



Linguistic Linked Data and Ontolex lemon

Jorge Gracia (University of Zaragoza, Spain)
Fahad Khan (Institute for Computational Linguistics «A. Zampolli»)

Motivation of LLD

LLD - Motivation

Electronic [Language Resources \(LRs\)](#) on the Web

- Dictionaries & Lexica
- Terminologies & Thesauri
- Corpora
- Encyclopedic resources/Knowledge Bases
- Wordnets
- etc.



LLD - Motivation

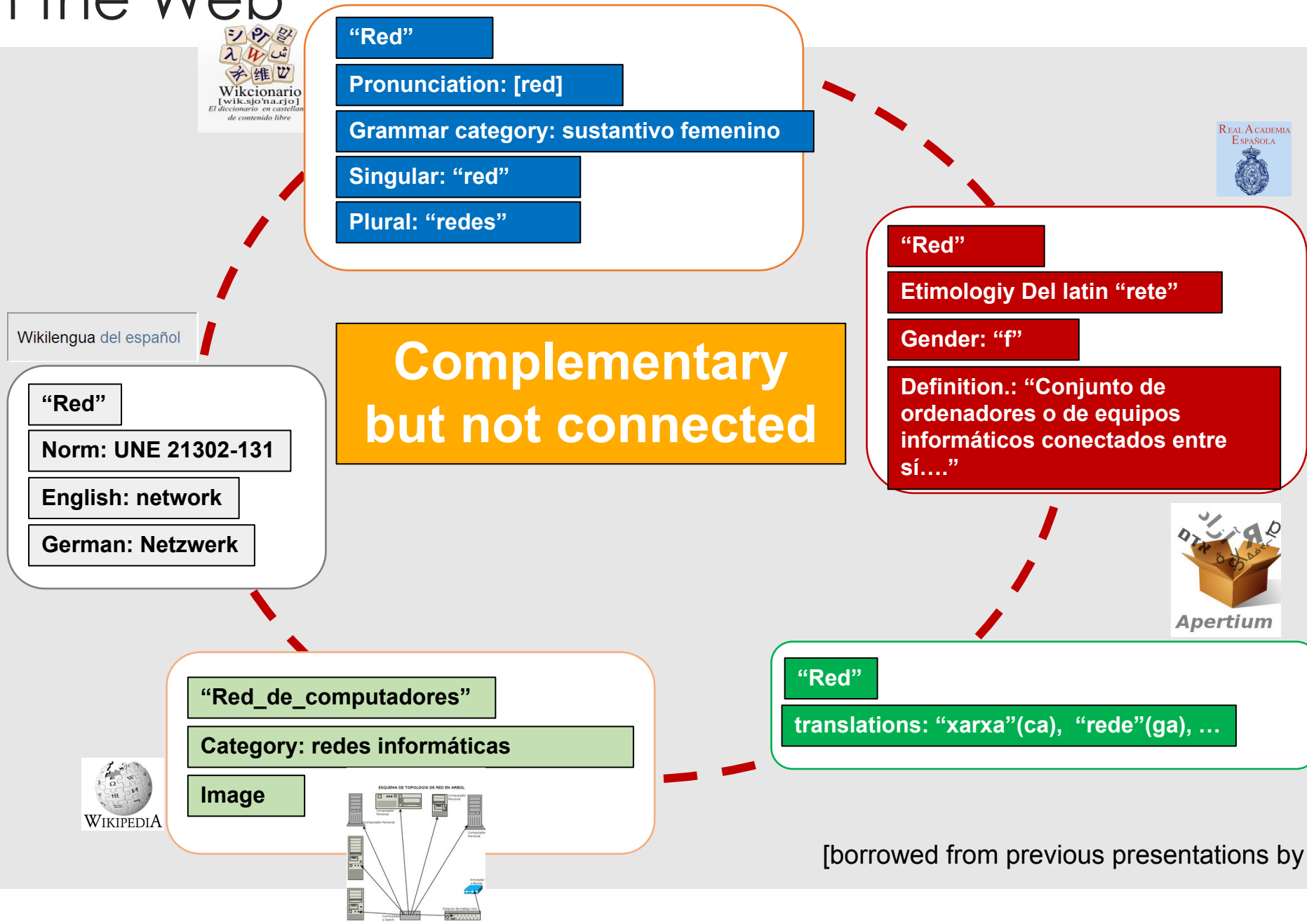
LRs Currently...

- **Disconnected** from other resources (silos of LRs)
- Proprietary and heterogeneous **formats**
- Different **representation schemes**
- Non-standard **access means (APIs)**
- Different **access levels** (from “write me an email” to web services)
- Several LRs repositories with **different metadata** and schemas



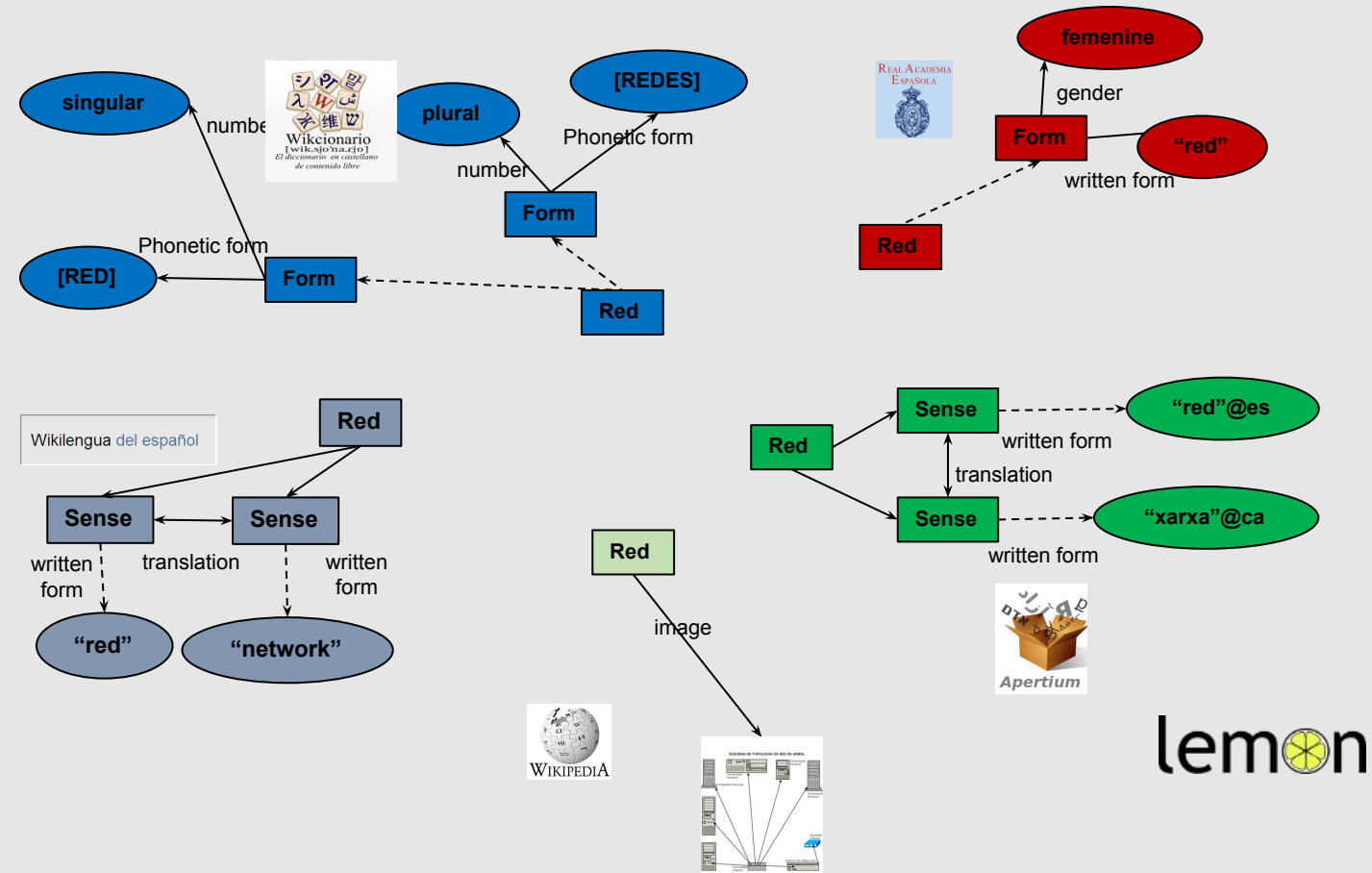
∴ Lack of interoperability across datasets that are potentially complementary & that could be combined together

LRs on the Web



[borrowed from previous presentations by OEG]

LLD - Motivation



[borrowed from previous presentations by OEG]

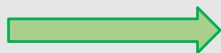
LLD - Motivation

Some BENEFITS of LRs as Linked Data

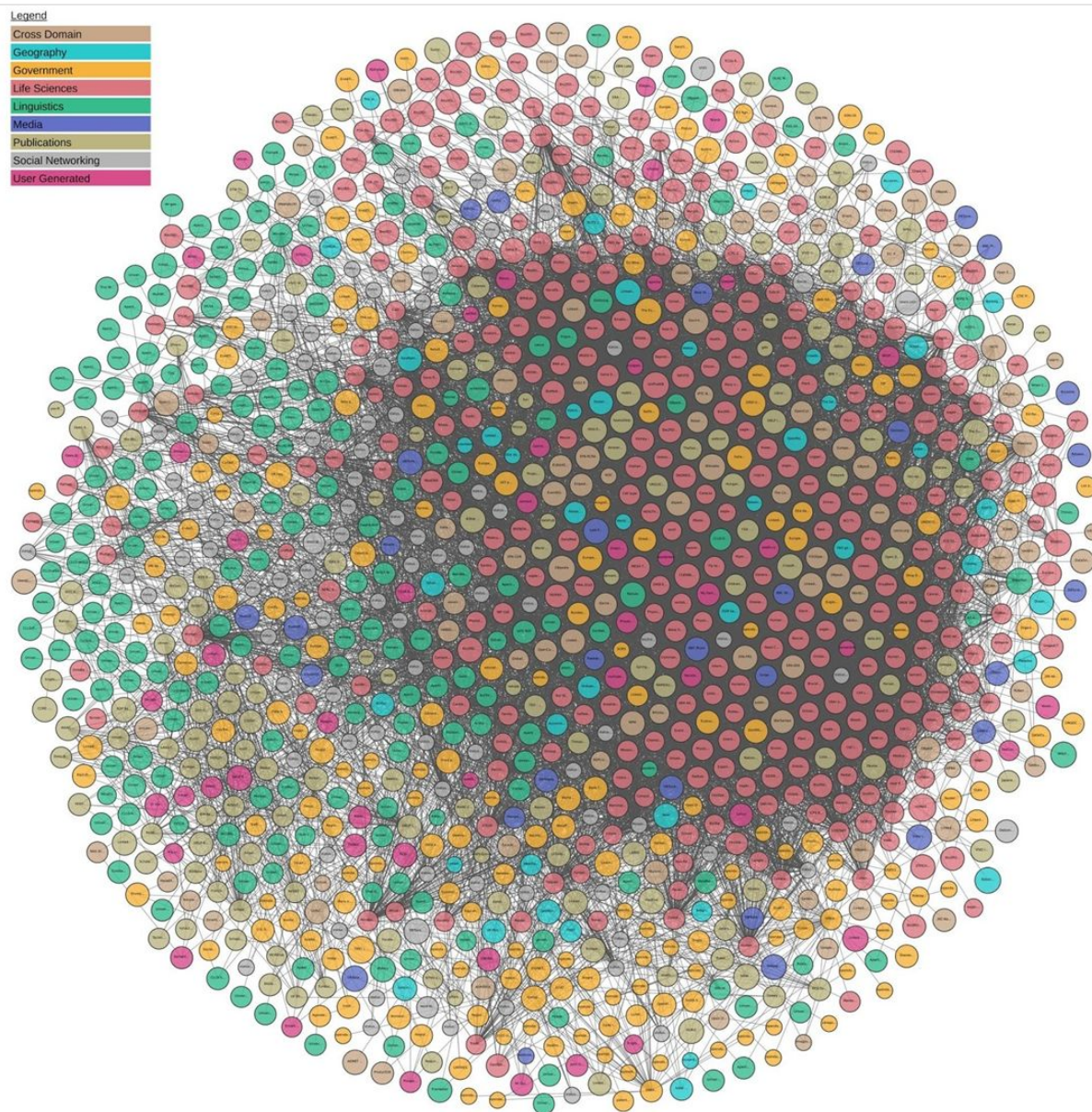
- Aggregation and **integration** of linguistic resources (with each other & kinds of datasets) using a common data model (RDF)
- Resources are explicitly **linked**
- Data **exposed** in a standardized way (SPARQL)
- Improved **discovery** of dataset and services
- Use of common **vocabularies** for representing language content and metadata
- In short: Linked Data helps to make LR's **Findable**, **Accessible**, **Interoperable**, and **Reusable**



