

THE MORDIGITAL PROJECT AND ITS APPROACH TO LEXICOGRAPHY AND LINGUISTIC DATA: Experiences in (Retro-)Digitising an Historical Dictionary Using Lexical Standards

Anas Fahad Khan, CNR-ILC, Italy
Ana Salgado, NOVA CLUNL, Portugal
Rute Costa, NOVA CLUNL, Portugal

Agenda

- ❑ This presentation aims to introduce MORDigital – Digitisation of *Dicionário da Língua Portuguesa* by António de Moraes Silva [PTDC/LLT-LIN/6841/2020], a Portuguese national project, and provide updates on its current progress.
- ❑ The project incorporates some of the latest advancements in **computational lexicography, digital humanities, ontologies** and **linguistic linked data**, including the modelling of lexicographic resources using standards such as TEI Lex-0 and OntoLex.
- ❑ Our focus will centre on MORDigital's approach to lexicography and linguistic data, specifically the project's integration of TEI Lex-0 and OntoLex-Lemon.

Introduction

- ❑ MORDigital's primary objective is to transform the first three historical editions of the Portuguese-language dictionary, *Diccionario da Lingua Portuguesa* (1789; 1813; 1823), also known as Moraes, into a digital resource.
- ❑ The central goal of the project is to convert the three editions of Moraes into a structured lexical resource, available in both TEI-XML (compliant with the ISO LMF standard) and in RDF (following the OntoLex-Lemon model and its recent extensions).
- ❑ These editions will also be made available via TEI Publisher.

Project Overview

Lexicographic analysis:
analysis of macro and
microstructural components.

**Digitisation of dictionary
sources:** OCR (Optical
Character Recognition) using
ABBYY FineReader.

**Automatic structuring of
lexical content:** Using
GROBID-Dictionaries, a
machine learning library for
structuring digitised lexical
resources.

**Conversion and mapping to
standards:** Transforming TEI
content into the LMF standard
and its respective
serialisations. Alignment with
OntoLex-Lemon.

**Terminological methods and
ontologies:** Applying
terminological methods to
organise knowledge and build
ontologies, addressing the gap
in general language
dictionaries.

Data linking and alignment:
Connecting entries, senses,
and other lexicographic
content between the three
editions.

Platform development:
Creating a platform for Morais,
enriched with both
lexicographic and ontological
modules.

Digital Humanities

OCR Tasks: ABBYY FineReader was used for Optical Character Recognition, enabling the conversion of printed text into machine-readable content.

Structuring Digital Editions: We employ GROBID-Dictionaries, a machine learning library, to structure digitised editions, ensuring the data is well-organized and accessible.

Data Encoding: The data is encoded using TEI Lex-0 within the Oxygen framework, adhering to standards and facilitating consistency in the project.

XSLT Transformation: We are in the process of testing XSLT transformations to OntoLex and its extensions, allowing for efficient data integration and compatibility. The TEI-XML sources will subsequently be converted to OntoLex (both the original model and its follow-up modules) using an XSLT stylesheet.

Ontologies: We use Protegé to develop ontologies, enabling the representation of knowledge.

Data Publication: To share our results with the wider community, TEI Publisher has been chosen as the platform for publishing the data.

TEI Lex-0

- ❑ TEI Lex-0 is a customisation of Text Encoding Initiative (TEI P5): <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>
- ❑ It is a baseline encoding and a target format to facilitate the interoperability of heterogeneously encoded lexical resources (Tasovac et al. 2018)
- ❑ DARIAH Working Group for Lexical Resources: <https://bit.ly/tei-lex-0>

TEI Lex-0

- ❑ The att.datcat changes are now official in TEI 4.7.0:

<https://www.tei-c.org/Vault/P5/4.7.0/doc/tei-p5-doc/en/html/ref-att.datcat.html>

The [att.datcat](#) attributes can be used for any sort of taxonomies. The example below illustrates their usefulness for describing usage domain labels in dictionaries on the example of the *Dicionário da Língua Portuguesa* by António de Morais Silva, retro-digitised in the [MORDigital project](#).

```
<!-- in the dictionary header --><encodingDesc>
<classDecl>
  <taxonomy xml:id="domains">
<!--...-->
    <category xml:id="domain.medical_and_health_sciences">
      <catDesc xml:lang="en">Medical and Health Sciences</catDesc>
      <catDesc xml:lang="pt">Ciências Médicas e da Saúde</catDesc>
      <category xml:id="domain.medical_and_health_sciences.medicine"
        valueDatcat="https://vocabs.rossio.fcsh.unl.pt/pub/morais_domains/pt/page/0025">
        <catDesc xml:lang="en">
          <term>Medicine</term>
          <gloss>
<!--...-->
            </gloss>
          </catDesc>
          <catDesc xml:lang="pt">
            <term>Medicina</term>
            <gloss>
<!--...-->
              </gloss>
            </catDesc>
          </category>
        </category>
      </category>
    </taxonomy>
  </classDecl>
</encodingDesc>
<!--
      inside an <entry> element: -->
  <usg type="domain"
    valueDatcat="#domain.medical_and_health_sciences.medicine">Med.</usg>
```

