

Applying terminological methods to lexicographic work: terms and their domains

Ana Salgado^{1,2}, Rute Costa¹, Toma Tasovac³

¹ NOVA CLUNL, FCSH, Universidade NOVA de Lisboa, Portugal

² Academia das Ciências de Lisboa, Portugal

³ Belgrade Center for Digital Humanities, Serbia

Motivation

- to improve the lexicographic work carried out at the Academia das Ciências de Lisboa (ACL)
- to improve metalanguage and lexicographic description in general
- to guarantee a consistent representation of lexicographic data



Terminological Methods in Lexicography:
Conceptualising, Organising and Encoding Terms in
General Language Dictionaries

Ana Maria de Castro Faria Salgado

PhD in Translation and Terminology
Specialisation in Terminology

December 2021

<https://run.unl.pt/handle/10362/137023>



Objectives

- to study the inclusion and treatment of terms in academy dictionaries, discussing certain decisions traditionally taken by lexicographers
- to facilitate the organisation, description and consistent modelling of lexicographic components, namely domain labels
- to help lexicographers with the drafting of definitions
- to develop a mixed methodology for the treatment of terms
- to show how the currently recommended TEI practice of representing domain labels as flat values is not robust enough to deal with more complex domain structures
- to contribute to the urgent issue of solving problems that affect the sharing, alignment and linking of lexicographic data



Terminological Methods in Lexicography:
Conceptualising, Organising and Encoding Terms in
General Language Dictionaries

Ana Maria de Castro Faria Salgado

PhD in Translation and Terminology
Specialisation in Terminology

December 2021

<https://run.unl.pt/handle/10362/137023>



Outline

1. Framework issues
2. *Dicionário da Língua Portuguesa* (ACL) as a case-study
3. Methodological proposal for the treatment of terms
4. Improving definitions
5. Concluding remarks
6. Future work and research



Framework issues

- Lexicography and terminology are considered autonomous disciplines in the broad digital humanities field
- The digital revolution has challenged adopting new methods concerning the traditional research methodology on lexicography
- Lexicography has a twofold nature: a practical and theoretical components
- Terminology has a double dimension (linguistic and conceptual)
- Our approach assumes that terminology and lexicography can be complementary



Framework issues

CONCEPTUAL DIMENSION

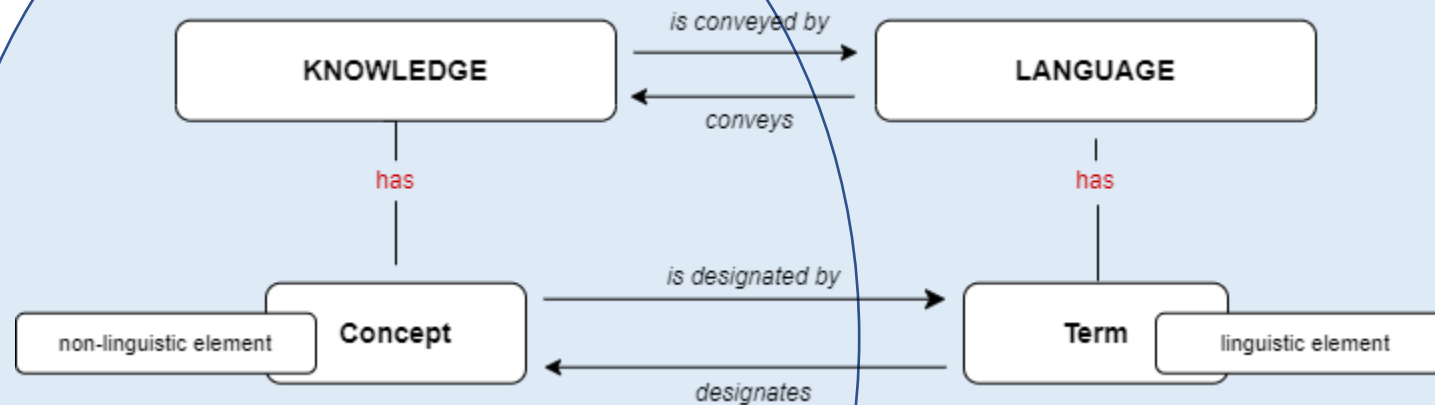
Concept

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

LINGUISTIC DIMENSION

Term

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



*The Relationship of Concept and Term mirroring the double dimension of terminology
(Salgado, 2021, adapted from Costa, 2021)*