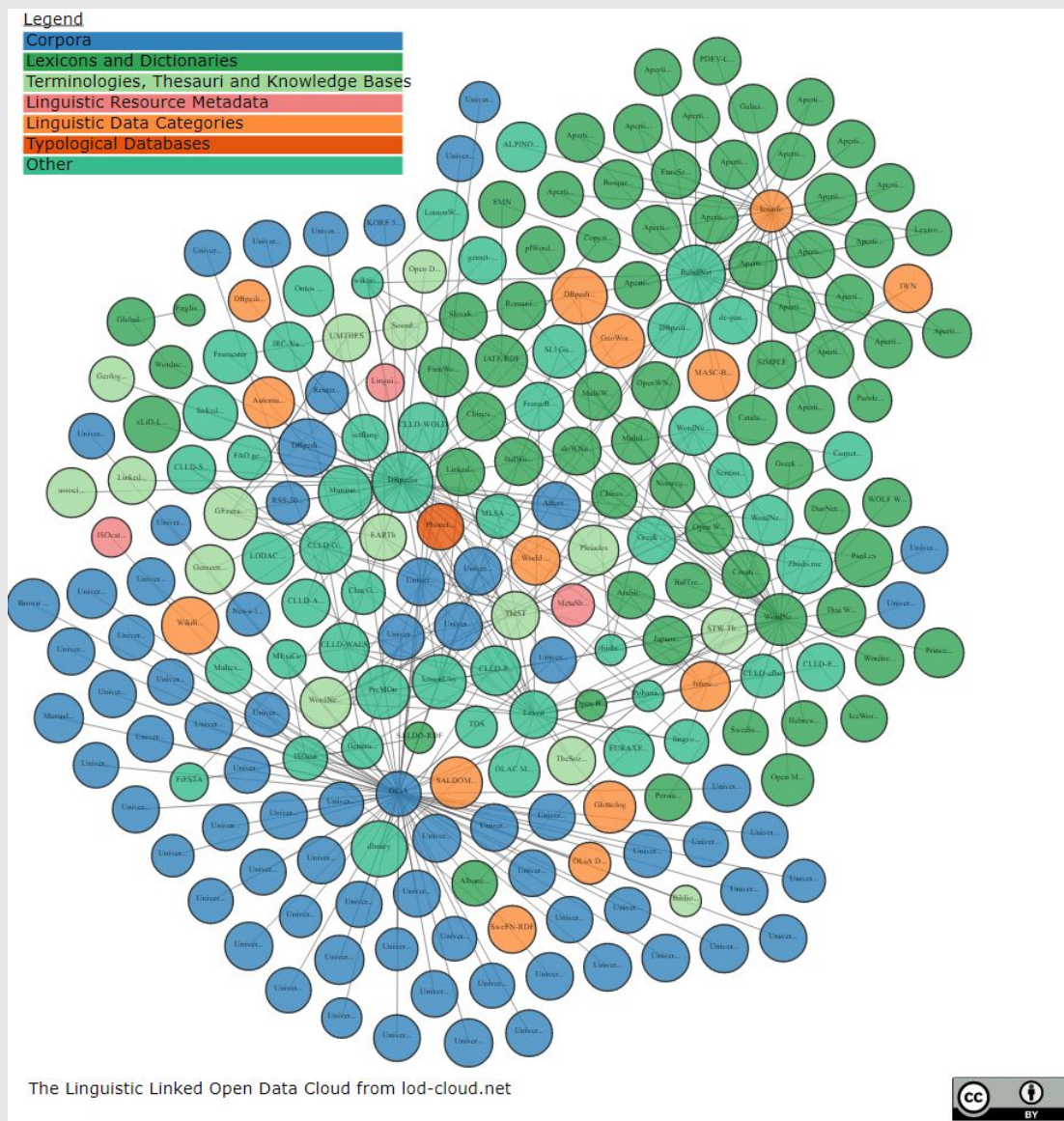
LOD cloud today



“Linguistics” in green



Linguistic Linked Open Data Cloud



<http://linguistic-lod.org/>

Well-known Vocabularies for Linguistic Linked Data

- Lexicons and Dictionaries:
 - [OntoLex-Lemon](#), [lexicog](#)
- Metadata vocabularies:
 - [Basic Metadata: Dublin Core](#) , [FOAF](#), [DCTERMS](#), [Prov-O](#)
 - [Linguistic Metadata: lime](#), [METASHARE](#)
- Terminology and Thesauri:
 - [SKOS \(-XL\)](#)
- Corpora and Annotation:
 - [NIF](#), [Web Annotation](#)
- Data Categories:
 - [lexinfo](#)

The Ontolex lemon model



General Requirements

R1. OWL and RDF

R2. Multilinguality

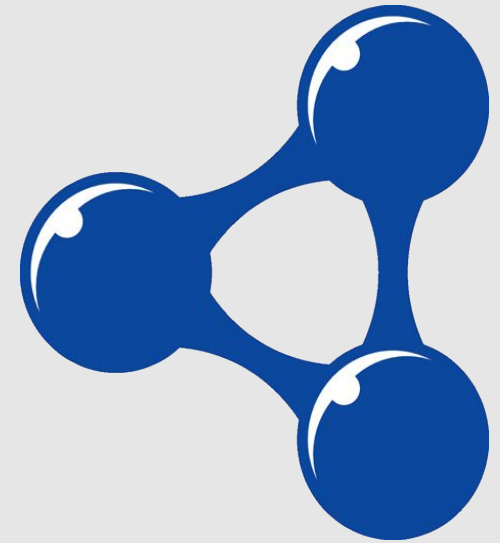
R3. Semantics by Reference

R4. Openness

R5. Reuse relevant standards

RDF and OWL

- RDF models are labelled directed graphs
- Each entry has a URI
- Reuse of lexicon data
- Reasoning



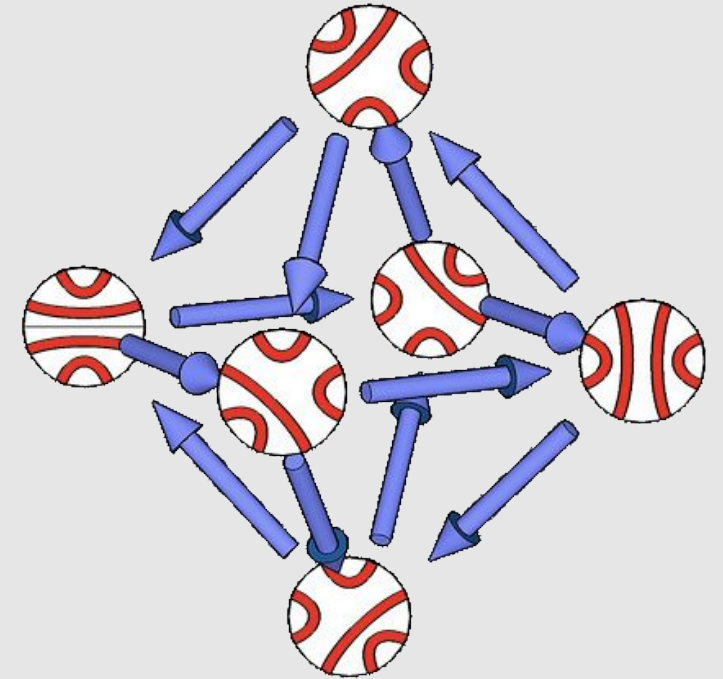
Multilinguality

- Support any language
- Do not make language-specific assumptions
 - Part-of-speech values
 - Gender
 - These can be specified using external vocabularies like lexinfo
- Translation and variation



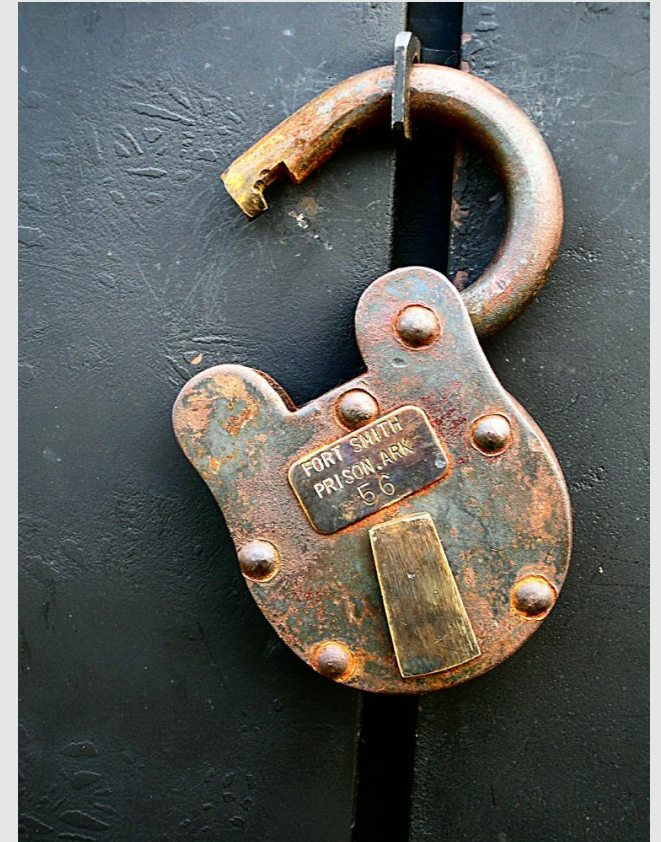
Semantics by Reference

- Meaning of a word given by **reference**
- Reference captures **semantic** information
- Disambiguation is performed relative to the ontology
- No (traditional) word senses



Openness

- Extensible with new models
- No unnecessary choices of linguistic categories
- No payment or restrictions in using the model



Reuse standards

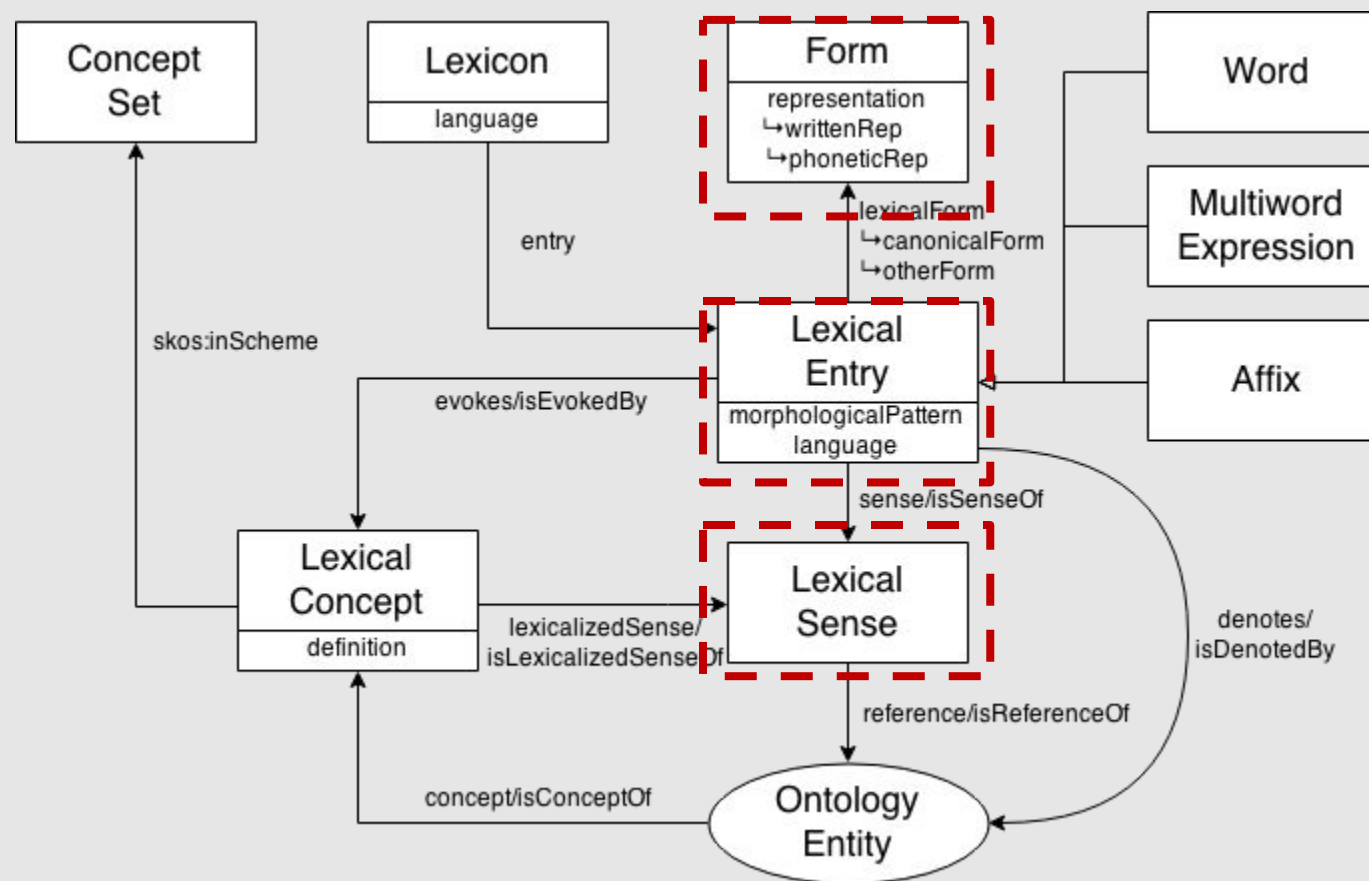
Reuse of standards whenever possible

- OWL
- RDF(S)
- SKOS
- Dublin Core
- LMF
- ...