In the Morais dictionary, the relevant domain labels are in the header, getting referenced inside the dictionary, from [usg](#) elements. The vocabulary used for dictionary-internal labelling is in turn anchored in the [MorDigital controlled vocabulary service](#) of the NOVA University of Lisbon – School of Social Sciences and Humanities (NOVA FCSH).

```
<taxonomy xml:id="domains">
  <category xml:id="domain.mathematical_sciences"
    valueDatcat="http://www.semanticweb.org/OntoDomLab-Math#MathematicalSciences"
    http://vocabs.rossio.fcsh.unl.pt/morais_domains/0036">
    <catDesc xml:lang="en">
      <term>Mathematical Sciences</term>
      <!-- Source of gloss:
      https://dbpedia.org/page/Mathematical_sciences-->
      <gloss>Group of areas of study that includes, in addition
        to mathematics, those academic disciplines that are primarily
        mathematical in nature but may not be universally considered
        subfields of mathematics proper.</gloss>
    </catDesc>
    <catDesc xml:lang="pt">
      <term>Ciências Matemáticas</term>
    </catDesc>
    <category xml:id="domain.mathematics"
      valueDatcat="http://www.semanticweb.org/OntoDomLab-Math#Mathematics"
      http://vocabs.rossio.fcsh.unl.pt/morais_domains/0024">
      <catDesc xml:lang="en">
        <term>Mathematics</term>
      </catDesc>
      <catDesc xml:lang="pt">
        <term>Matemática</term>
      </catDesc>
    </category>
  </category>
</taxonomy>
```

```
<entry xml:id="MORAIS.DLP.1.TRIGONOMETRIA" type="mainEntry" xml:lang="pt">
  <form type="lemma">
    <orth>TRIGONOMETRIA</orth>
  </form>
  <metamark function="lemmaDelimiter">,</metamark>
  <gramGrp>
    <gram type="pos" norm="NOUN">f.</gram>
    <gram type="gen">f.</gram>
  </gramGrp>
  <sense xml:id="MORAIS.DLP.1.TRIGONOMETRIA.s.1">
    <usg type="domain" valueDatcat="#domain.mathematics" resp="#Salgado"/>
    <def>parte da Mathematica , que enfina a refolver os triangulos planos ,
    e esfericos</def>
  </sense>
  <pc>.</pc>
</entry>
```


TEI Publisher

- TEI Lex-0 Publisher, a customisation of the TEI Publisher for lexicographic resources.

- Users can seamlessly navigate through the dictionary, locate specific entries through basic queries, or conduct more intricate searches by combining multiple parameters.

- Collapsed and expanded entry.

- The incorporation of facsimile display adds an extra layer of depth to the user experience.



The screenshot displays the TEI Publisher web interface, which is a customised version of the TEI Publisher for lexicographic resources. The interface features a top navigation bar with links for Start, Documentation, News, DTS, Download, and a search bar. The main content area shows a facsimile of a dictionary page with entries for ITI, JAB, and J. The entries are presented in a two-column layout, with the left column showing the entry text and the right column showing the facsimile image. The facsimile image is a high-resolution scan of a printed dictionary page, showing the original text and layout. The entries are as follows:

ITI
ITEM, *adv. lat.* significa também ; ufamos delle, quando fe fazem varios articulos , e enumeração de coifas , nas leis v. g. ,, *prohibo que entrem chapeos , item meias de seda , item joias , etc. § sublt. Effar aos itens com alguem , i. e. á conta com elle , e f. em alterações ; em recados , e repoltas . Cajlan. 3. f. 136. § f. Por-se o espirito aos itens com a carne , disputar-lhe a victoria , ou tomar contas a consciencia ás paixões Conspiração f. 333.*
ITINERARIO , *f. m.* livro em que fe contém a descrição da jornada , ou viagem que fe fez v. g. ,, *o Itinerario da Terra Santa de Anronto Tenreiro . Barros 1. f. 171. v. a modo de itinerario maritimo.*
ITINERARIO , *adj.* que respeita a caminhos v. g. ,, *medida —*

JAB
dezeje ; nunca já tal farei ; já disse são sofregas ,, *Enfr. f. 207. § Talvez se repete o adv. para dar a entender que caimo s no que não nos occorria v. g. ,, já , já , disse o cavalleiro , entendido fols vós . B. Clar. f. 146. col. 1. Vilhalp. Aro 5. sc. 2. Ferreira. Ciofo Aro 4. sc. 6. § Já ufa-fe subltant. , ou com preposição exprella , v. g. ,, desde já , ou desde este momento.*
JABOTICABA , *f. f.* fruto da jaboticabeira , Brasil. , he redondo como huma grande cereja negra ; a casca não fe come , e he mui altringente ; tem hum succo mui doce , e caroço esponjofo ; nasce pegado immediatamente aos troncos , e ramos , da arvore. *Vafconcellos Not. f. 265.*
JABOTICABEIRA , *f. f.* arvore grande , de tronco , e ramos mui lifos , casca delgada , que perde annualmente ; tem a folha pequena , da feição de lança mui aguda ; da a jaboticaba , e vive no Brasil.