ONOMASIOLOGICAL APPROACH

SEMASIOLOGICAL APPROACH

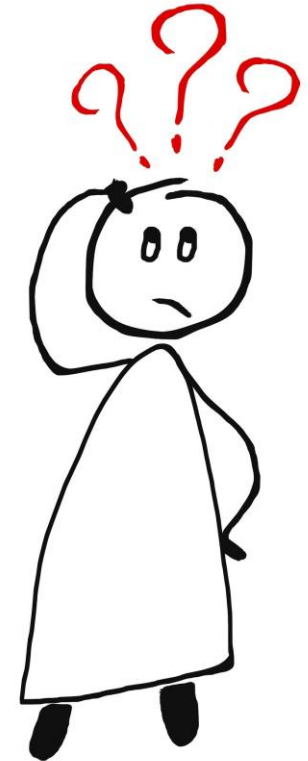


Framework issues

Defining terms? Or Defining concepts?

- identifying concepts and concept relations;
- analysing and modelling concept systems based on identified concepts and concept relations;
- establishing representations of concept systems through concept diagrams;
- **defining concepts;**
- attributing designations (predominantly terms) to each concept in one or more languages;
- recording and presenting terminological data

ISO 704:2009, p. v



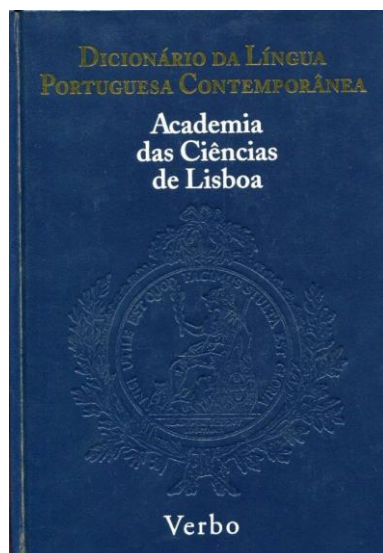
The lexicographer must propose a linguistic description of the concept designated by a term.



Dicionário da Língua Portuguesa (ACL) as a case-study

DLPC (2001): *Dicionário da Língua Portuguesa Contemporânea*, ACL, coord. João Malaca Casteleiro, 2 vols.

