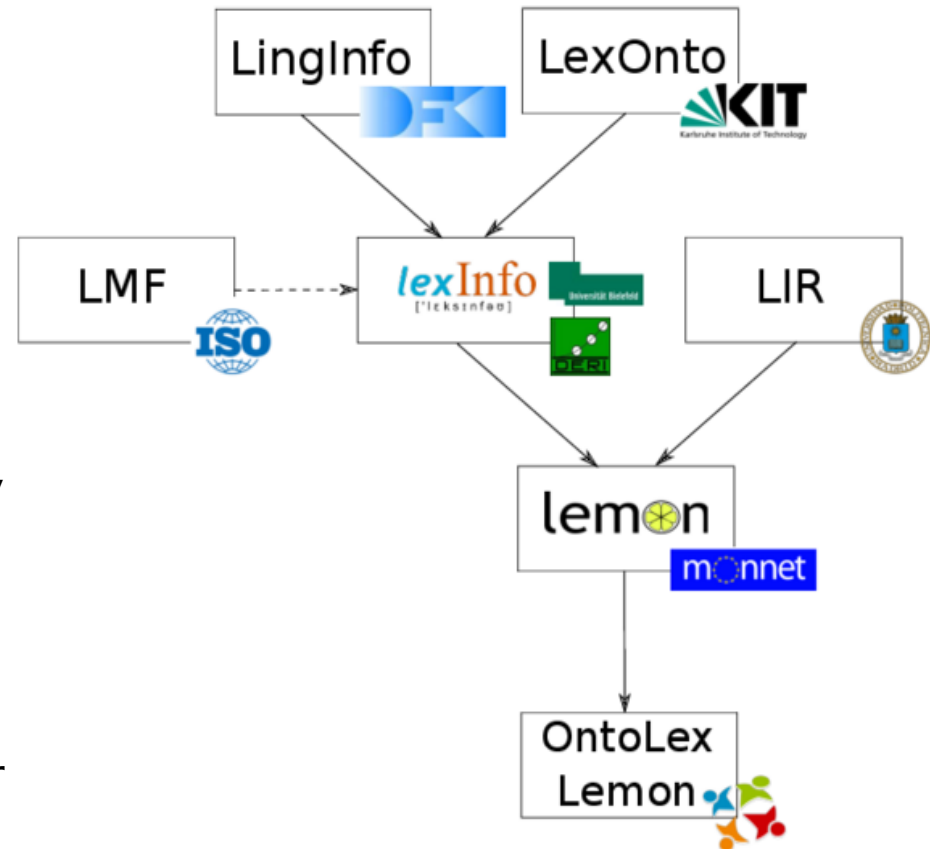
lemon-ontolex

¿Qué es y cómo surge?
Modelo de representación
de **información**
lingüística
multilingüe con
respecto a una ontología



Los orígenes...

- LMF -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



HOY

- Modelo de referencia para transformar recursos lingüísticos al formato de datos enlazados
(Ver <http://linghub.lider-project.eu/>)

From McCrae, J., "lemon: The Lexicon Model for Ontologies"

SD-LLOD-15



lemon-ontolex

¿Qué necesidades intenta cubrir?

- Escasez de información lingüística en ontologías
 1. Información morfosintáctica
 2. Variación terminológica
 3. Traducciones
 4. ...
- Falta de modelos o estándares para representar recursos lingüísticos en RDF

RDF(S), OWL

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

SKOS

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

↑
rdfs:SubPropertyOf



limitación 1

Ejemplo 1: “author” – **escasez de info léxica asociada a entidades de ontologías o vocabularios**

Ejemplo: en DBpedia, “author” es una propiedad con este identificador:

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



limitación 1

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

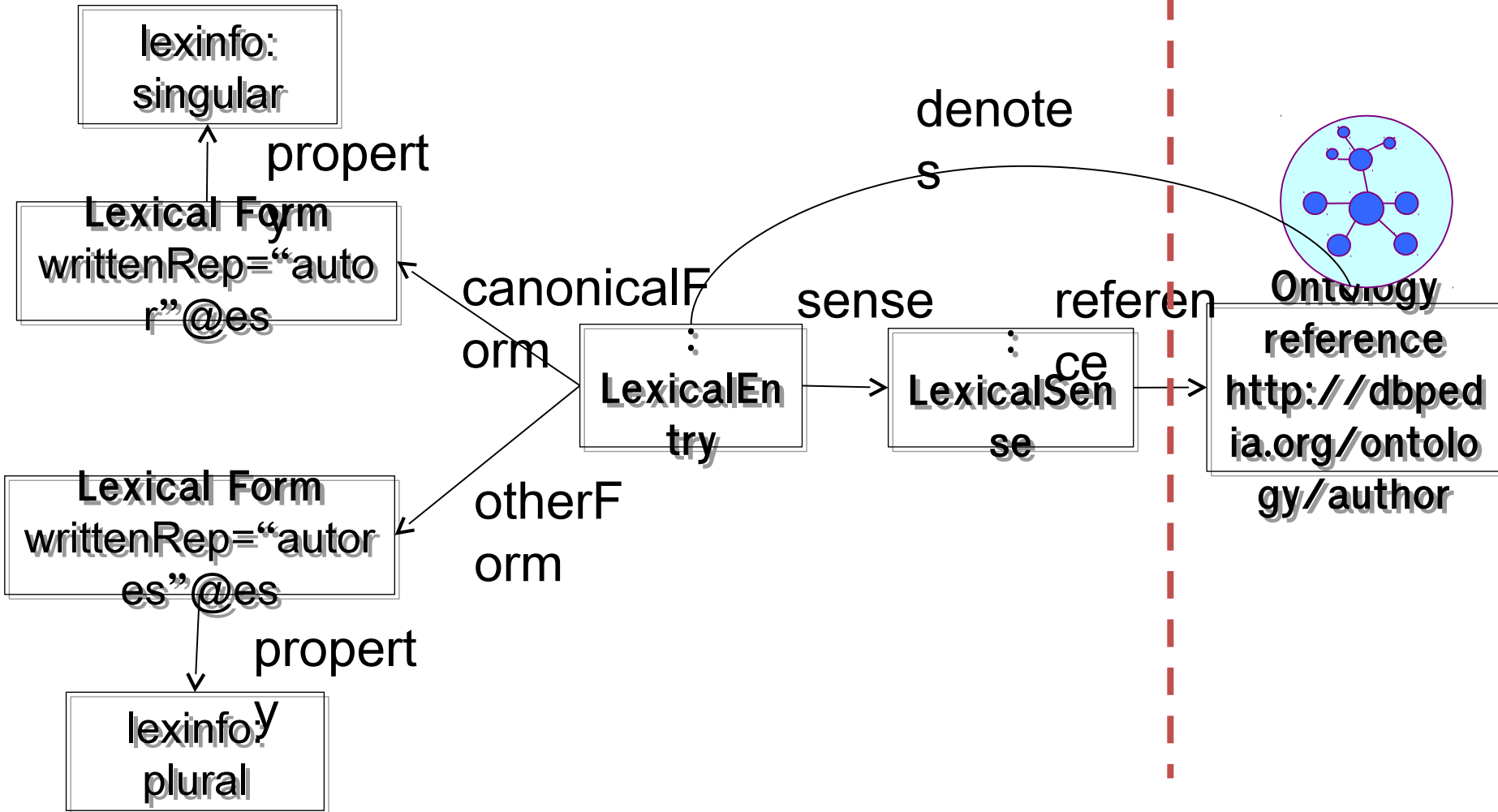
“autora”, “autores”, “authors”, “Verfasserin”