J
J , *f. m.* confoante , que modifica o som das vogaes a que procede do mesmo modo , que o g antes do e , e do i vulgarmente lhe chamão *i confoante* ; denominação absurda , porque estas letras nada tem de commum , nem na figura , nem na essencial differença , porque i representa hum som , ou vogal ; e j representa a modificação de hum som , ou confoante : melhor fe lhe chamaria *je*.
JA' , *adv.* neste tempo , a este momento v. g. ,, *já vejo , já está feito . § Já mais , nunca , em nenhum tempo. Ulisses 2. 79. § Neste momento , sem demora v. g. ,, saia , parta já , e não se demore . Noutro tempo , quando se une*

Linked (Open) Data

- ❑ “[L]inked data is **structured data** which is **interlinked** with other data so it becomes more useful through **semantic queries**. It builds upon standard Web technologies such as **HTTP, RDF and URIs**, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be **read automatically by computers**. Part of the vision of linked data is for the Internet to become a global database.”
 - Source: [Wikipedia](#) (emphases mine)
- ❑ Emphasis is here on the **use of standards** in creating datasets, **linking between datasets** (to avoid data silos) using **web technologies**, and making the **meaning of data** clear enough that it can be easily processed by machines.
- ❑ The **Semantic Web** is a web of datasets that are structured and linked together using a common set of **standards and technologies** so that they can be more easily processed by computers in terms of what they ‘mean’, unlike normal hypertext documents which are designed to be read by humans. Linked Data is one very important way of making the semantic web a **reality**.

Linked (Open) Data

Tim Berners-Lee's statement of the principles of linked data

- ❑ **All conceptual things** should have a name starting with **HTTP**.
- ❑ Looking up an HTTP name should return **useful data** about the thing in question in a **standard format**.
- ❑ Anything else that that same thing has a **relationship** with through its data should also be given a name beginning with HTTP.

- ❑ These names are **URIs**: identifiers that uniquely identify resources across the whole Semantic Web (**global IDs**)
 - **Uniform Resource Locators (URLs)** are also a kind of URI

- ❑ If in addition to these previous prerequisites we make the data available under an open license then our data becomes **Linked Open Data (LOD)**.
 - The collection of *linked data datasets* published with an open license constitutes the linked open data cloud.

Resource Description Framework

- ❑ With linked data we make **everything a resource** and give it an **URI** that we **can look up** ('dereference') in the same way we do with web pages. But we would like to talk about such resources, describe them and the relationships between them. To say things like, e.g.,
 - **'Paris is a city, located in France'** or
 - **'Emmanuel Macron is the President of France'**
- ❑ Where 'Paris', 'city', 'France', 'Emmanuel Macron', 'President of France' are all represented as resources with URIs.
- ❑ **The Resource Description Framework (RDF)** is a model, for describing resources and relating them together.

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

http://ur.dbpedia.org/resource/ایمانویل_میکخواں

[rdf:type](#)

[owl:sameAs](#)

https://dbpedia.org/resource/Emmanuel_Macron



<http://dbpedia.org/property/title>

http://dbpedia.org/resource/President_of_France

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

http://ur.dbpedia.org/resource/ایمانویل_میکخواں

[rdf:type](#)

[owl:sameAs](#)

https://dbpedia.org/resource/Emmanuel_Macron



http://dbpedia.org/resource/Emmanuel_Macron

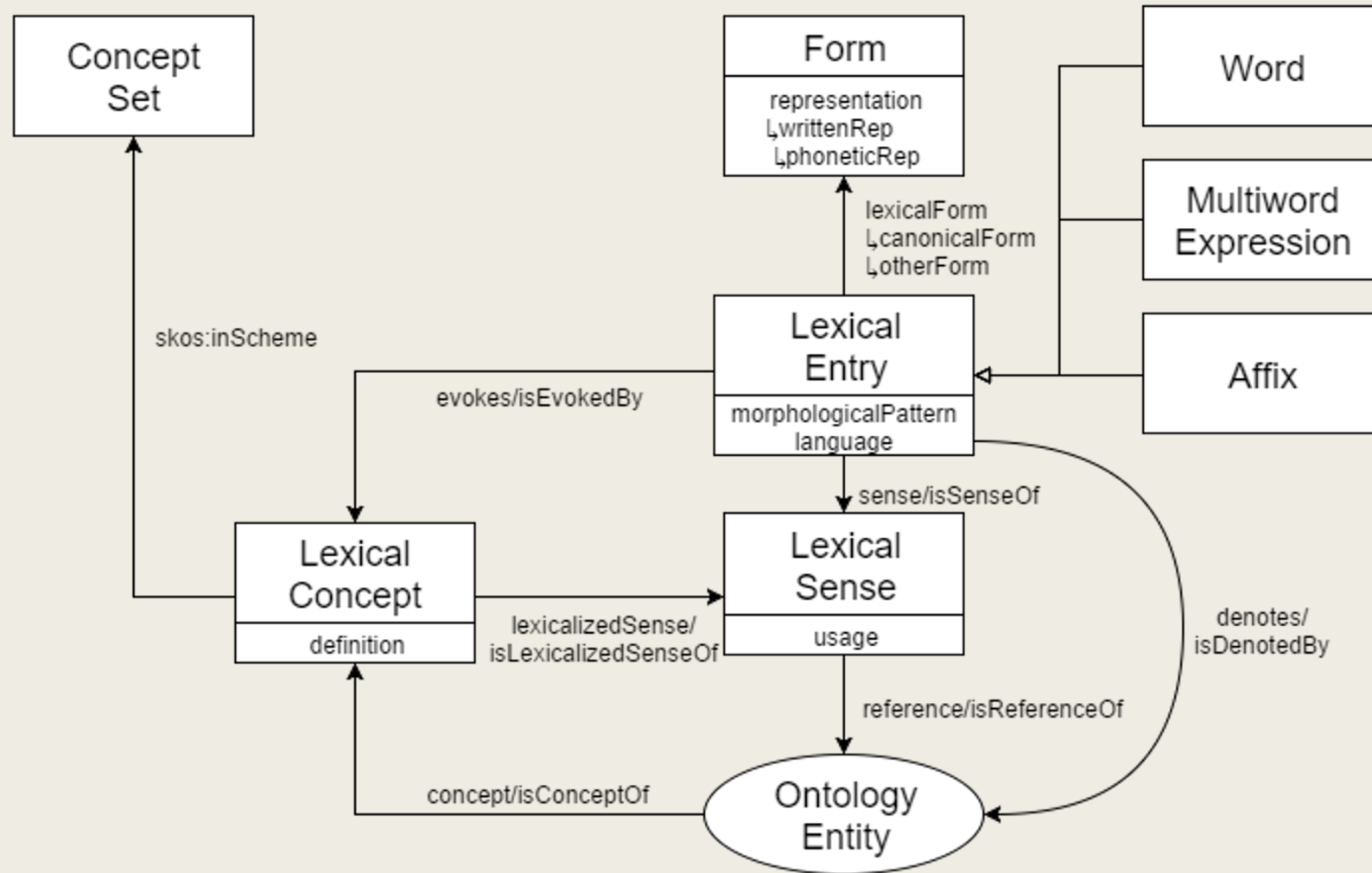
http://dbpedia.org/resource/Emmanuel_Macron