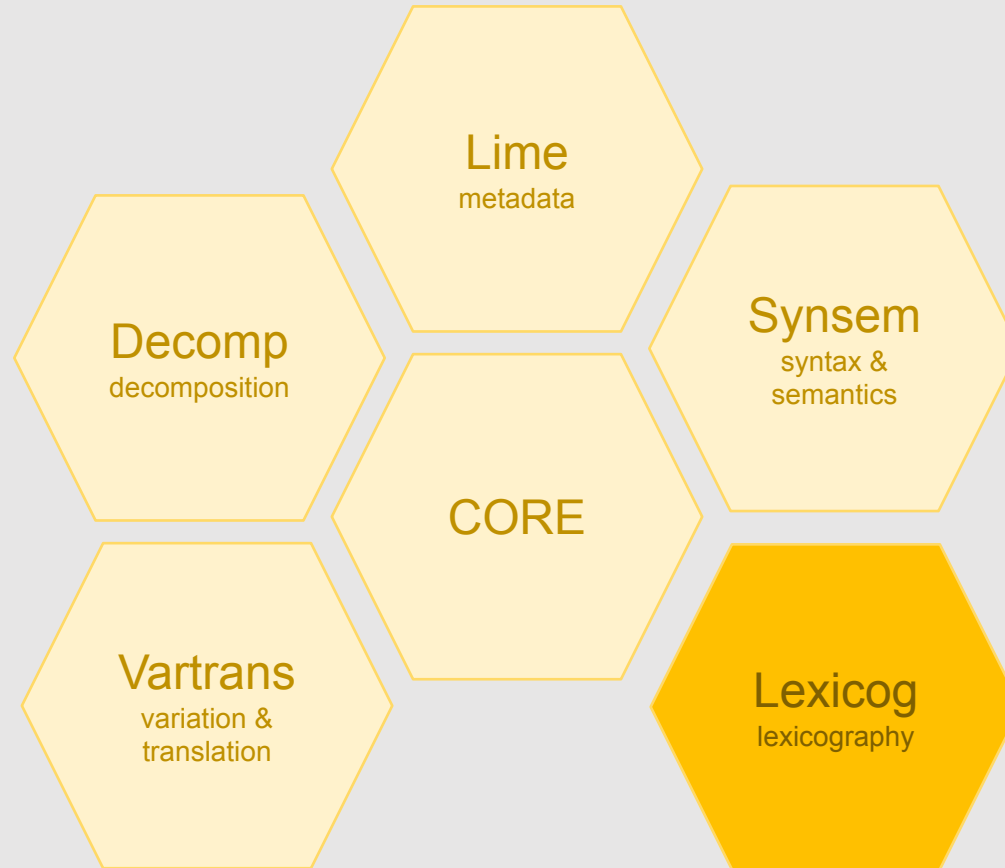
LLD – The Ontolex lemon model

Core of the model

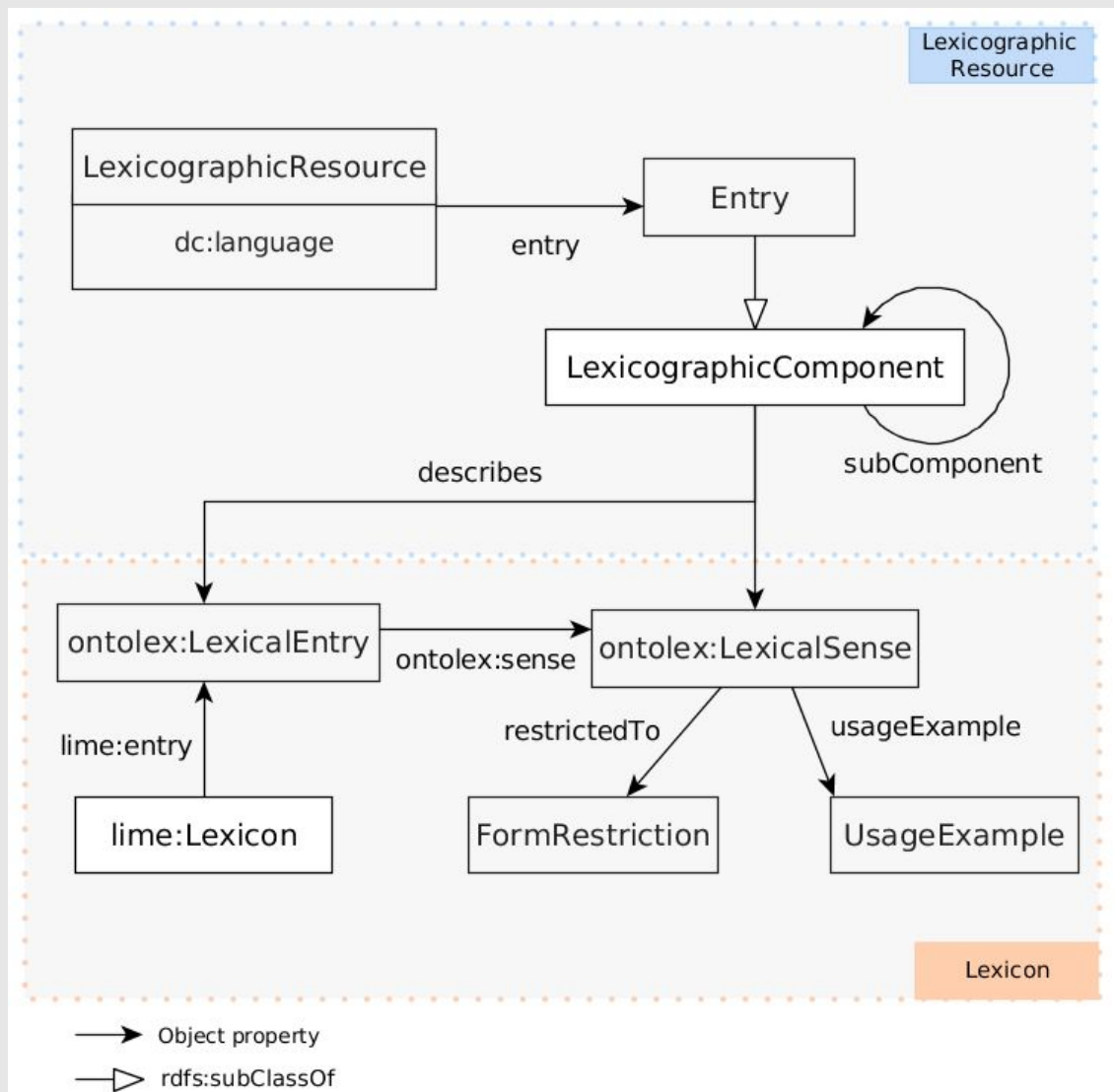


<https://www.w3.org/2016/05/ontolex/>

Lemon Modules



The lexicog module



Examples of LLD applications

Metadata

- [Linghub](#): aggregation of catalogues of language resources (Meta-Share, CLARIN, LREmap, etc.)
- [IULA catalogue](#) of language technologies, specially for Humanities and Social Sciences.
- [Linguistik.de](#): Virtual library which offers scientific information on [linguistic subjects](#)

Lexicography and terminology

- Linked data based [dictionaries](#) (Kernerman Dictionaries, Apertium RDF...)
- Building and indexing [diachronic textual information](#) [Fahad Khan et al.]
- [Language contact](#) studies (LiODi project)
- European [infrastructure](#) for lexicography (ELEXIS project)
- LiLa project: Linking [latin](#)
- TIAD initiative: [translation inference](#) across dictionaries
- Terminoteca: a hub for [terminologies](#) in Spain

Other topics

- [Sentiment analysis](#) (Eurosentiment, MixedEmotions projects)
- Semantic enrichment of [digital content](#) (Frema project)

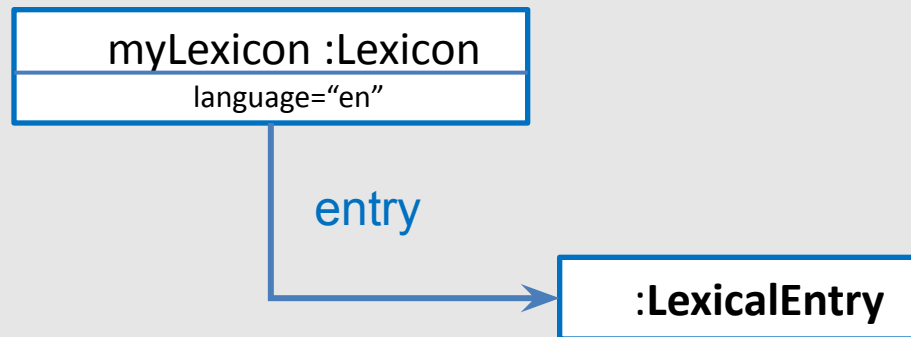
LLD – Ontolex-lemon: an example

Lexicon: The object representing the lexicon as a whole.

myLexicon :Lexicon
language="en"

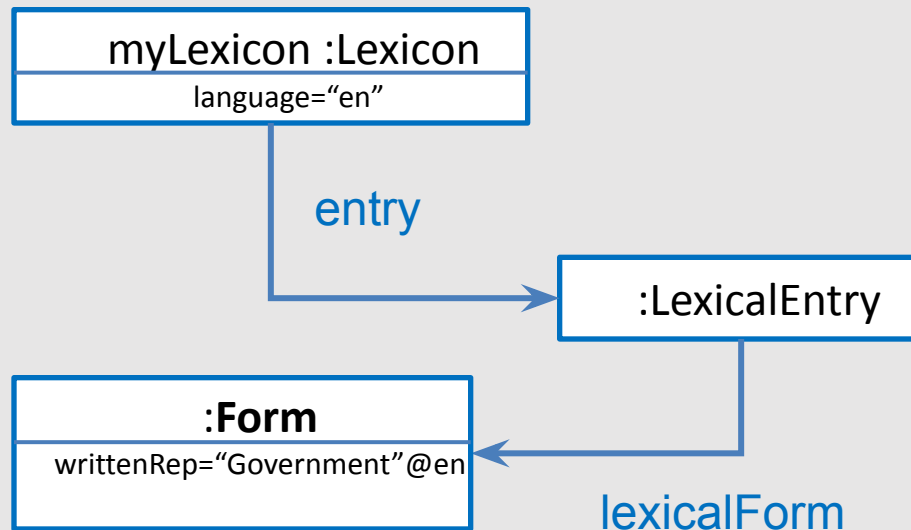
LLD – Ontolex-lemon: an example

Lexical Entry: An entry in a lexicon is a container for one or several **forms** and one or several **meanings** of a lexeme.



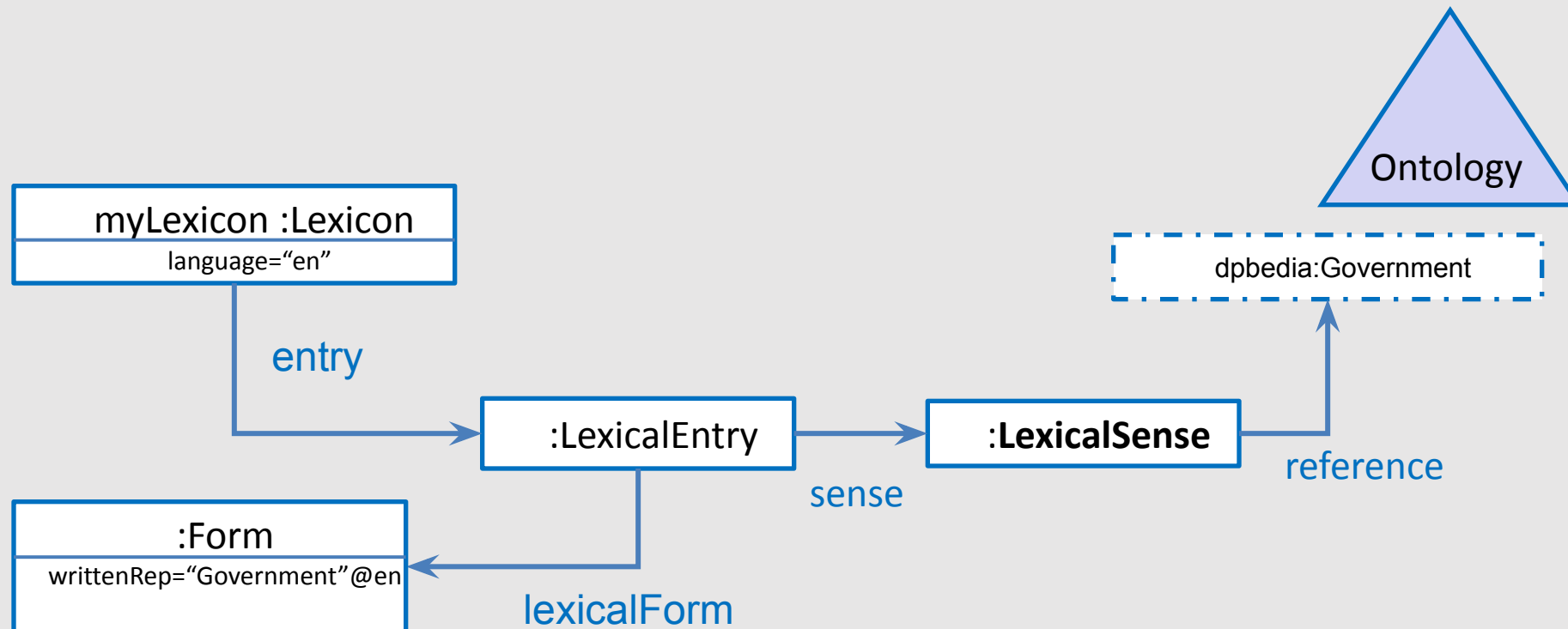
LLD – Ontolex-lemon: an example

Lexical Form: An inflectional form of an entry. A given lexical form may have several **representations** in different orthographies.