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

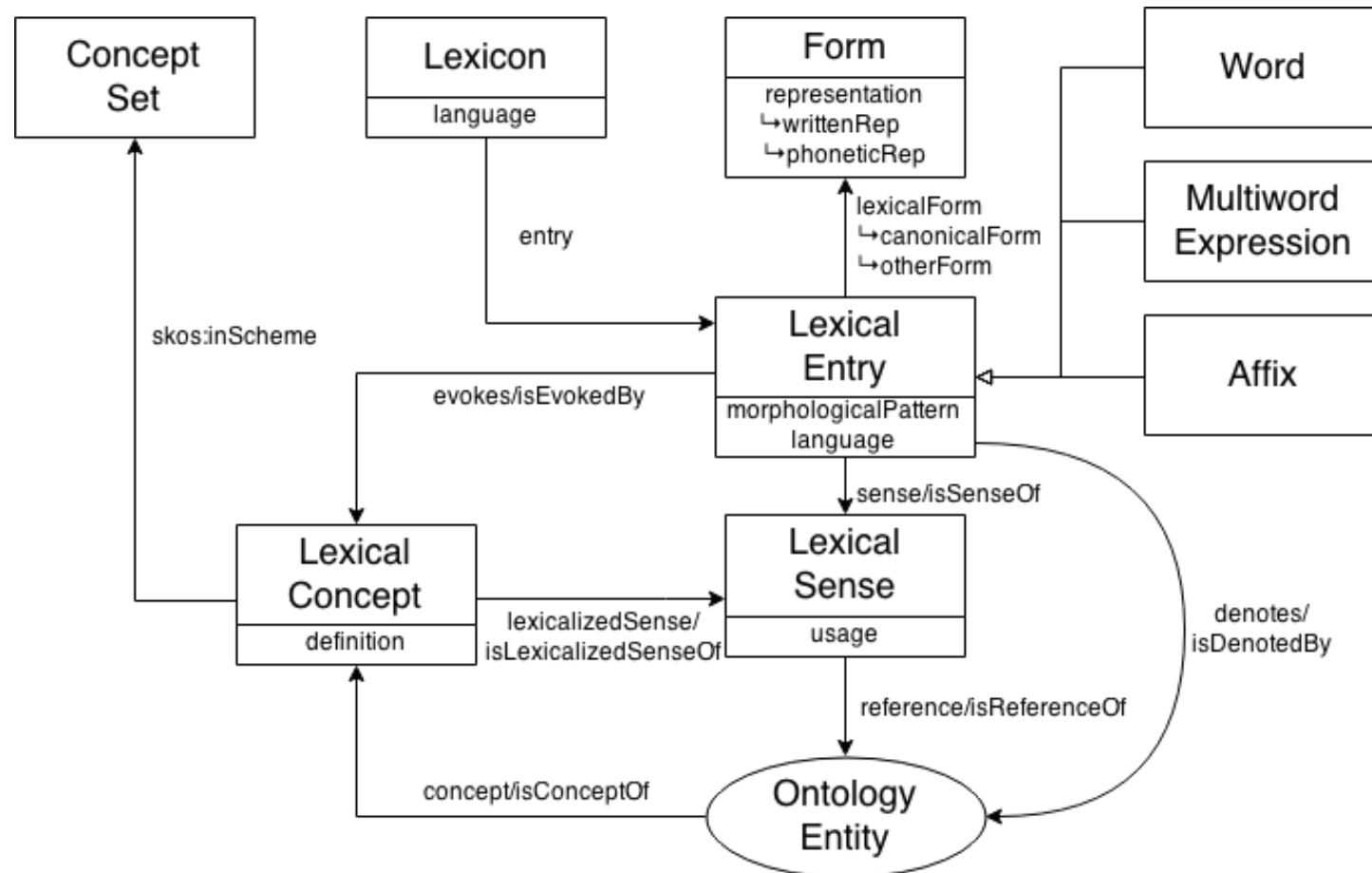
Así que necesitamos que las palabras sean entidades!!



limitación 1



lemon-ontolex Core



limitación 2

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

En DBpedia “escritor” es una clase:

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

Pero en la mayoría de vocs aparece como “autor” y es una propiedad:

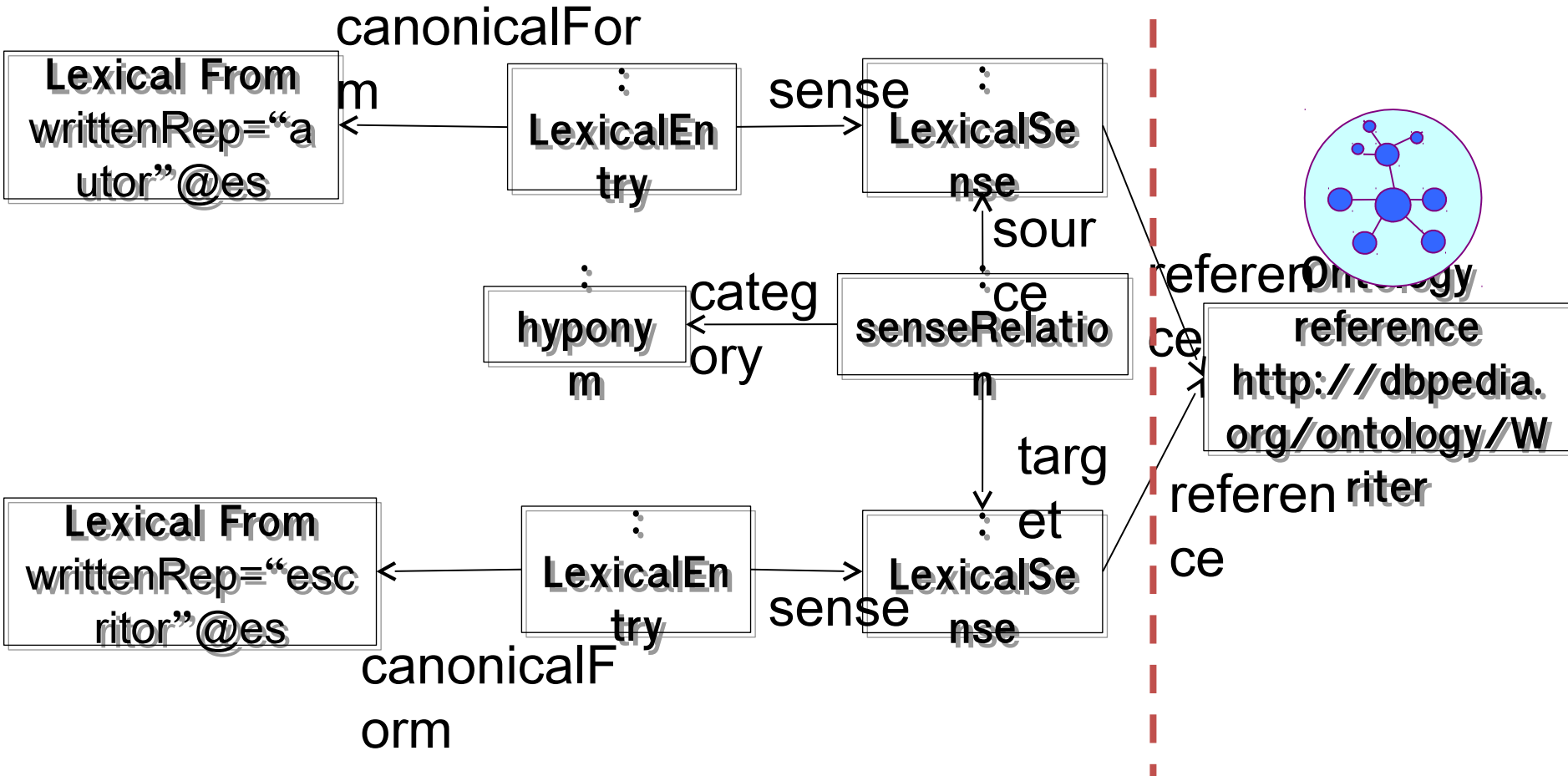
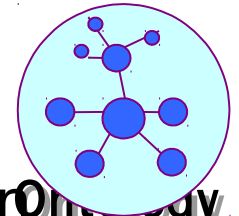
<http://d-nb.info/standards/elementset/gnd#author>

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

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



limitación 2



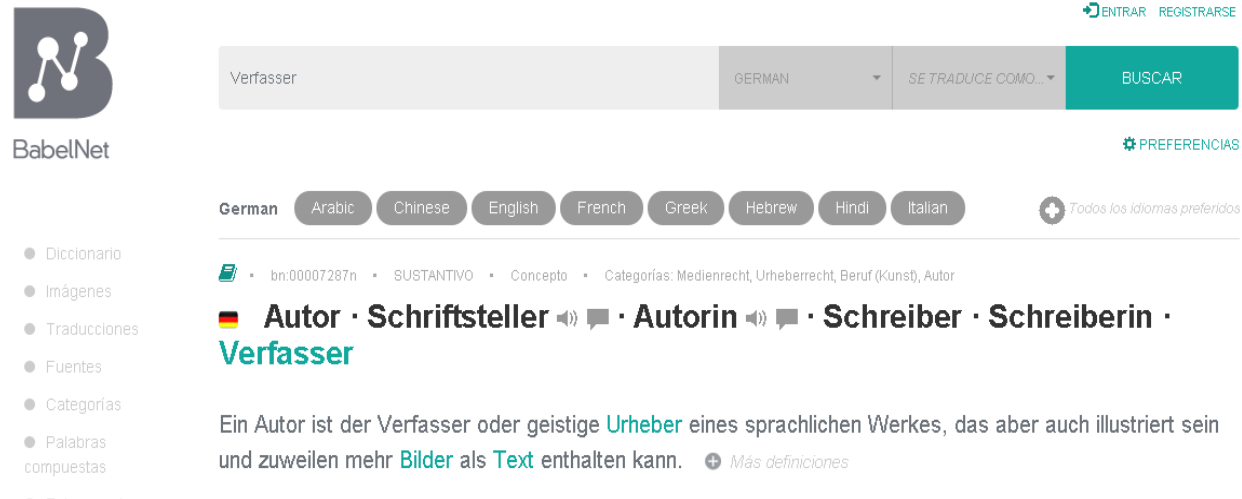
limitación 3

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?

