

MorDigital

Problèmes terminologiques dans les dictionnaires anciens

Rute Costa & Ana Salgado

NOVA FCSH, Lisbonne

8 février 2023

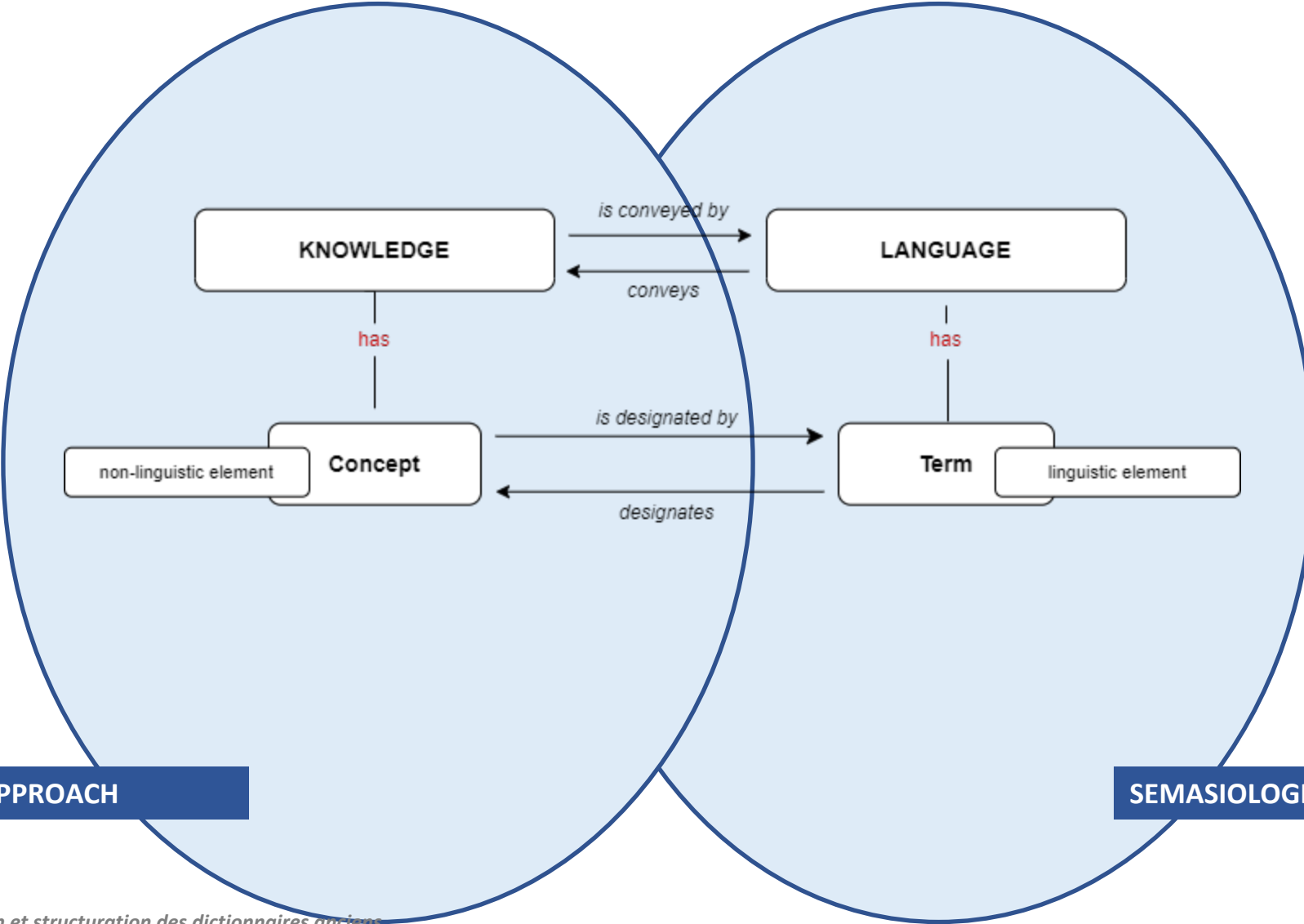
Theoretical framework

- Lexicography has undergone a radical change in the past two decades.
- This paradigm shift is directly related to the advancement of digital humanities.
- The perspective underpinning the construction of lexical resources presupposes rethinking the methodologies of the Portuguese lexicographic tradition, perceiving lexicography, terminology, ontologies and computational linguistics as an integral part of the digital humanities, which will imply a paradigm shift in the construction of dictionary resources.
- In this new paradigm, ontologies will play a key role in organising and representing linguistic and metalinguistic knowledge (Carvalho, Costa & Roche, 2018; Almeida, Costa & Roche, 2019), as well as supporting its operationalisation and, therefore, its preservation in the long term.