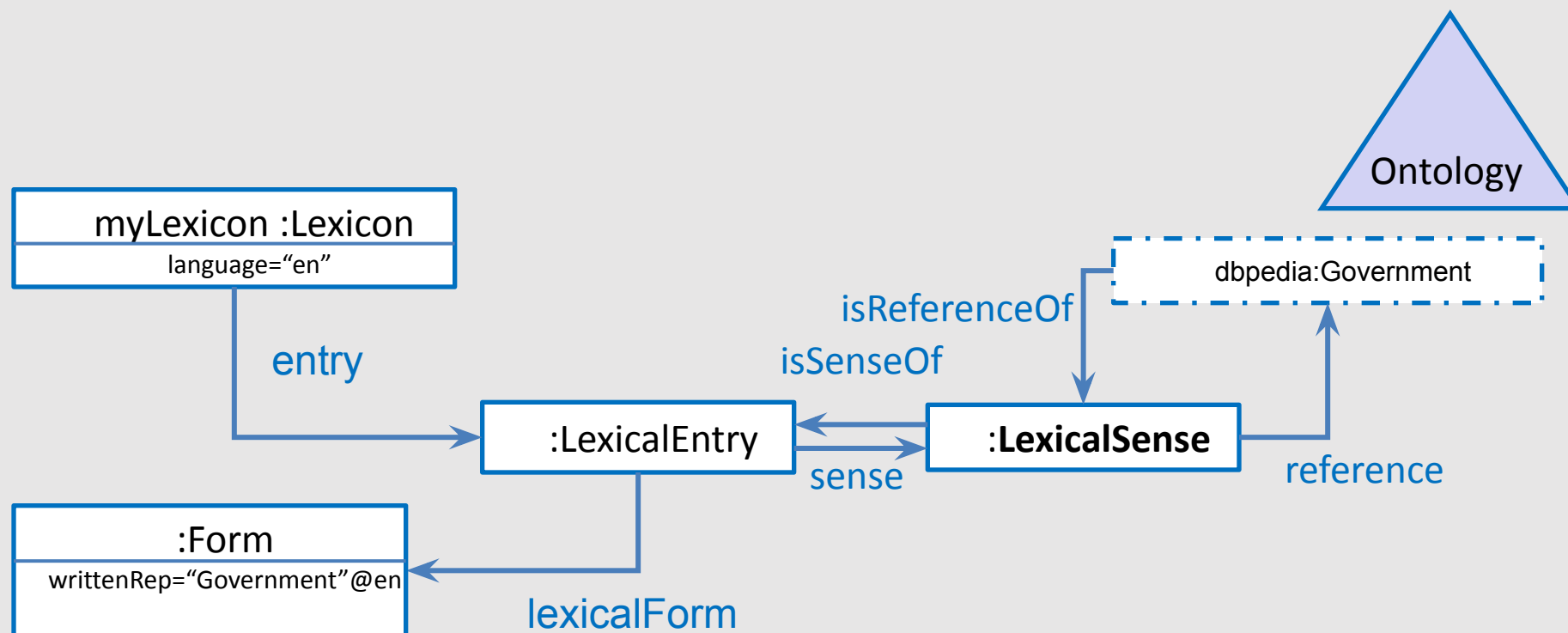
LLD – Ontolex-lemon: an example

Lexical Sense: A sense links the **lexical entry** to the **reference** (ontology term) used to describe its meaning.



LLD – Ontolex-lemon: an example

Relation between **lexical entry** and **reference** ontology term can be in any direction



Entry in Turtle (using blank nodes)

```
@prefix ontollex: <http://www.w3.org/ns/lemon/ontollex#>
@prefix lime: <http://www.w3.org/ns/lemon/lime#> .
@prefix dbpedia: <http://dbpedia.org/resource/>.
```

Namespace



```
:myLexicon a lime:Lexicon ;
  lime:language "en";
  lime:entry :government .
```

Lemma



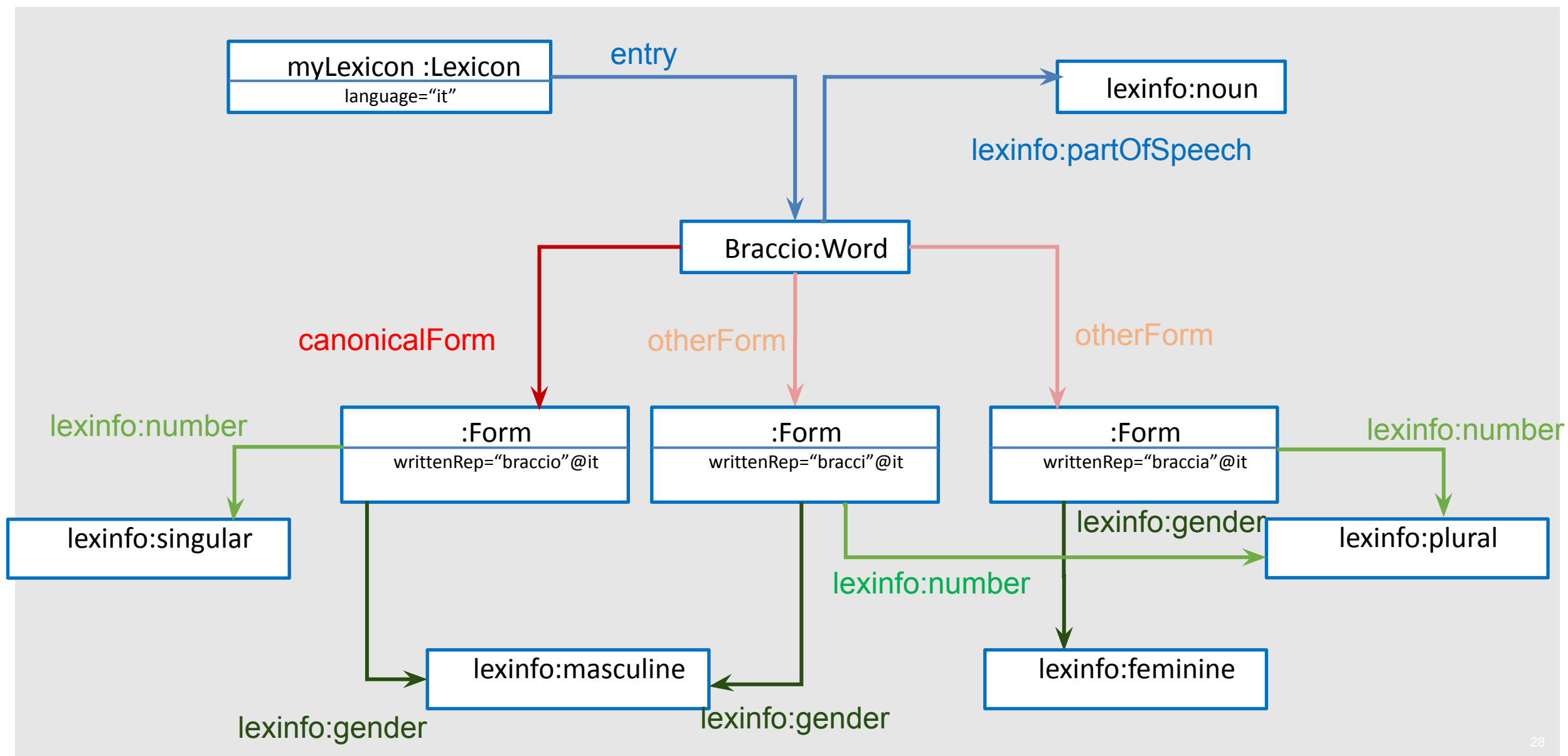
```
:government a ontollex:LexicalEntry ;
  ontollex:canonicalForm [
    ontollex:writtenRep "Government"@en ] ;
```

```
ontollex:sense [
  ontollex:reference dbpedia:Government] .
```

Sense



LLD – Ontolex-lemon: Basic Grammatical Information



Adding Phonetic Information

```
@prefix ontollex: <http://www.w3.org/ns/lemon/ontollex#> .
```

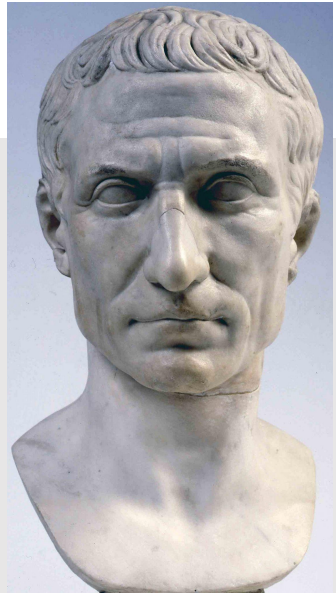
```
lex_tomato a ontollex:LexicalEntry;  
  ontollex:lexicalForm :form_tomato.
```

```
:form_tomato a ontollex:Form;  
  ontollex:writtenRep "tomato"@en;  
  ontollex:phoneticRep "'tʰə. 'meɪrou"@en-US-fonipa;  
  ontollex:phoneticRep "'tə'mɑː.tʊ"@en-GB-fonipa.
```



Basic Morphological Information

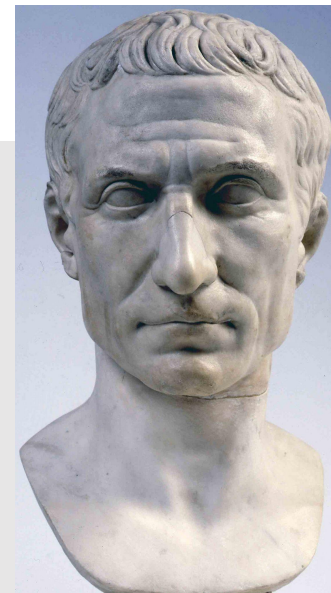
```
:venio ontalex:morphologicalPattern :latin_fourth_conjugation ;  
  ontalex:canonicalForm :venio_form ;  
  ontalex:otherForm :veni_form .  
:venio_form ontalex:writtenRep "veniō"@la;  
:veni_form ontalex:writtenRep "vēnī"@la .
```



Basic Morphological Information

```
:venio ontalex:morphologicalPattern :latin_fourth_conjugation ;  
  ontalex:canonicalForm :venio_form ;  
  ontalex:otherForm :veni_form .  
:venio_form ontalex:writtenRep "veniō"@la;  
:veni_form ontalex:writtenRep "vēnī"@la .
```

```
:video ontalex:morphologicalPattern :latin_second_conjugation ;  
  ontalex:canonicalForm :video_form ;  
  ontalex:otherForm :vidi_form .  
:video_form ontalex:writtenRep "videō"@la ;  
:vidi_form ontalex:writtenRep "vīdī"@la.
```

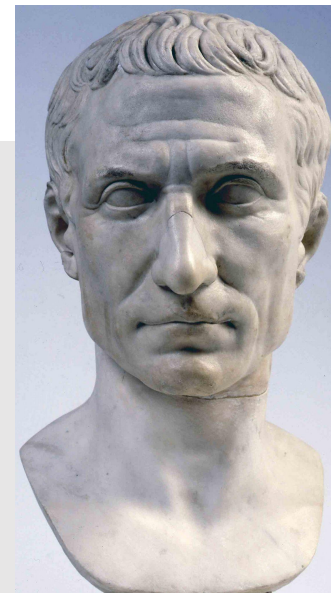


Basic Morphological Information

```
:venio ontolex:morphologicalPattern :latin_fourth_conjugation ;  
  ontolex:canonicalForm :venio_form ;  
  ontolex:otherForm :veni_form .  
:venio_form ontolex:writtenRep "veniō"@la;  
:veni_form ontolex:writtenRep "vēmī"@la .
```

```
:video ontolex:morphologicalPattern :latin_second_conjugation ;  
  ontolex:canonicalForm :video_form ;  
  ontolex:otherForm :vidi_form .  
:video_form ontolex:writtenRep "videō"@la ;  
:vidi_form ontolex:writtenRep "vīdī"@la.
```

```
:vinco ontolex:morphologicalPattern :latin_third_conjugation ;  
  ontolex:canonicalForm :vinco_form ;  
  ontolex:otherForm :vinco_form.  
:vinco_form ontolex:writtenRep "vincō"@la  
:vici_form ontolex:writtenRep "vīcī"@la .
```



Basic Semantic Information

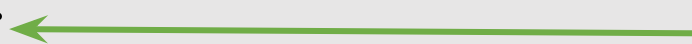
@prefix ontollex: <<http://www.w3.org/ns/lemon/ontollex#>> .

@prefix dbpedia: <<http://dbpedia.org/resource/>> .