Idea: Let's create Linked Data graphs for lexical resources!

Linguistic Linked (Open) Data

- ❑ Linguistic Linked Open Data refers to linguistic datasets published as linked open data
 - Linguistic Linked Open Data has its own sector in the Linked Data Cloud
- ❑ Kinds of resources available on the linguistic Linked Open Data cloud include **corpora**, **terminologies**, **linguistic data categories**, and **lexicons**
- ❑ For lexicons the *de facto* standard vocabulary/ontology used to model and publish linked data lexicons is **OntoLex-Lemon** which is based on the Lexical Markup Framework but adapted to the RDF data model
- ❑ OntoLex-Lemon consists of a number of different modules, including a core module, a module for syntax and semantics, translations, and a module for lexicography (lexicog)

OntoLex-Lemon



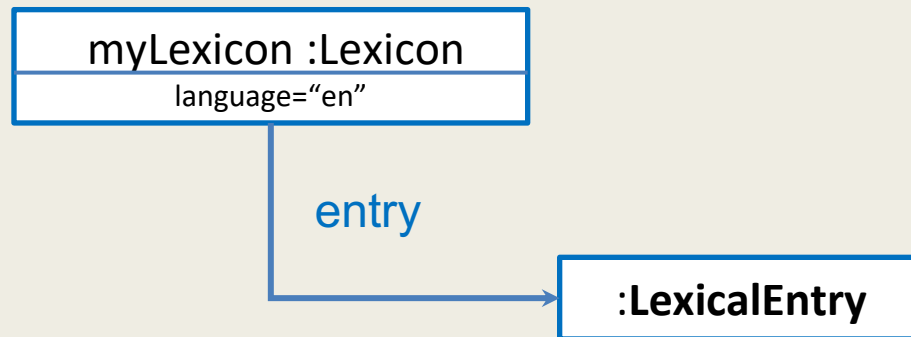
Ontolex-lemon: an example

Lexicon: The object representing the lexicon as a whole.

myLexicon : Lexicon
language="en"