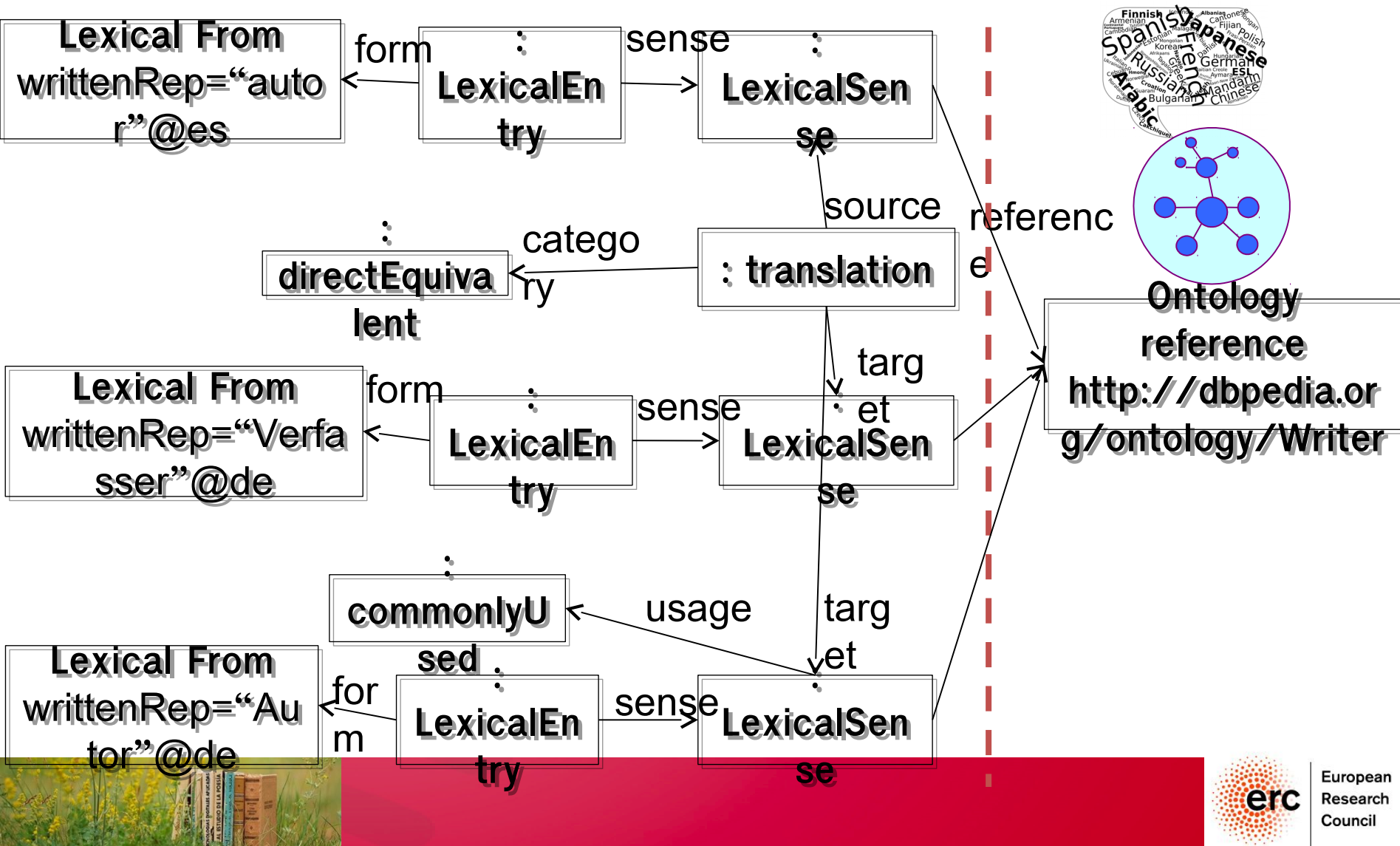


The screenshot shows the BabelNet interface. On the left is the BabelNet logo and a sidebar with navigation links: Diccionario, Imágenes, Traducciones, Fuentes, Categorías, and Palabras compuestas. The main area features a search bar with 'Verfasser' entered, a language dropdown set to 'GERMAN', and a 'BUSCAR' button. Below the search bar are buttons for various languages: German, Arabic, Chinese, English, French, Greek, Hebrew, Hindi, and Italian. The results section shows the word 'Autor' with its German equivalent 'Schriftsteller' and English equivalents 'Autorin', 'Schreiber', and 'Schreiberin'. The word 'Verfasser' is highlighted in green. 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.' A link to 'Más definiciones' is also present.