@prefix dbo: <<http://dbpedia.org/ontology/>> .

```
:autumn a ontollex:Word ;  
  ontollex:sense [  
    ontollex:reference dbpedia:Autumn ;  
    ontollex:usage [ rdf:value "British English" ] ] ;  
  ontollex:denotes dbpedia:Autumn .
```

sense \circ reference
=
denotes

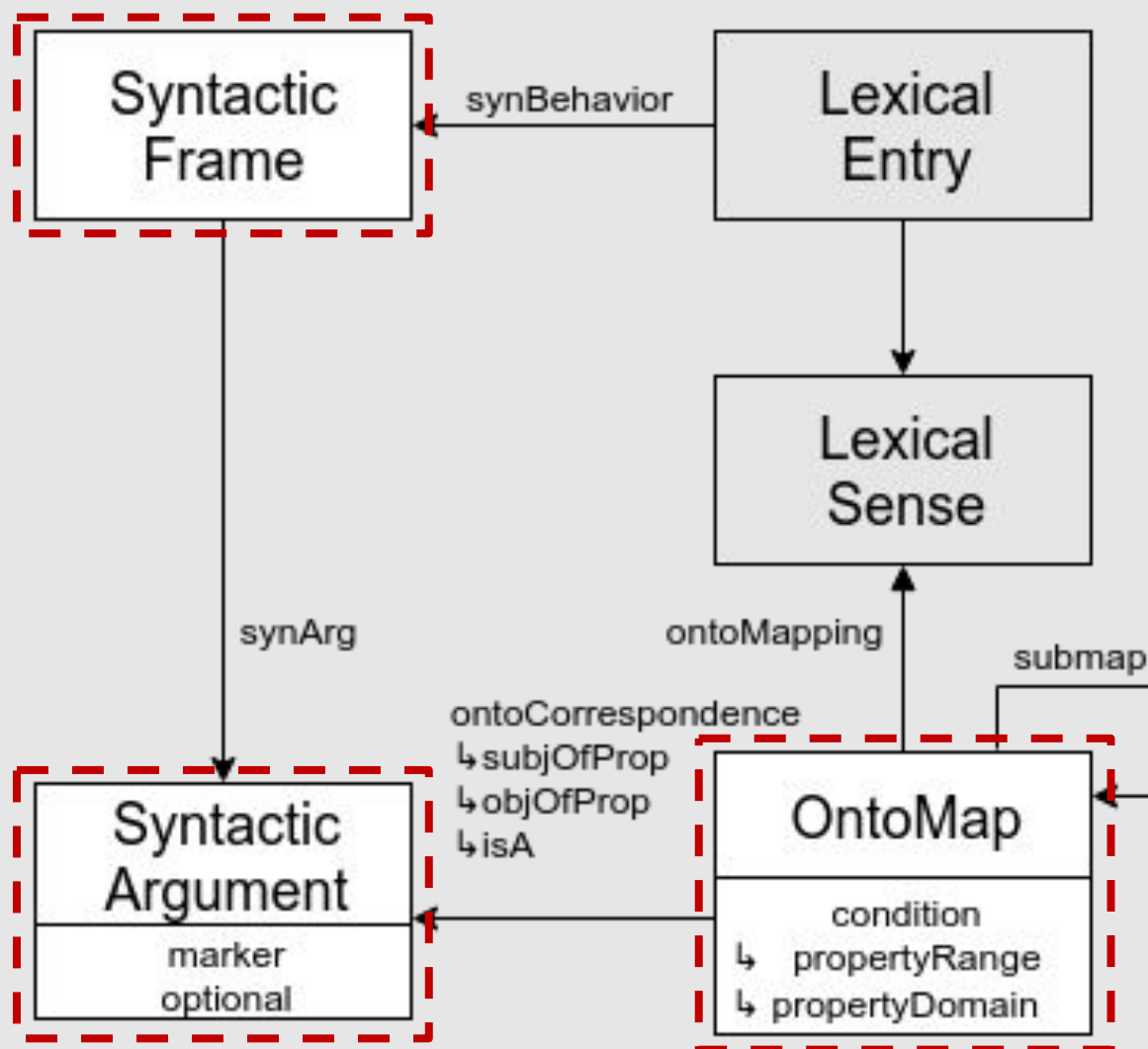


```
:fall a ontollex:Word ;  
  ontollex:sense [  
    ontollex:reference dbpedia:Autumn ;  
    ontollex:usage [ rdf:value "American English" ] ] ;  
  ontollex:denotes dbpedia:Autumn .
```

Restriction on
Lexical Sense



Syntax and Semantics



Syntactic Frames

Synsem
Module

```
@prefix ontollex: <http://www.w3.org/ns/lemon/ontollex#> .  
@prefix synsem: <http://www.w3.org/ns/lemon/synsem#> .  
@prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#> .
```

```
:know a ontollex:Word ;  
      synsem:synBehavior :know_transitive .
```


```
:know_transitive a synsem:SyntacticFrame, lexinfo:TransitiveFrame  
;  
lexinfo:subject :know_subject ;  
lexinfo:directObject :know_directObject .
```

Frame

Syntactic and Semantic Frames

```
@prefix ontollex: <http://www.w3.org/ns/lemon/ontollex#> .  
@prefix synsem: <http://www.w3.org/ns/lemon/synsem#> .  
@prefix lexinfo: <http://www.lexinfo.org/2.0/lexinfo#> .  
@prefix foaf: <http://xmlns.com/foaf/> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
```


Lexical sense is
an ontology
mapping



```
:know a ontollex:Word ;  
  ontollex:sense :know_sense ;  
  synsem:synBehavior :know_transitive .
```

```
:know_sense a ontollex:LexicalSense , synsem:OntoMap ;  
  synsem:ontoMap :know_sense ;  
  ontollex:reference foaf:knows ;  
  synsem:subjOfProp :know_subject ;  
  synsem:objOfProp :know_directObject .
```

Identifiers
from syntactic
frame

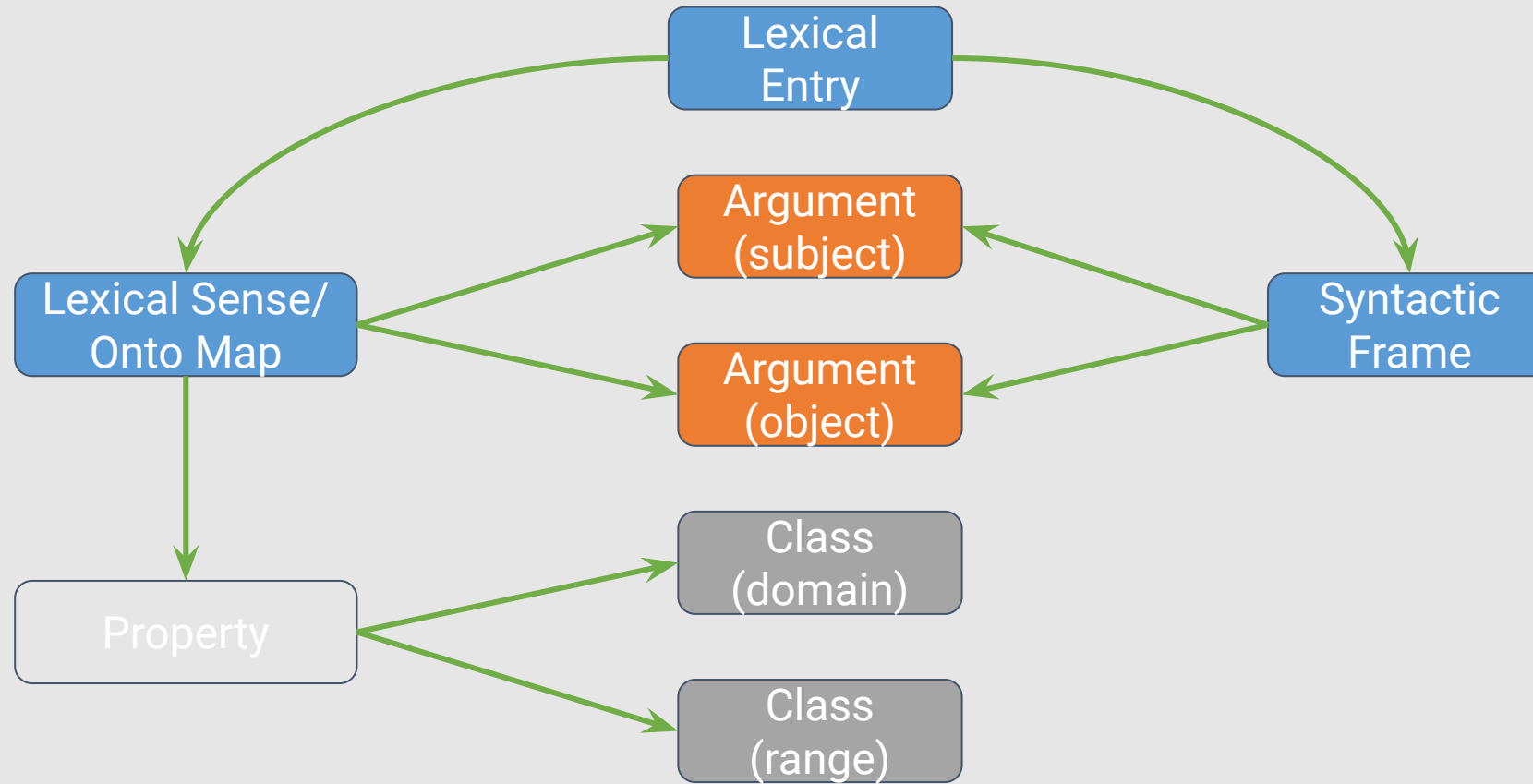


```
foaf:knows a rdf:Property ;  
  rdfs:domain foaf:Person ;  
  rdfs:range foaf:Person .
```

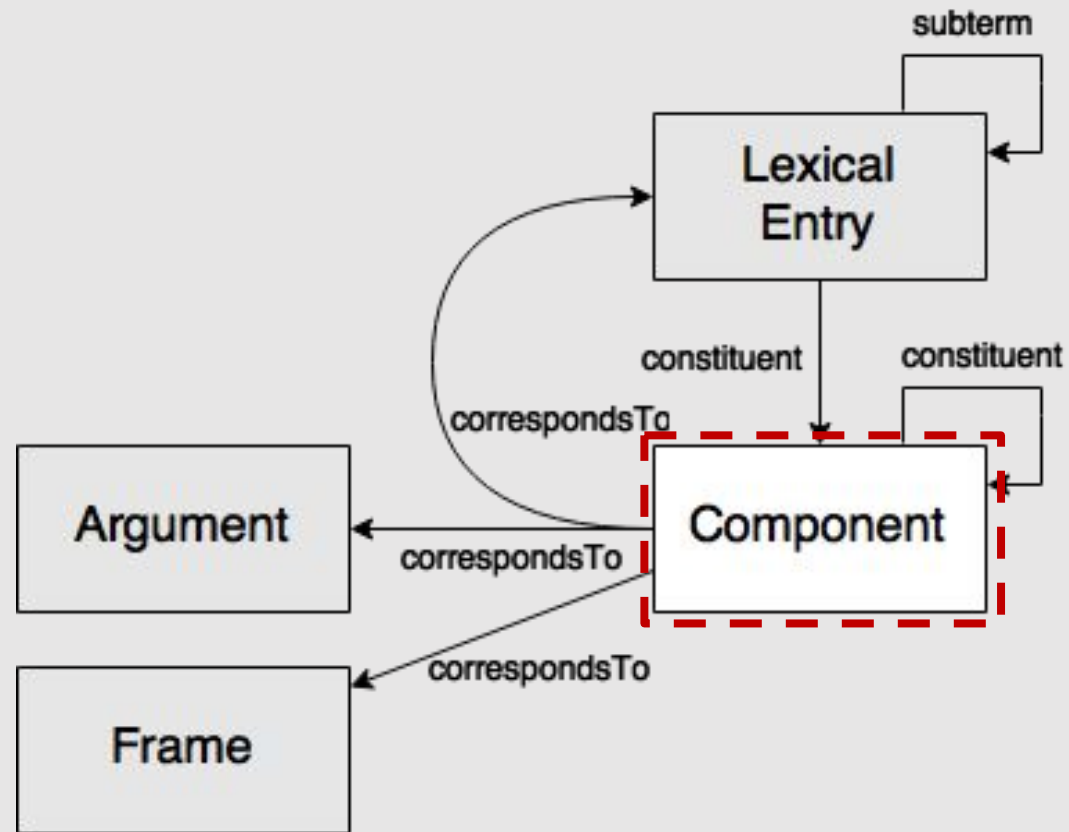
Ontological
definition of
semantic frame



Syntactic-Semantic Mapping



Decomposition



Decomposition

constituent \circ correspondsTo
=
subterm

@prefix ontollex: <<http://www.w3.org/ns/lemon/ontollex#>>

@prefix decomp: <<http://www.w3.org/ns/lemon/decomp#>>

:summer_school a ontollex:MultiWordExpression ;
decomp:subterm :summer, :school .

:école_d'été a ontollex:MultiWordExpression ;
decomp:constituent :école_d'été_école_comp ,
:école_d'été_de_comp ,
:école_d'été_été_comp ;

rdf:_1 :école_d'été_école_comp ;

rdf:_2 :école_d'été_de_comp ;

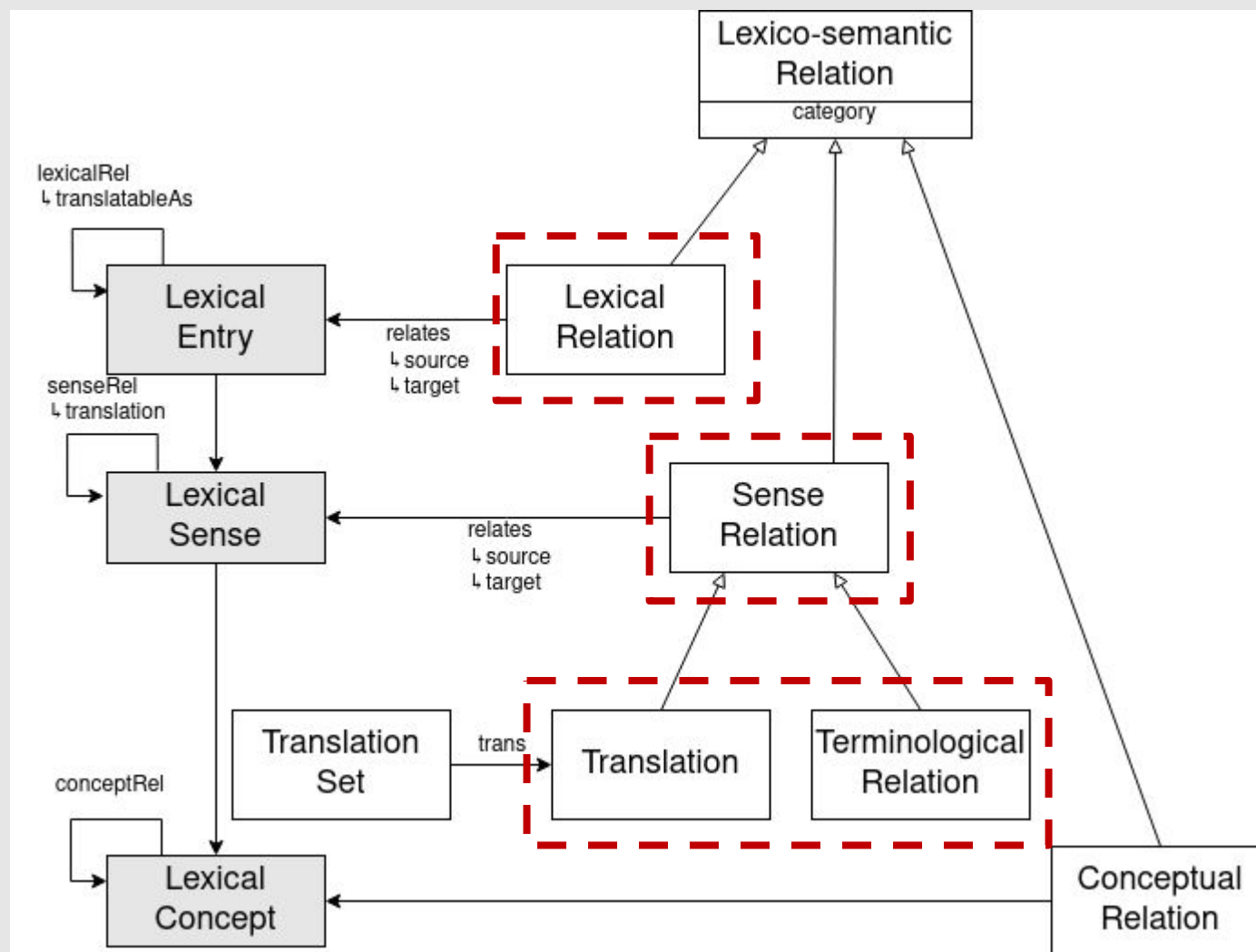
rdf:_3 :école_d'été_été_comp ;