Theoretical framework

Concept
'unit of knowledge created by a unique combination of *characteristics*' (ISO 1087, 2019, p. 3)

Term
'*designation* that represents a *general concept* by linguistic means' (ISO 1087, 2019, p. 7)



ONOMASIOLOGICAL APPROACH

SEMASIOLOGICAL APPROACH

Theoretical framework

- The European lexicographic scenario is currently quite heterogeneous, in what concerns
 - (1) the types of existing lexicographic resources;
 - (2) their structural component, which relates to how the data are represented, the adopted models, as well as the respective applied formats.
- The diversity of incompatible formats creates severe problems in the digital landscape, making it impossible to interconnect resources and their respective metadata and lexical data ☐ importance of following compatible standards and formats such as LMF (ISO 24613: 2008), TEI Lex-0 (Tasovac and Romary et al., 2018) and Ontolex-Lemon (McCrae et al., 2017).

Organising knowledge

EXPLICAÇÃO D A S ABBREVIATURAS USADAS NESTE DICCIONARIO.

adj.	termo	Adjectivo.
adv.		Adverbio, ou adverbial.
Agric.		Agricultura.
Anat.		Anatomia, ou Anatomico.
Ant. ou antiq.		antiquado.
Archit.		d'Architectura.
Arithm.		Arithmetico.
Artelh.		d'Artelharia.
Afiat.		usado na India Portug.
Astrol.		Astrologico.
Astron.		Astronomico.
At.		Verbo ativo.
Aument.		augmentativo.
Botan.		Botanico.
Bras.		do Brasão.
(C. ou		Capitulo.
(Cap.		
Chim.		Chimico.
Cirurg.		Cirurgico.
Com.		Commun de dois.
Compar.		Comparativo.
Conj.		Conjunção.
(Ch.		r. Chulo.
(Chul.		
(Chron. ou		Chronica.
(Cron.		
Dim.		Diminutivo.
Ed.		edicação : ult. ultima.
Escult.		Escultura.
F.		femenino.
Fam.		familiar.
Fr.		fraz.
Fraze prov.		fraze proverbial.
Filof.		Filosophico.
Fific.		da Fifica.
Fortif.		da Fortificação.
Freq.		frequentemente.
Geogr.		Geografico.
Geometr.		Geometrico.
Grammat.		da Grammatica.
I. e.		isto he.
Interj.		Interjeição.
Irreg.		Irregular.
Jurid.		Juridico.
Jurisp.		da Jurisprudencia.
L.		Livro nas citações dos Autores.

Sample of the flat front matter list of Morais dictionary (Morais Silva, 1789, p. xi)

Organising knowledge

- TEI Lex-0 recommends that canonical labels should be defined in the <teiHeader> and then pointed to from the individual entries or senses in which these labels are used. Domain labels inside a sense are documented in <encodingDesc> (encoding description).

```
<encodingDesc>
  <classDecl>
    <taxonomy xml:id="domain">
      <category xml:id="domain.mathematical_sciences">
        <catDesc xml:lang="en">Mathematical Sciences</catDesc>
        <catDesc xml:lang="pt">Ciências Matemáticas</catDesc>
        <catDesc xml:lang="es">Ciencias Matemáticas</catDesc>
        <catDesc xml:lang="fr">Sciences mathématiques</catDesc>
      <category xml:id="domain.mathematical_sciences.mathematics">
        <catDesc xml:lang="en">Mathematics</catDesc>
        <catDesc xml:lang="pt">Matemática</catDesc>
        <catDesc xml:lang="es">Matemáticas</catDesc>
        <catDesc xml:lang="fr">Mathématiques</catDesc>
      <category xml:id="domain.mathematical_sciences.mathematics.arithmetic">
        <catDesc xml:lang="en">Arithmetic</catDesc>
        <catDesc xml:lang="pt">Aritmética</catDesc>
        <catDesc xml:lang="es">Aritmética</catDesc>
        <catDesc xml:lang="fr">Arithmétique</catDesc>
        <[...]>
      </category>
    </category>
  </category>
</taxonomy>
</classDecl>
</encodingDesc>
```

This hierarchical organisation constitutes the foundation of the domain ontology.