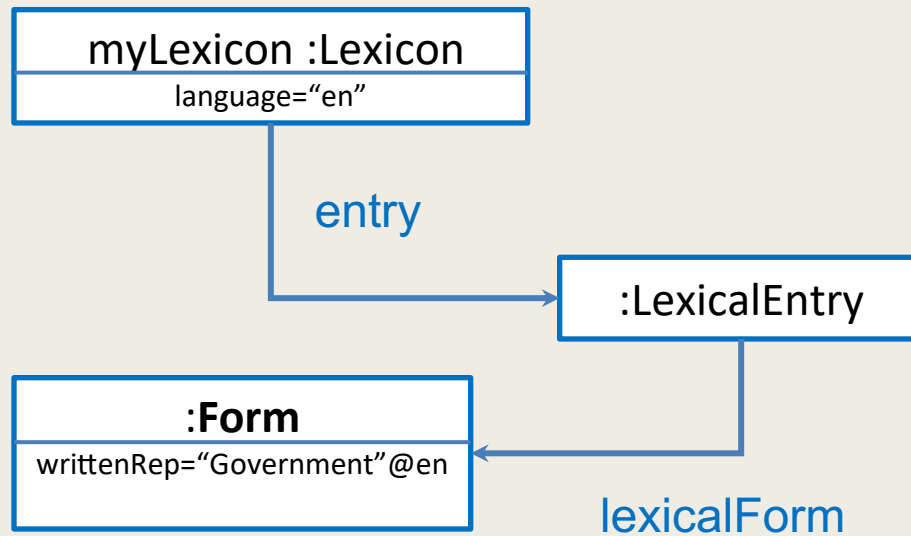
Ontolex-lemon: an example

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



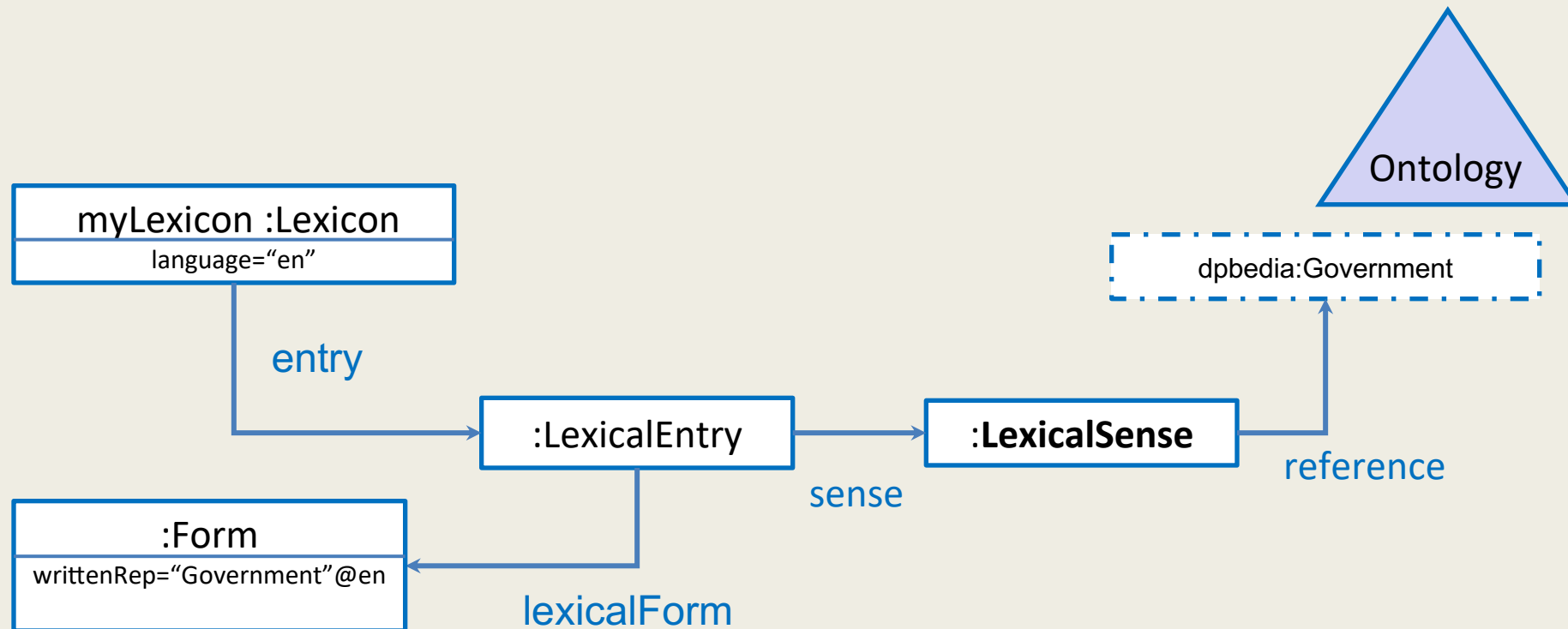
Ontolex-lemon: an example

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



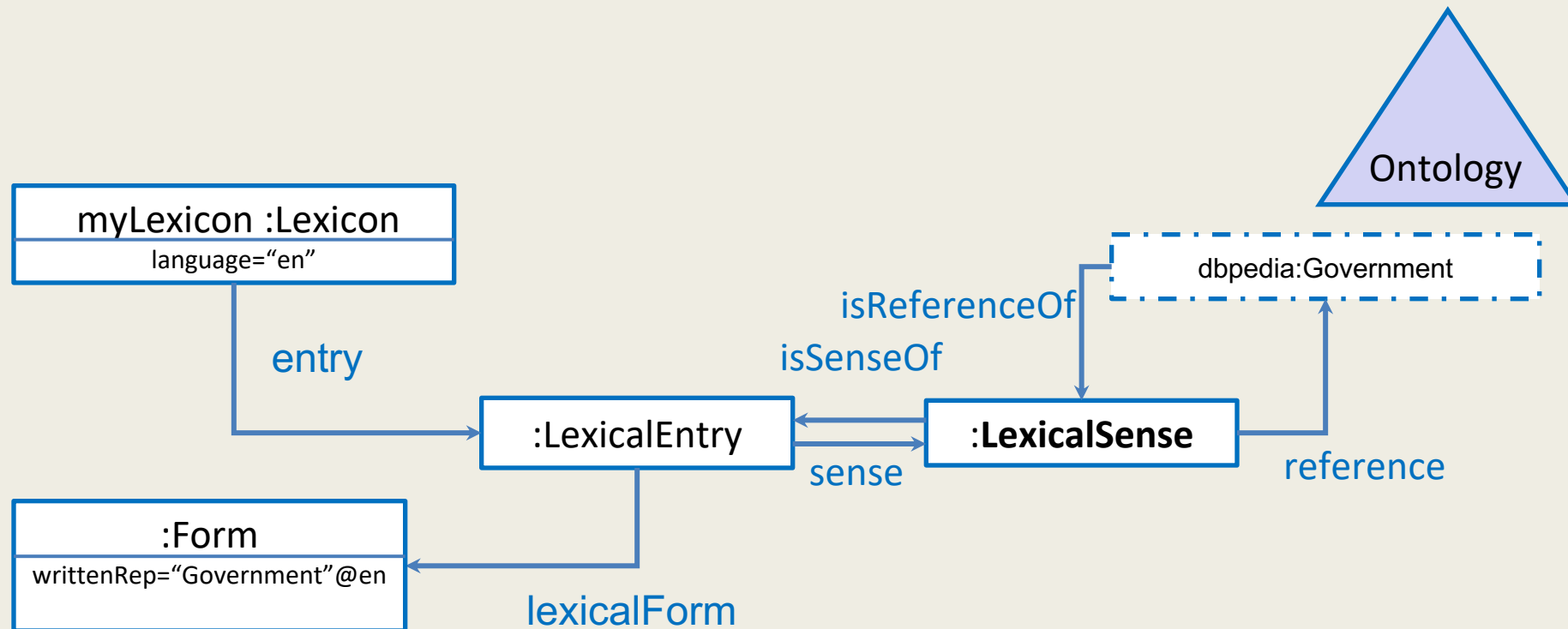
Ontolex-lemon: an example

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



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



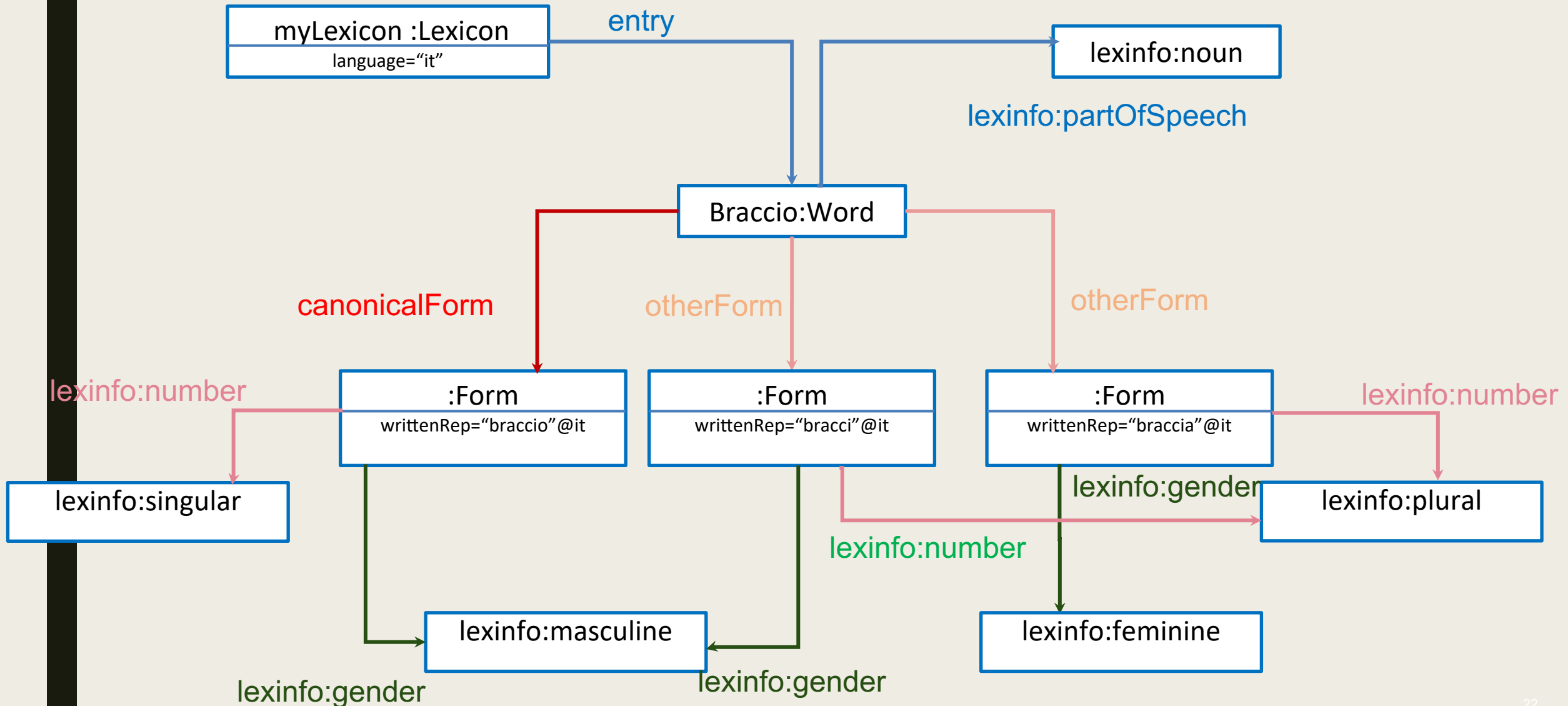
```
:government a ontollex:LexicalEntry ;  
    ontollex:canonicalForm [  
        ontollex:writtenRep "Government"@en ] ;
```

```
    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.'meɪtəʊ"@en-US-form;  
  ontollex:phoneticRep "'tə'mɑ: .təʊ"@en-GB-form.
```



Motivation for Publishing Lexicons in Linked Data

- ❑ Once we have our lexicons published as linked data we can use SPARQL to write queries on top of it -- and allow others to do the same remotely.
- ❑ E.g., the LiLa project made the whole legacy Lewis-Short Latin-English dictionary available online in linked data, and queryable via a SPARQL endpoint. Here are the results of a recent query I wrote to this endpoint to look for semantic shifts in medical terms:
 - <https://lila-erc.eu/sparql/>