69,426 entries
167,556 senses

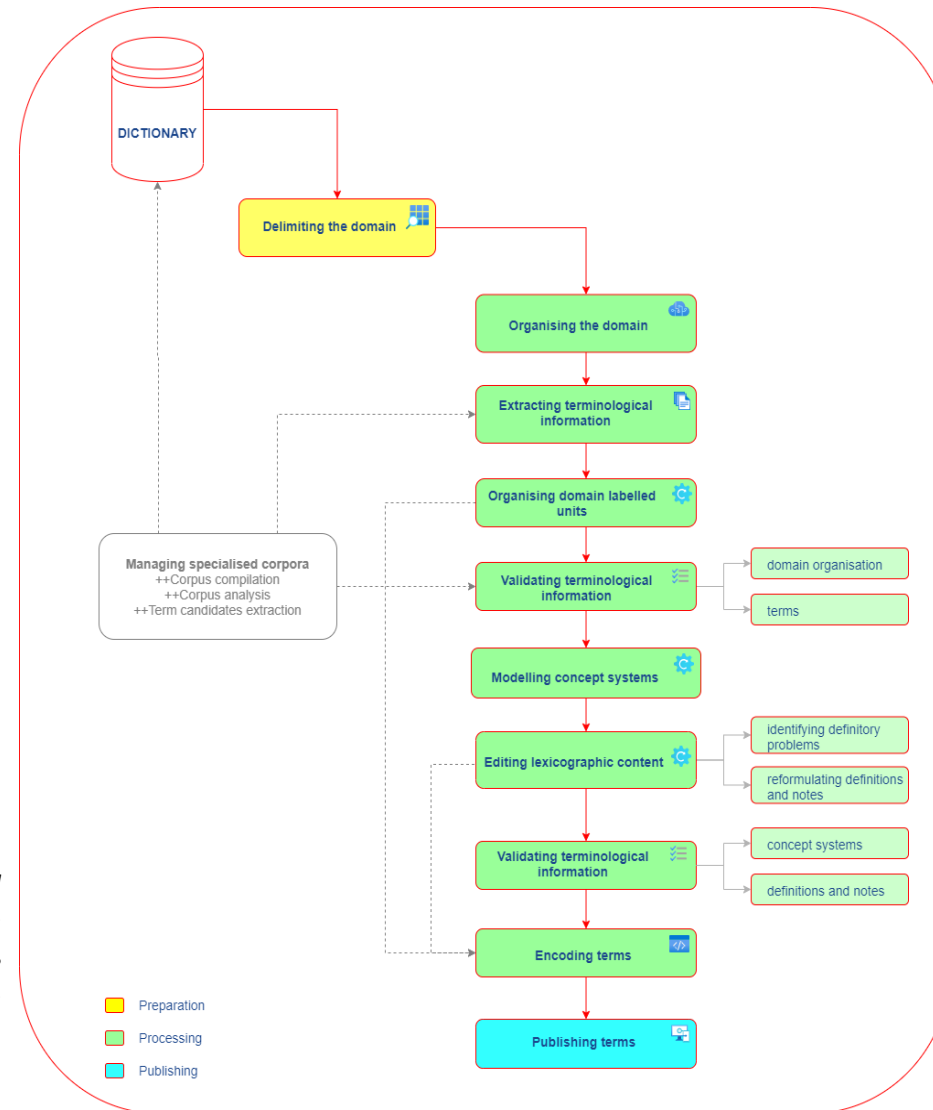


DLP (2022): *Dicionário da Língua Portuguesa*, ACL, coord. Ana Salgado, digital version
100,000 entries...



Methodological proposal for the treatment of terms

Applying terminological methods when treating terms in general language dictionaries (Salgado, 2021)

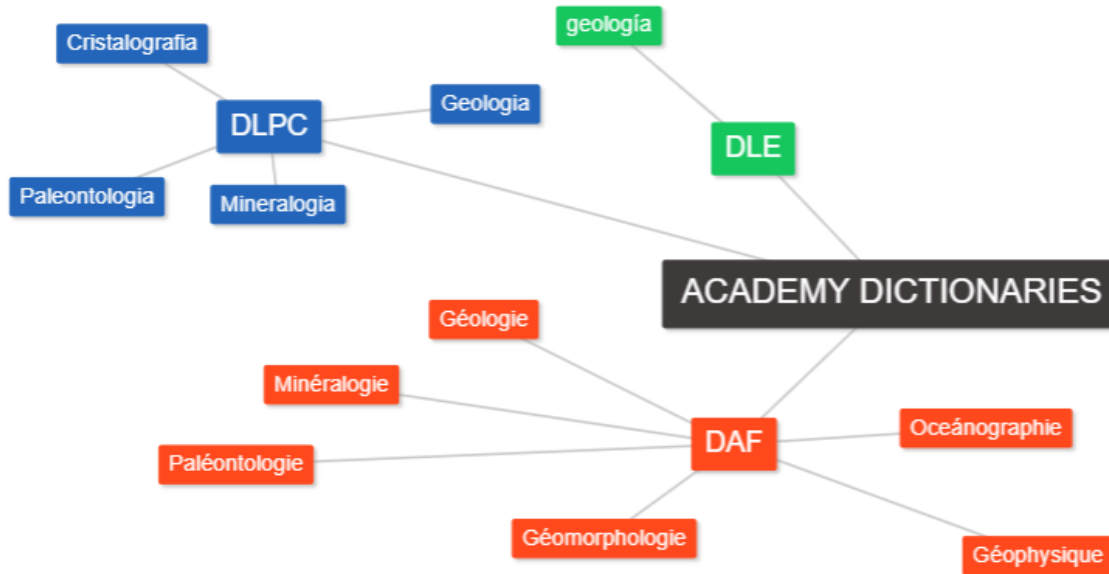


ISO 704 (2009), Terminology work – Principles and methods

ISO 1087 (2019), Terminology Work - Vocabulary - Part 1: Theory and Application

Methodological proposal for the treatment of terms

GEOLOGY



Domain labels in the three academy dictionaries (ACL, AF, RAE)

Rocks Chronostratigraphic Units	Time Geochronologic Units
Eonothem (<i>Eonotema</i>)	Eon (<i>Eon</i>)
Erathem (<i>Eratema</i>)	Era (<i>Era</i>)
System (<i>Sistema</i>)*	Period (<i>Período</i>)
Series (<i>Série</i>)*	Epoch (<i>Época</i>)
Stage (<i>Andar</i>)**	Age (<i>Idade</i>)
Substage	Subage
Subandar)/Chronozone (<i>Cronozona</i>)	(<i>Subidade</i>)/Chron (<i>Crono</i>)

* If additional categories are needed, the prefixes sub- and super- can be used for this purpose.