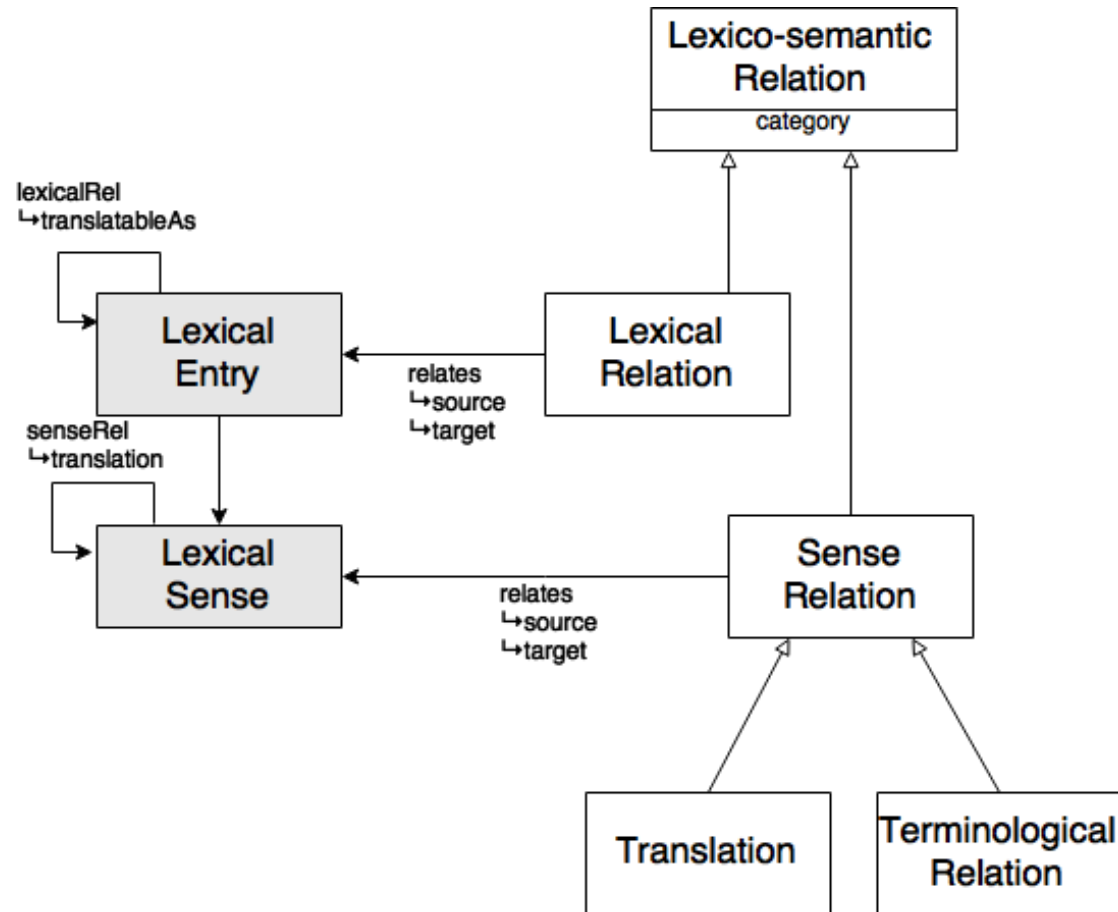
Queremos añadir info sobre frecuencia de uso en las traducciones, etc.



limitación 3



lemon-ontolex vartrans module



Resumen de limitaciones que lemon-ontolex pretende cubrir:

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





DH@Madrid Summer School 2016

La nube de datos lingüísticos (multilingües) enlazados y *ontolex-lemon*

Elena Montiel Ponsoda

emontiel@fi.upm.es

Agradecimientos a Jorge Gracia

Madrid, 29 junio de 2016



European
Research
Council