Organising knowledge

Morais	METALABEL	Encyclopedia of Diderot and d'Alembert	Dewey Decimal Classification (DDC)	EuroVoc	BabelNet
Arithm. Arithmetico	arithmetics	Reason Philosophy Science of Nature Mathematics Pure Arithmetics	500 Natural sciences & mathematics 510 Mathematics 513 Arithmetic	BT1 pure mathematics BT2 mathematics BT3 natural sciences	IS A pure mathematics area of mathematics
Geometr. Geometrico	geometry	Reason Philosophy Science of Nature Mathematics Pure Geometry	500 Natural sciences & mathematics 510 Mathematics 516 Geometry	BT1 pure mathematics BT2 mathematics BT3 natural sciences	IS A pure mathematics area of mathematics mathematics
Mathem. Mathematico	mathematics	Reason Philosophy Science of Nature Mathematics	500 Natural sciences & mathematics 510 Mathematics	BT1 natural sciences NT1 applied mathematics NT1 pure mathematics	IS A Science major Universal language Academic discipline formal science

Comparison of Morais domain labels and existing classification systems

Organising knowledge

ARITHMETICA, f. f. arte de calcular por algarismos.

GEOMETRIA, f. f. parte da Mathematica que ensina a conhecer a grandeza, razões, e proporções das grandezas continuas, ou sejam linhas, ou figuras, ou sólidos, ou superficies.

MATHEMATICA, f. f. a sciencia, que ensina a conhecer as grandezas de toda sorte, suas razões, relações, e proporções: *Mathematica mista* (oppõe-se ás puras) a que ensina a applicar os principios de calculo, e geometria aos corpos.

ACUTANGULO, adj. Geometr. que tem tres
angulos agudos v. g. *triangulo*.—

Lexicographic article of ACUTANGULO (Morais Silva, 1789, vol. I, p. 24)

DIAMETRO, f. m. a linha recta que tirada
de hum ponto do circulo a outro passa polo seu
ponto central. *P. Pereira* 2. f. 21. usa deste ter-
mo significando a recta em contraposição da li-
nha curva.

Lexicographic article of DIAMETRO (Morais Silva, 1789, vol. I, p. 435)

Organising knowledge

SO'LIDO , adj. que não he fluido ; o corpo cujas partes tem firme união , e não se desfunem de si mesmas ; v. g. o pão , pedra , os metaes , &c. § Não fragil , que resiste ao embate , ou força sem se quebrar v. g. „ *solido edificio , ponte sólida. Ulissea.* § f. Real , effectivo , duravel , que tem força , he bem fundado v. g. „ *doutrina— ; amizade— ; razões— ; devoção—* § *Solido* , em Mathem ; se diz substantivamente , o corpo que tem as 3 dimensões de largura , altura , e longer ; oppõe-se a linha , e superficie. § *Número solido* , v. cubico. § *Em solido* v. *sólidum*. *F. Mendes c. 151.*

Lexicographic article of SO'LIDO (Morais Silva, 1789, vol. I, pp. 414–415)

Organising knowledge

- The Morais dictionary includes labels associated to senses, usually through **abbreviations** typeset in italics.
- Some definitions also include usage information, which is given in a more verbose form.
- The analysis of the lists of abbreviations allowed to put forward a typology of usage information in the Morais dictionary.

EXPLICAÇÃO

D A S

ABBREVIATURAS USADAS NESTE DICCIONARIO.

[illegible]

Organising knowledge

Typology of usage labels – TEI Lex-0 types of <usg>

- **Diatopic information (geographic)**: usage restricted to a regional variety (e.g. *Asiat.*)
- **Diachronic information (temporal)**: usage associated with a period in history (e.g. *Ant.*)
- **Diaintegrative information (hint)**: indicating a loanword (e.g. *Lat.*)
- **Diafrequentative information (frequency)**: indicating frequency of occurrence (e.g. *P. us.*)

Organising knowledge

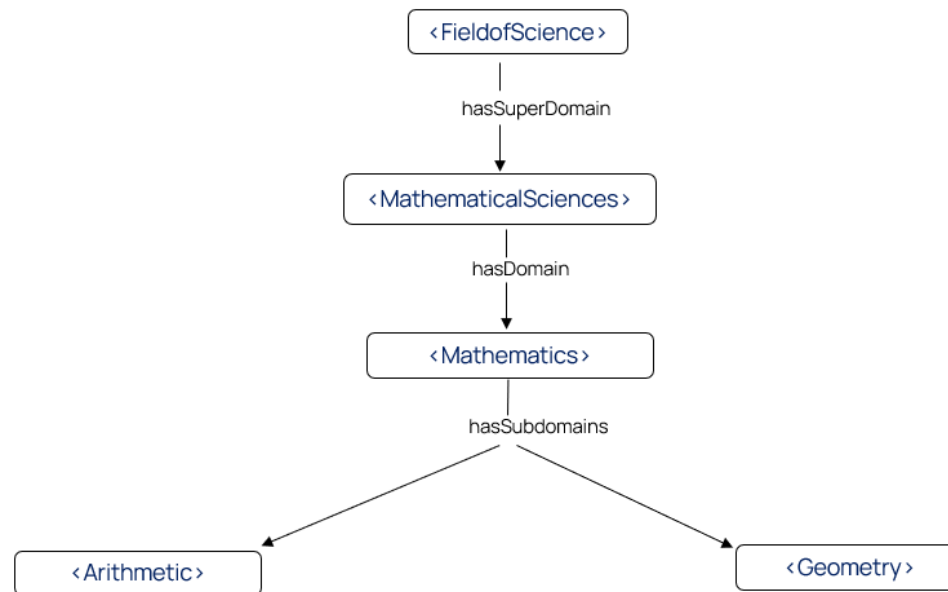
META'STASE , ou *Metastasis* , f. f. *Med.* degeneração de huma doença em outra , especie de Crife. § *na Rhet.* figura pela qual o Orador attribue alguma coiza a outrem , desonerando-fe della.