Order

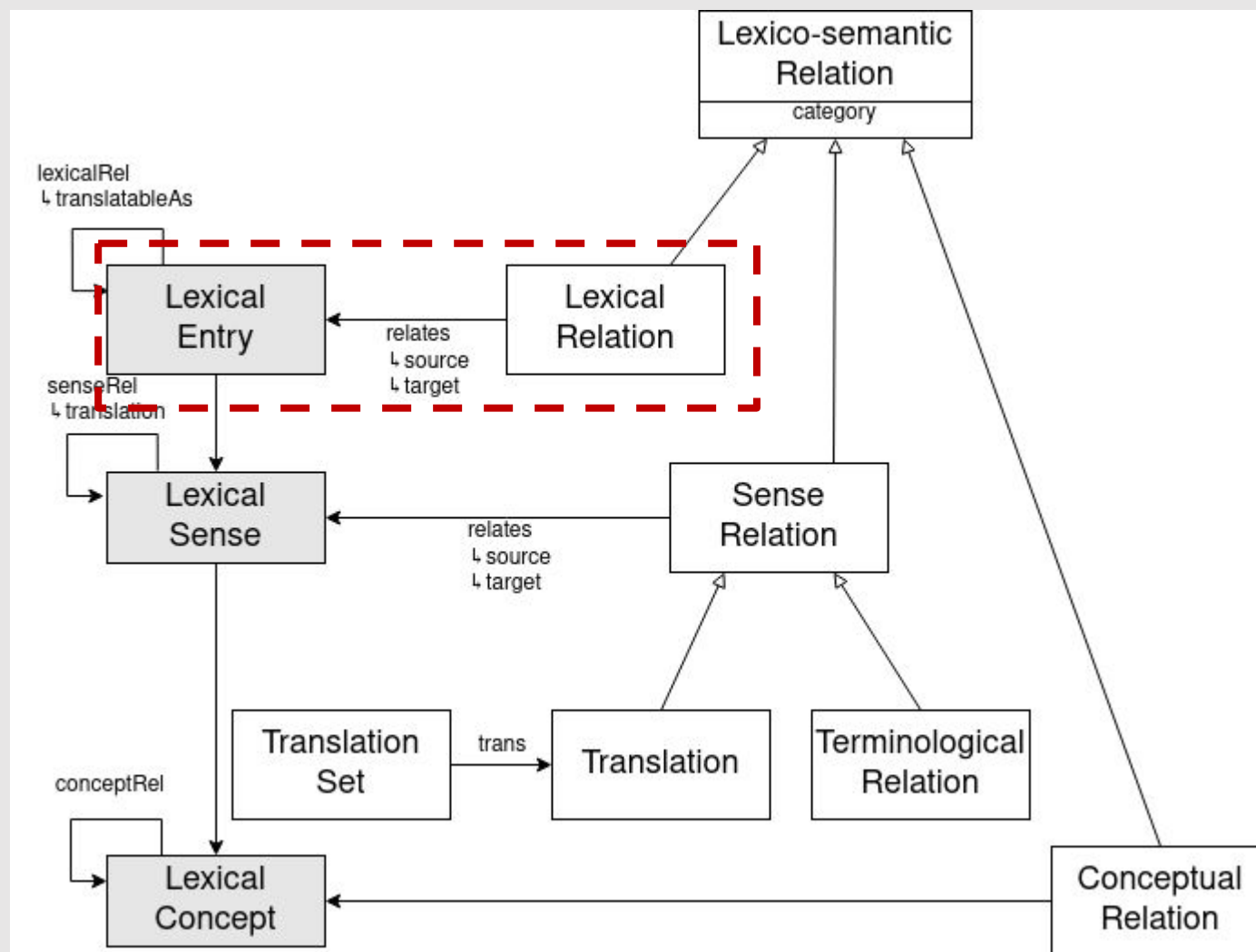
:école_d'été_de_comp a decomp:Component ;
decomp:correspondsTo :de ;
lexinfo:lexTermType lexinfo:contraction .

Component
Properties

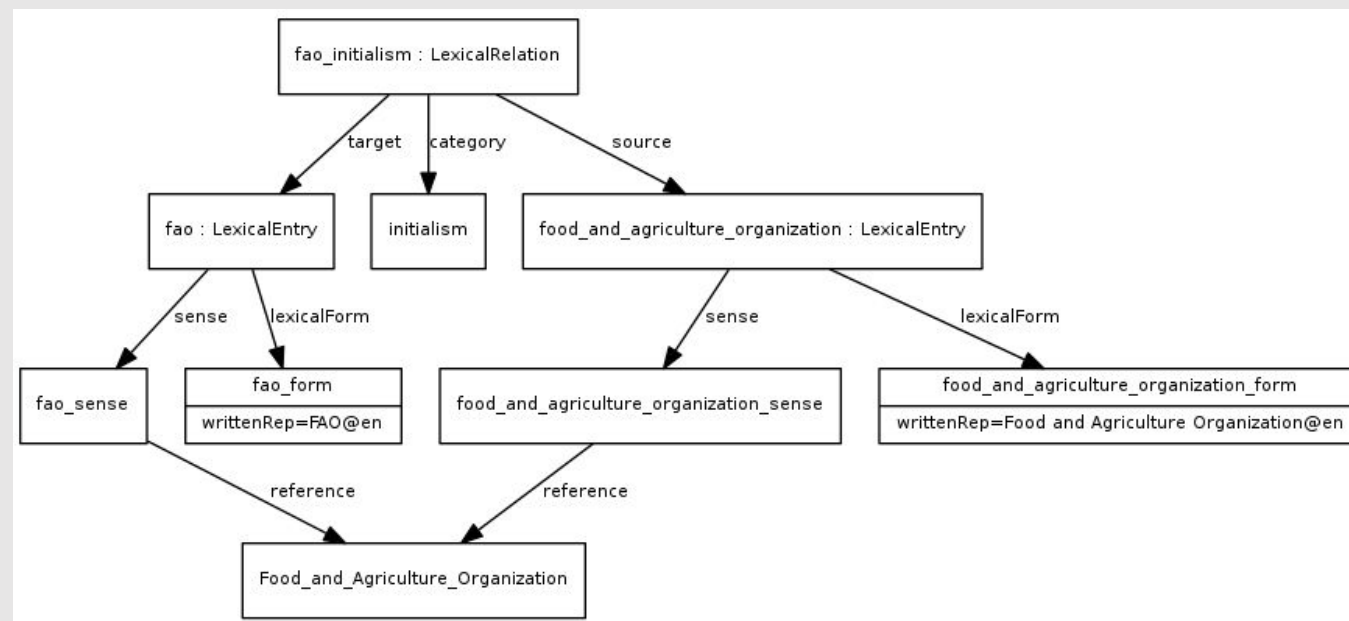
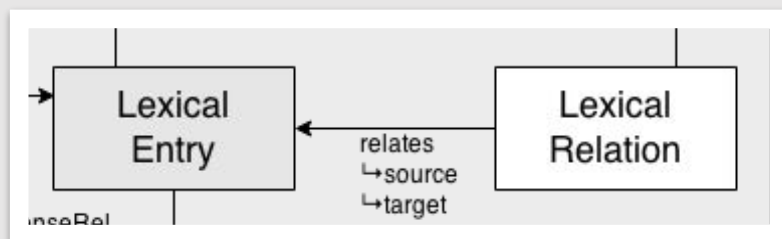
Vartrans module



Vartrans module



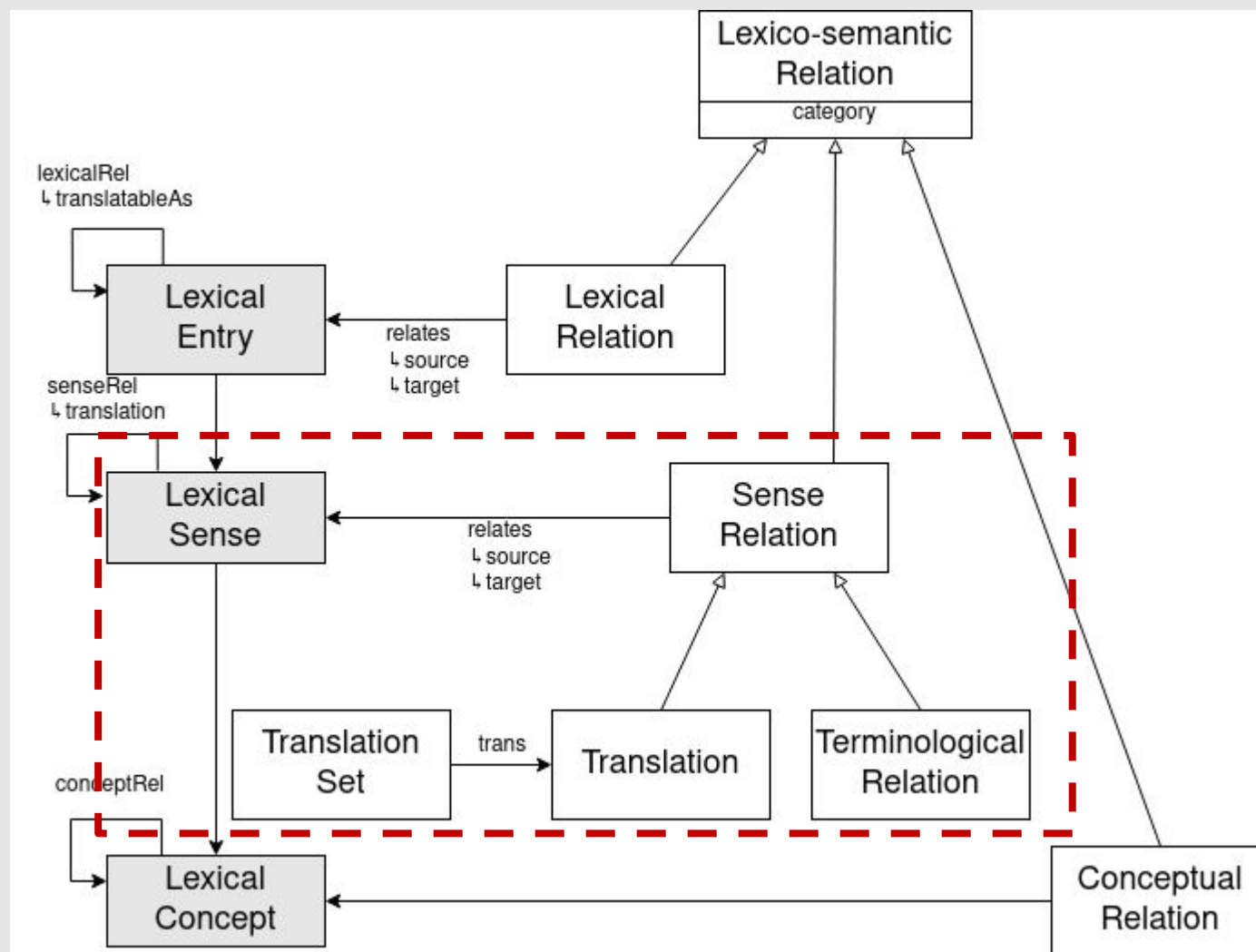
Lexical relations



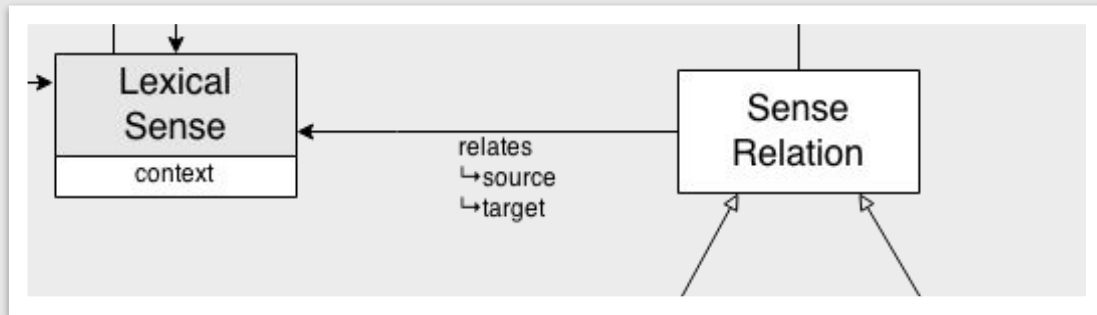
Examples of **lexical relations** are the following:

- **Derivational** relation (e.g., adjective → adverb variation: quick vs. quickly)
- **Morphosyntactic** relation (e.g. ecological tourism vs. eco-tourism)
- **Abbreviation** relation (including acronyms, e.g., peer to peer and p2p; WYSWYG, FAO, UNO)

Vartrans module

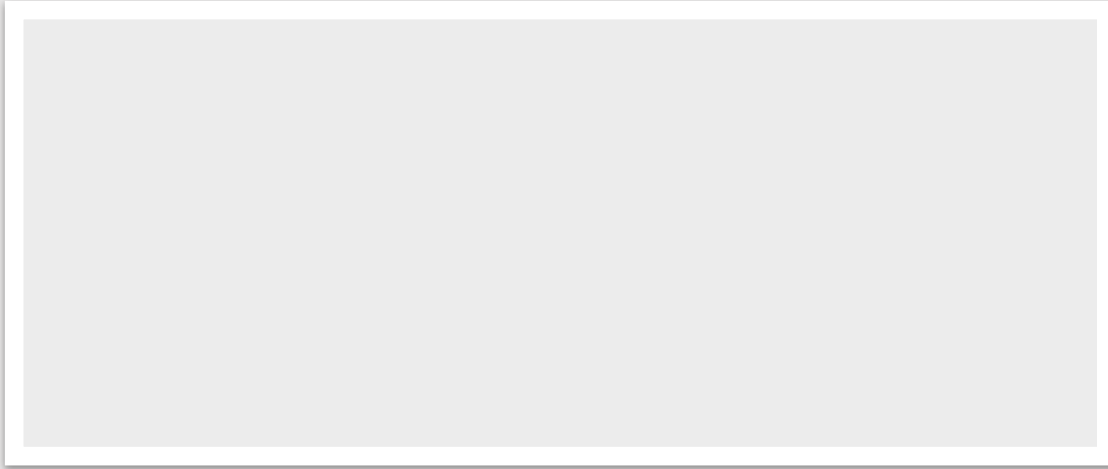


Semantic relations



Examples of **semantic relations** are the [equivalence](#) relation between two senses, [hyponymy](#) and [hyponymy](#) relations, [synonymy](#), [antonymy](#), [translations](#), etc..

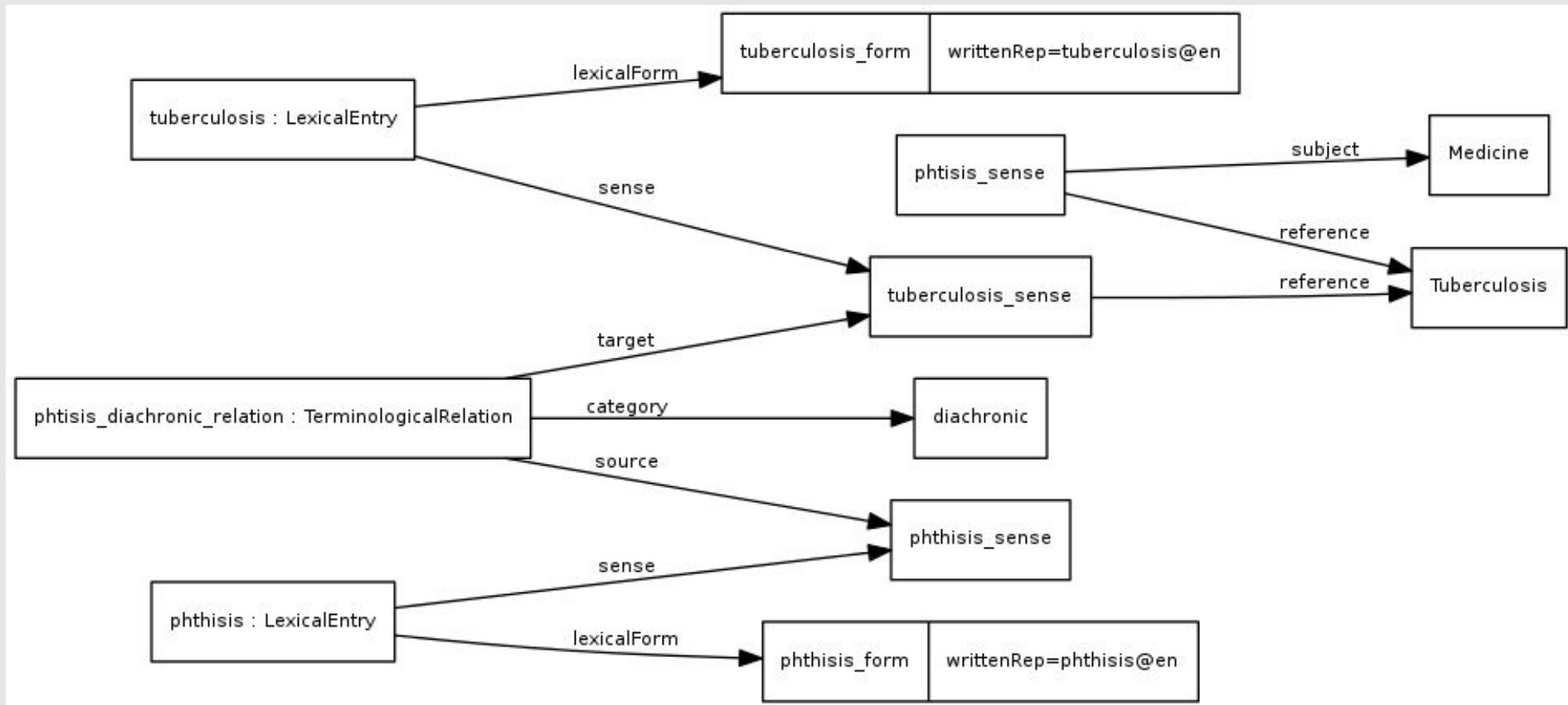
Terminological variants



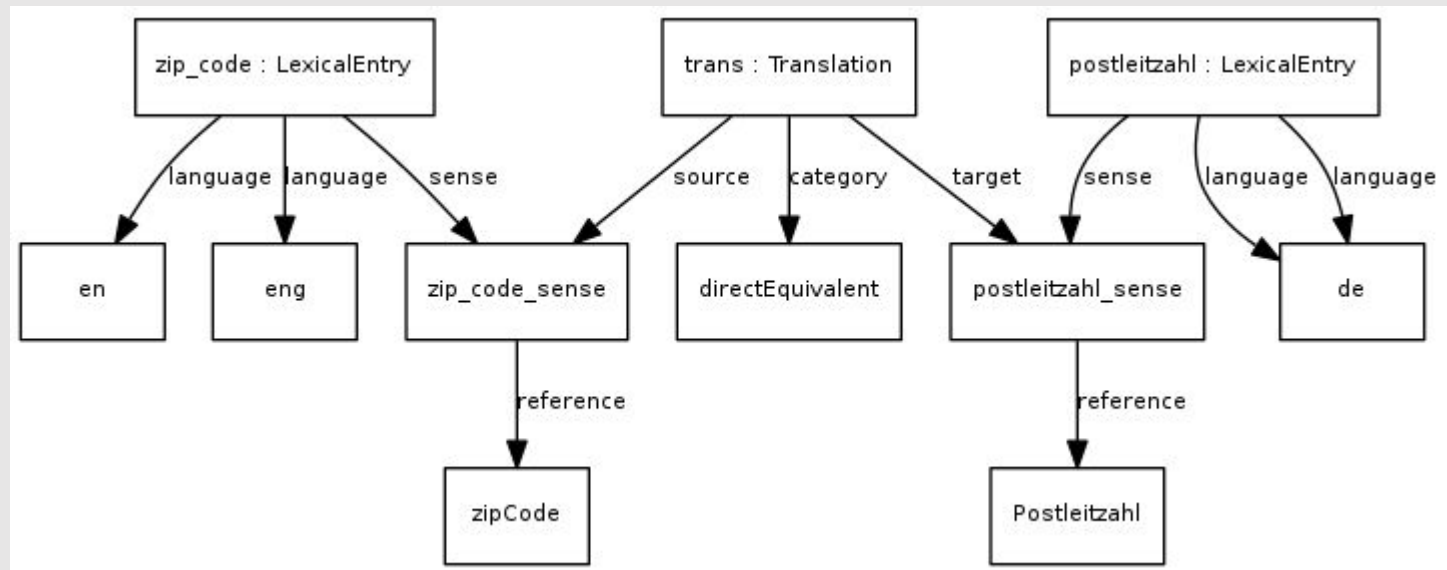
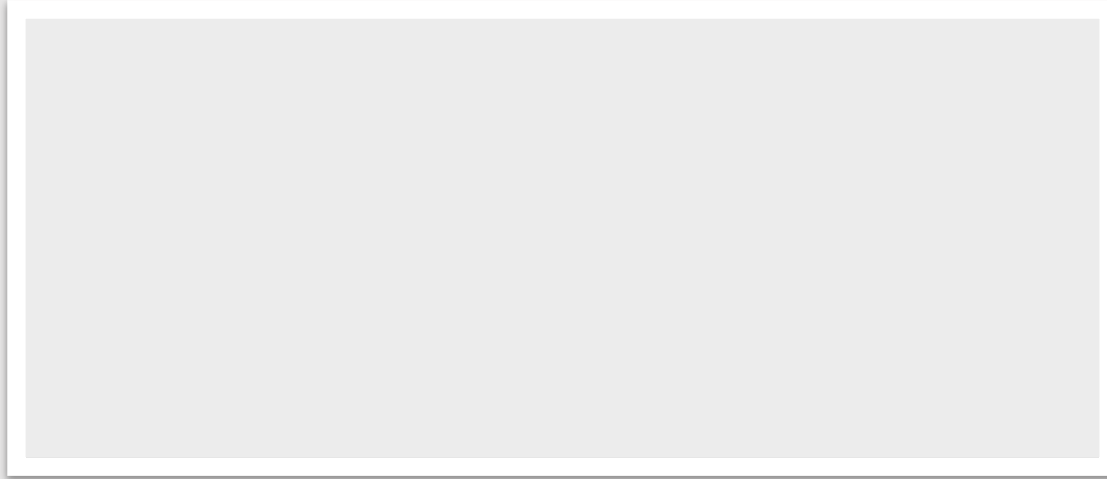
Examples of categories of **terminological variants (terminological relations)** include:

- **Diatopic** (dialectal or geographical variants) (e.g., gasoline vs. petrol)
- **Diaphasic** (register) (e.g., headache vs. cephalalgia; swine flu vs. pig flu vs. H1N1 vs. Mexican pandemic flu)
- **Diachronic** (or chronological variants) (e.g., tuberculosis vs. phthisis)
- **Diastratic** (discursive or stylistic variants) (e.g., man vs. bloke)
- **Dimensional** variants: the terms point to the same concept but highlight a different property or dimension of the concept (e.g., *bio-sanitary waste* vs. *hospital waste*)