PREFIX lila: <http://lila-erc.eu/ontologies/lila/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX lime: <http://www.w3.org/ns/lemon/lime#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ontollex: <http://www.w3.org/ns/lemon/ontollex#>

Searches for all lemmas in LiLa LemmaBank whose definition in the LS dictionary includes the word "grow"

```
SELECT DISTINCT ?lemmaLabel ?leid ?sense ?ps ?gn ?original_definition ?definition ?le ?lemma ?lc WHERE {  
  VALUES ?lemmaLabel {"anus" "aranea" "cancer" "caninus" "causa" "caulis" "fistula" "folliculus" "impetus" "lenticula" "lumen" "malum" "menstrua" "mola"  
"musculus" "patella" "pecten" "plaga" "potio" "pupilla" "scrotum" "spina" "spiritus" "tibia" "vitium" "uitium"}
```

```
  ?le ontollex:canonicalForm ?lemma ;  
    ontollex:sense ?lc .  
  ?lc <http://www.w3.org/2004/02/skos#definition> ?def.  
  ?lemma ontollex:writtenRep ?wf .  
  ?lemma rdfs:label ?lemmaLabel .  
  <http://lila-erc.eu/data/lexicalResources/LewisShort/Lexicon> lime:entry ?le.
```

```
OPTIONAL {  
  ?lemma <http://lila-erc.eu/ontologies/lila/hasPOS> ?pos ;  
    <http://lila-erc.eu/ontologies/lila/hasGender> ?gen .  
  BIND(REPLACE(STR(?pos), "^.*(?:/[^/]*)$", "$1") AS ?ps)  
  BIND(REPLACE(STR(?gen), "^.*(?:/[^/]*)$", "$1") AS ?gn)  
}
```

```
BIND(REPLACE(STR(?lc), "^.*(?:/[^/]*)$", "$1") AS ?sense)  
BIND(REPLACE(STR(?le), "^.*(?:/[^/]*)$", "$1") AS ?leid)  
BIND(REPLACE(STR(?lc), "(\\.[0-9]+)(?!\\.\\.[0-9]+)", "\\0") AS ?zeros)  
BIND(URI(?zeros) AS ?zero_uri)
```

```
  ?zero_uri <http://www.w3.org/2004/02/skos#definition> ?original_definition .  
  BIND(IF(?def = ?original_definition, "", ?def) AS ?definition)  
}  
ORDER BY ?lemmaLabel ?sense
```


1	anus	n2940	n2940.0	noun	ma...	the posteriors , fundament.	
2	anus	n2940	n2940.2	noun	ma...	the posteriors , fundament.	Meton. , disease of the anus, piles , hemorrhoids Lat.): bis;
3	anus	n2941	n2941.0	noun	ma...	an iron ring for the feet	
4	anus	n2942	n2942.0	noun	fe...	an old woman (married or unmarried), a matr...	
5	anus	n2942	n2942.2	noun	fe...	an old woman (married or unmarried), a matr...	Transf. , old , aged (cf. senex, old; old man , sometimes ...
6	aranea	n3328	n3328.0	noun	fe...	A spider	
7	aranea	n3328	n3328.2	noun	fe...	A spider	A spider's web , cobweb
8	aranea	n3328	n3328.3	noun	fe...	A spider	Also, for threads similar to spiders' webs
9	cancer	n6451	n6451.0	noun	ma...	gen. ; acc. plur. (neutr. , cornu], a crab , a river-...	
10	cancer	n6451	n6451.2	noun	ma...	gen. ; acc. plur. (neutr. , cornu], a crab , a river-...	Meton. , hands that cling fast like the claws of crabs
11	cancer	n6451	n6451.3	noun	ma...	gen. ; acc. plur. (neutr. , cornu], a crab , a river-...	As nom. propr. , the Crab , the sign of the zodiac in whic...
12	cancer	n6451	n6451.4	noun	ma...	gen. ; acc. plur. (neutr. , cornu], a crab , a river-...	Poet. for the region of the south , the south
13	cancer	n6451	n6451.5	noun	ma...	gen. ; acc. plur. (neutr. , cornu], a crab , a river-...	To designate great or violent heat
14	cancer	n6451	n6451.6	noun	ma...	gen. ; acc. plur. (neutr. , cornu], a crab , a river-...	In medicine, a crawling , eating , suppurating ulcer , mal...
15	cancer	n6452	n6452.0	noun	ma...	a lattice	
16	caninus	n6504	n6504.0			of or pertaining to a dog , canine , dog-.	
17	caninus	n6504	n6504.3			of or pertaining to a dog , canine , dog-.	Trop. : prandium, in which no wine is drunk , mean (v. th...
18	caulis	n7180	n7180.0	noun	ma...	the stalk or stem of a plant .	
19	caulis	n7180	n7180.2	noun	ma...	the stalk or stem of a plant .	Pennae, a quill
20	caulis	n7180	n7180.3	noun	ma...	the stalk or stem of a plant .	The stem or bony part of an ox ' s tail
21	caulis	n7180	n7180.4	noun	ma...	the stalk or stem of a plant .	In insects, a tube by which eggs are deposited
22	caulis	n7180	n7180.5	noun	ma...	the stalk or stem of a plant .	= membrum virile, Lucil. ap.
23	causa	n7197	n7197.0	noun	fe...	that by , on account of , or through which any t...	