The different senses of the lexicographic article are constrained to individual subject fields (medicine and rhetoric).

METASTASIS , n. f. *Med.* degeneration of a disease into another , a kind of Crisis. § in *Rhet.* figure by which the Speaker attributes something to another , disowning it.

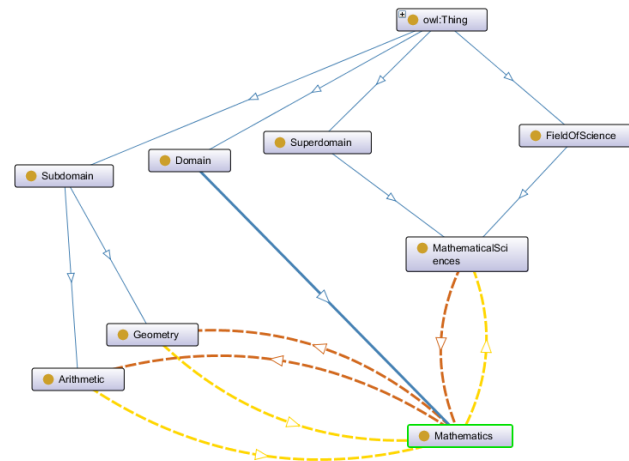
[Freely translated into English]

Organising knowledge



Domain labels within the <MathematicalSciences> superdomain
showing <Mathematics> as a domain
and identifying its subdomains, <Arithmetic> and <Geometry>

Organising knowledge



Representation of the (current) knowledge organisation of MATHEMATICS,
a branch of MATHEMATICAL SCIENCES, in Protégé (OntoGraf)