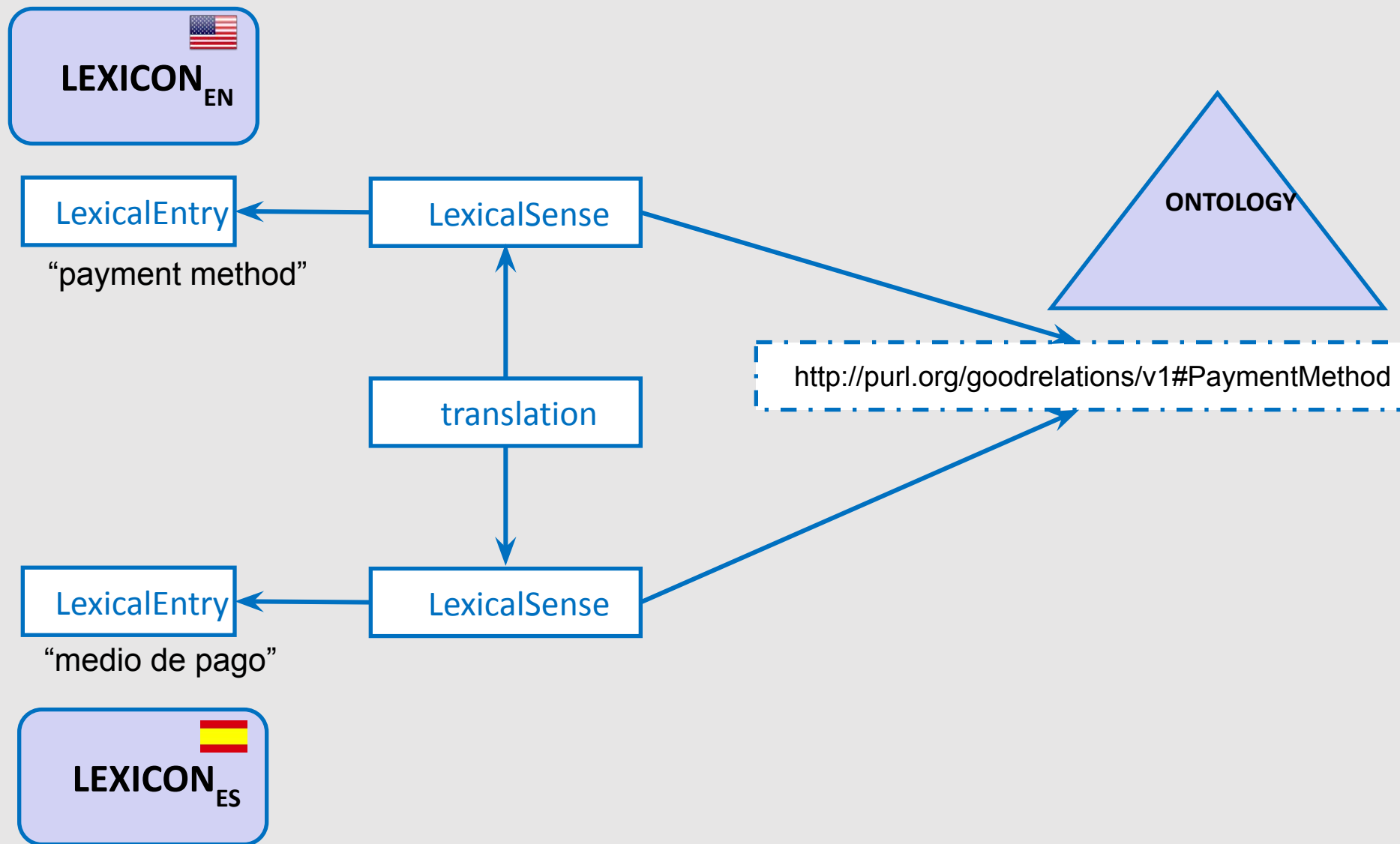
Terminological variants



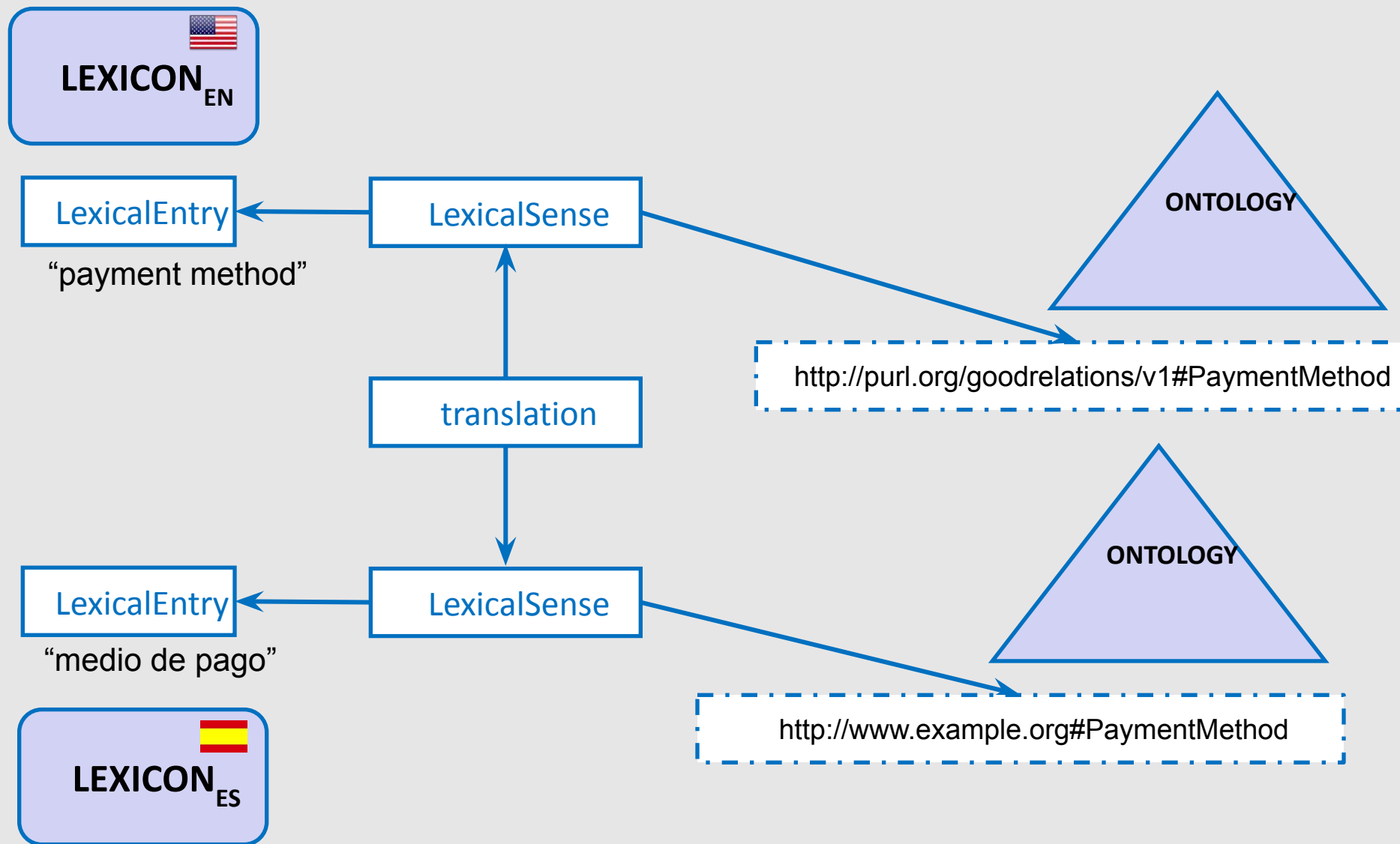
Translations



LLD – Ontolex-lemon: translations



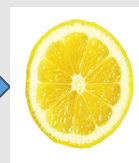
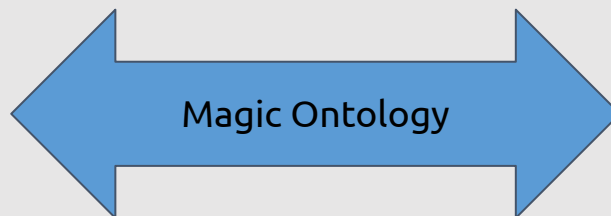
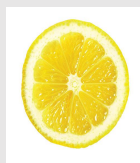
LLD – Ontolex-lemon: translations



Lexicon Metadata

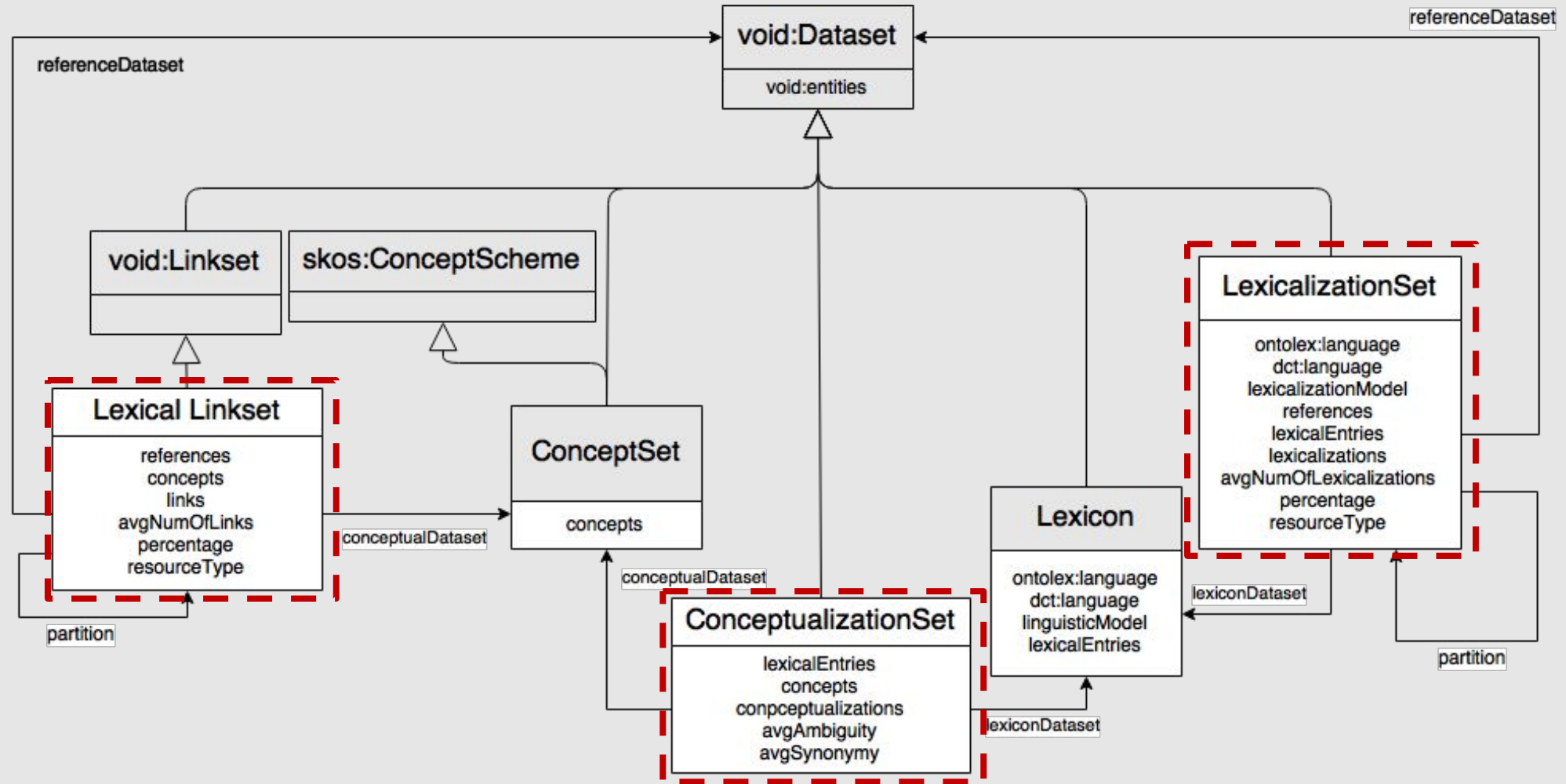


Jace the Wizard



Erhnam the Djinn

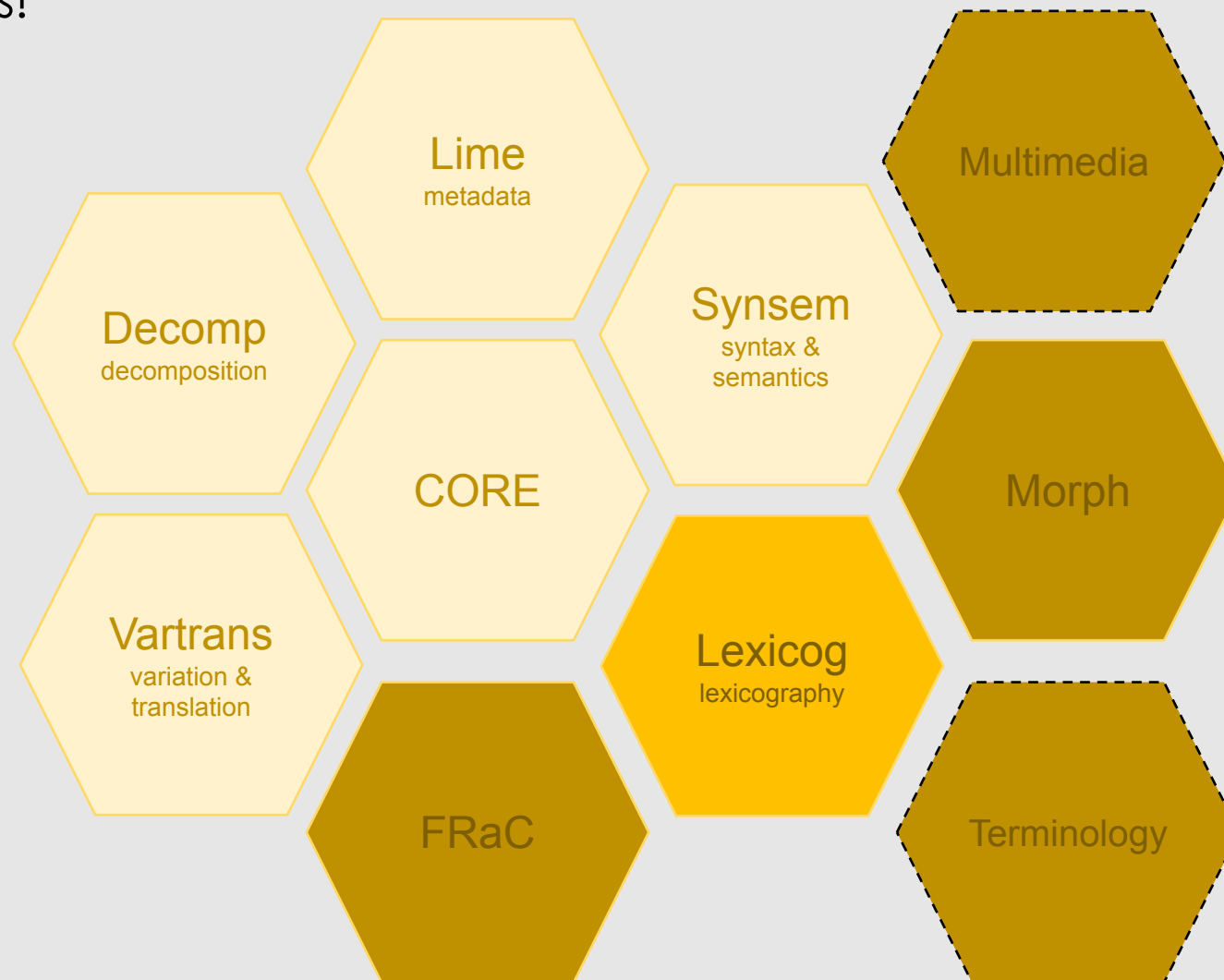
LIME



Coming Soon



New Lemon Modules!



Thanks!

Jorge Gracia (University of Zaragoza, Spain)
Fahad Khan (Institute for Computational
Linguistics «A. Zampolli»)