TEI to OntoLex

- ❑ Both TEI and OntoLex seem to have their own advantages and disadvantages, e.g., TEI allows us to remain closer to the original text whereas OntoLex makes things easier for machine processing, makes linking to other resources (linguistic and non linguistic) simpler, makes remote querying with a powerful query language easier (i.e., all using standard out of the box technologies)
- ❑ In MORDigital we decided to publish the Morais dictionary as both TEI and OntoLex to take advantage of both formats
- ❑ We are using an XSLT stylesheet to move between from TEI-XML to OntoLex (serialised in RDF-XML) with the output being post-edited by hand. In future work we are looking at the use of RDFa to insert RDF triples into a TEI source
- ❑ In the next few slides we will look at example encoding both in TEI (the original) and OntoLex

TEI to Ontolex

- ❑ We look at the entry for *citerior* from Moraes, adjective that means ‘situated this side of some place or location’.

CITERIOR, adj. que fica áquém de algum posto, ou sitio. *M. Lus.* usa-se na *Geograf.* „
Hespanha citerior, e ulterior.

TEI to Ontolex

```
<entry xmlns="http://www.tei-c.org/ns/1.0" xml:id="MORAIS.1.DLP.CITERIOR" type="mainEntry"
xml:lang="pt">
  <form type="lemma">
    <orth>CITERIOR</orth>
  </form>
  <metamark function="lemmaDelimiter">,</metamark>
  <gramGrp>
    <gram type="pos" norm="ADJECTIVE">adj.</gram>
  </gramGrp>
  <sense xml:id="MORAIS.1.DLP.CITERIOR.s.1">
    <def>que fica áquem de algum pofto, ou fítio</def>
    <pc>.</pc>
    <cit type="example">
      <bibl type="attestation" corresp="mor.bibl.authr.titl">
        <!-- point to Monarchia Lufitana -->
        <title>M. Luf.</title>
        <metamark function="quotationDelimiter">,</metamark>
        <lbl>usa-se na</lbl>
        <usg type="domain">t. Geograf.</usg>
        <quote>Hespanha citerior, e ulterior</quote>
        <pc>.</pc>
      </bibl>
    </cit>
  </sense>
</entry>
```

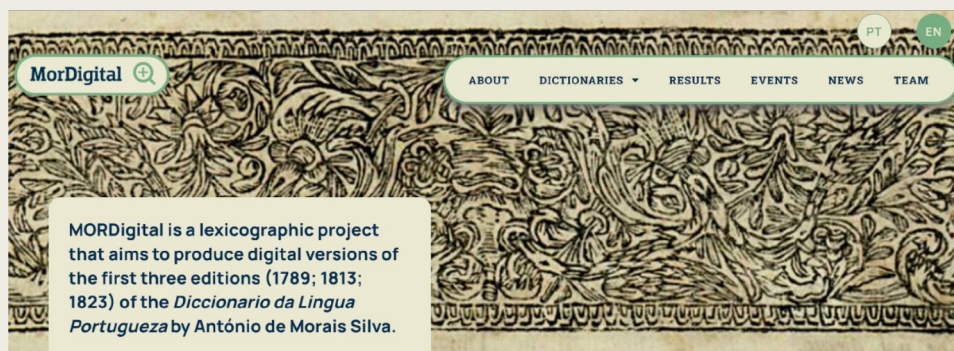

TEI to Ontolex

```
morais:citerior a ontolex:LexicalEntry ;
                    ontolex:canonicalForm [
ontolex:writtenRep "CITERIOR"@pt ] ;
                    lexinfo:partOfSpeech
lexinfo:adjective ;
                    ontolex:sense [
skos:definition "que fica áquem de
algum pofto, ou fitio"@pt ;
                    lexicog:usageExample [
dcterm:source " M. Luf. ,, usa-se na
t. Geograf. Hespanha citerior, e
ulterior . ";
dcterm:subject
<http://example.org/class/geografia>]
] .
```


Conclusion

- ❑ Knowledge organisation is crucial to improve the lexicographic content.
- ❑ TEI enables a thorough and structured markup of textual content, while Ontolex facilitates representation as linked data, allowing the content to be integrated with data from external lexical resources, which may not be limited to dictionaries.
- ❑ The linked data conversion of the Morais dictionary enables a further way to explore and enrich its data, and will constitute an important contribution to the Portuguese section of the linguistic linked open data cloud.

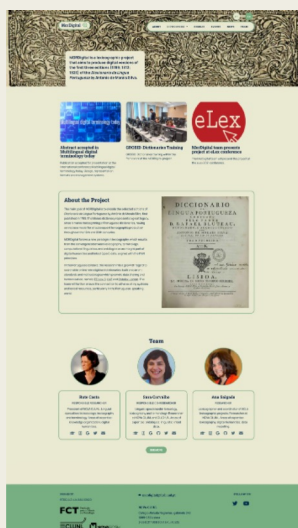
If you want to know more...



<https://mordigital.fcsh.unl.pt/>

Acknowledgements

- MORDigital – Digitalização do Dicionário da Língua Portuguesa de António de Moraes Silva [PTDC/LLT- LIN/6841/2020] project financed by the Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia
- Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia as part of the project Centro de Linguística da Universidade NOVA de Lisboa – UID/LIN/03213/2020



Thank You!

Anas Fahad Khan: fahad.khan@ilc.cnr.it

Ana Salgado: anasalgado@fcsh.unl.pt

Rute Costa: rute.costa@fcsh.unl.pt