Organising knowledge

```
<!-- http://www.semanticweb.org/OntoDomLab-Math#MathematicalSciences -->

<owl:Class rdf:about="http://www.semanticweb.org/OntoDomLab-Math#MathematicalSciences">
  <rdfs:subClassOf>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <rdf:Description rdf:about="http://www.semanticweb.org/OntoDomLab-Math#FieldOfScience"/>
        <rdf:Description rdf:about="http://www.semanticweb.org/OntoDomLab-Math#Superdomain"/>
        <owl:Restriction>
          <owl:onProperty rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#hasBranch"/>
          <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#Mathematics"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </rdfs:subClassOf>
</owl:Class>

<!-- http://www.semanticweb.org/OntoDomLab-Math#Mathematics -->

<owl:Class rdf:about="http://www.semanticweb.org/OntoDomLab-Math#Mathematics">
  <rdfs:subClassOf>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <rdf:Description rdf:about="http://www.semanticweb.org/OntoDomLab-Math#Domain"/>
        <owl:Restriction>
          <owl:onProperty rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#branchOf"/>
          <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#MathematicalSciences"/>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#hasBranch"/>
      <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#Arithmetic"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#hasBranch"/>
      <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/OntoDomLab-Math#Geometry"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

The OWL file illustrating the formal definitions of <MathematicalSciences> and <Mathematics>

Organising knowledge

MathematicalSciences **is_a** FieldOfScience
MathematicalSciences **is_a** Superdomain
MathematicalSciences **hasBranch** Mathematics
Mathematics **is_a** Domain
Mathematics **hasBranch** Geometry
Mathematics **hasBranch** Arithmetic
Geometry **is_a** Subdomain
Arithmetic **is_a** Subdomain

Description: Mathematics

Equivalent To +

SubClass Of +

- Domain
and (branchOf some MathematicalSciences)
- hasBranch some Arithmetic
- hasBranch some Geometry
- Domain

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

Target for Key +

Disjoint With +

Disjoint Union Of +

Description of the owl:ObjectProperties defining <Mathematics> in Protégé

Organising knowledge

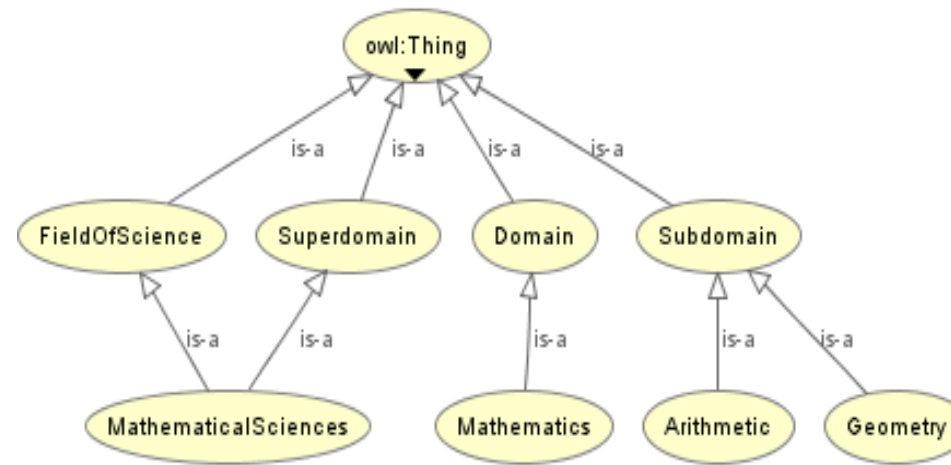


Illustration of the relationship subsumption inferred by the reasoner

Organising knowledge

An initial version of the MorDigital Domain Classification (CDMOR) was recently published as LOD. CDMOR covers the fields of knowledge used, in the form of domain labels, in the lexicographical articles.

VOCABS TOSSIO

Vocabularies About Help | em português

MorDigital Domain Classification

Content language English Search

Alphabetical Hierarchy

- agriculture
- anatomy
- architecture
- artillery
- astrology
- astronomy
- botany
- chemistry
- falconry
- fortification
- geography
- grammar
- heraldry
- horse training
- jurisprudence
- law
- logic
- mathematical sciences
- medical and health sciences
- military
- music
- optics
- orthography
- painting
- perspective
- philosophy
- physics
- rhetoric
- sculpture
- seamanship
- theology

Vocabulary information

TITLE	MorDigital Domain Classification
SUBJECT	Disciplines
DESCRIPTION	<p>The MorDigital Domain Classification (CDMOR) covers the fields of knowledge used, in the form of domain labels, in the lexicographic articles of "Dicionário da Língua Portuguesa" by António de Morais Silva (1789; 1813; 1823). The occurrence of these often abbreviated labels restricts the scope of usage of the underlying lexical units, identifying them as belonging to the lexicon of specialised fields. Thus, for example, the label "t. Anat." in the article "acelerador" (accelerator) (Morais Silva, 1789, p. 14) indicates that this is a term from anatomy, occurring frequently in texts from this specialised field.</p> <p>CDMOR is modelled based on Simple Knowledge Organization System (SKOS), the W3C recommendation for classifications and other knowledge organization systems. The spelling of the preferred terms assigned to each domain are modernised to facilitate search and accessibility. The original spelling of the domains and the respective labels are preserved, appearing in CDMOR as alternative terms and usage labels, with respective indication of their sources. Encyclopaedic information on each domain is presented through scope notes extracted from DBPedia. Finally, definitions of the relevant articles for each domain are presented, covering all three editions of the dictionary. For example, in the Military domain we have the full definitions of the adjective "militar" (military) and the noun "milícia" (militia) (which includes military science as a domain of knowledge).</p> <p>This resource is developed in the context of the project MorDigital – Digitisation of Dicionário da Língua Portuguesa by António de Morais Silva – PTDC/LLT-LIN/6841/2020.</p>
PUBLISHER	https://mordigital.fcsh.unl.pt/
LANGUAGE	English Portuguese
VERSION	0.5

Concluding remarks

- This project will contribute towards a more significant presence of lexicographic digital content in Portuguese through open tools and standards.
- The linking mechanisms of the resulting structured dictionary to other resources will constitute a prototype that can be replicated in other works, namely in the Portuguese-speaking world.
- We also propose combining semasiological and onomasiological approaches in our treatment of the different editions of Moraes. For this, we foresee the inclusion of ontologies (e.g. diasystematic marking, namely domain labels, registers and part of speech categories).

Thank you!

... questions?

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

Ana Salgado: anasalgado@fcsh.unl.pt