** When deemed appropriate, it is possible to group adjacent stages using the concept of superstage.

Conventional hierarchy of the chronostratigraphic/geochronologic units



Methodological proposal for the treatment of terms

METALABEL	DLPC	DLE	DAF
crystallography	Cristalog. Cristalografia	—	—
geology	Geol. Geologia	Geol. geología	géol. Géologie
mineralogy	Miner. Mineralogia	—	minér. Minéralogie
palaeontology	Paleont. Paleontologia	—	paléont. Paléontologie
sports	Desp. Desporto	Dep. deportes	—
football	Fut. Futebol	—	—

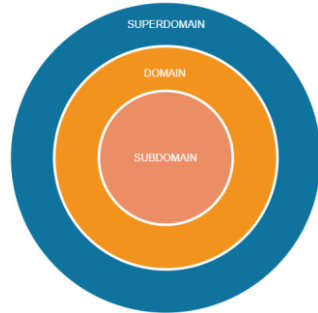
Domains and subdomains under study and their metalabel

DLPC	DLE	DAF	METALABEL	Dewey Decimal Classification (DDC)	Universal Decimal Classification (UDC)	EuroVoc	UNESCO Thesaurus
cristalografia			crystallography	540 Chemistry & allied sciences/548 Crystallography	54 Chemistry. Crystallography. Mineralogy/548/549 Mineralogical sciences. Crystallography. Mineralogy	<i>crystallography returned 0 results</i>	36 Science/2.20 Physical sciences/Cristallography
geologia	geología	géologie	geology	550 Earth sciences/551 Geology, hydrology & meteorology	55 Earth sciences. Geological sciences / 551 General geology. Meteorology. Climatology. Historical geology. Stratigraphy. Paleogeography	36 SCIENCE/3606 natural and applied sciences/NT1 geology	2 Science/2.35 Earth sciences/Geology
mineralogia		minéralogie	mineralogy	540 Chemistry & allied sciences/549 Mineralogy	54 Chemistry. Crystallography. Mineralogy/548/549 Mineralogical sciences. Crystallography. Mineralogy	36 SCIENCE/3606 natural and applied sciences/NT1 geology/NT2 mineralogy	2 Science/2.35 Earth sciences/Mineralogy
paleontologia		paléontologie	palaeontology	560 Paleontology; paleozoology	56 Paleontology	X	2 Science/2.35 Earth sciences/Paleontology

Comparison of academy dictionaries domain labels and classification systems (Salgado, Costa & Tasovac, 2021)

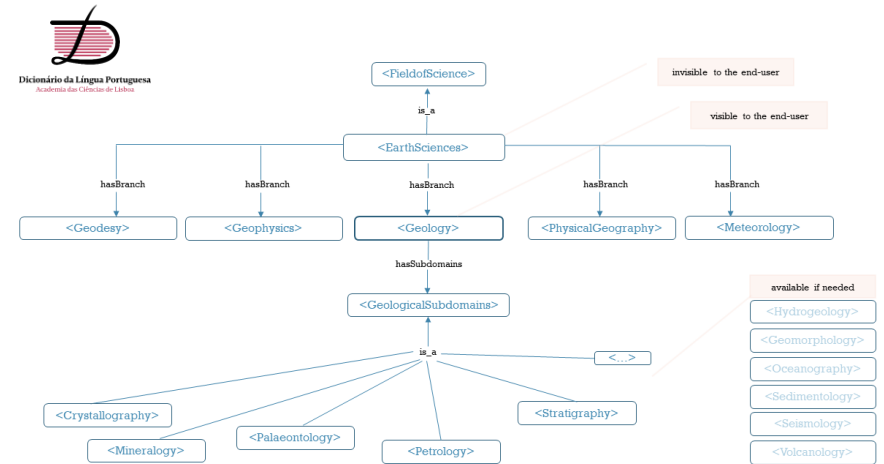


Methodological proposal for the treatment of terms



Domain hierarchy (Salgado, 2021)

- to organise an increasing amount of terminological data
- to provide the lexicographers greater control over specialised content
- to move from a non-hierarchical domain organisation to a hierarchical structure (cf. Salgado, Tasovac, Costa, Asialex, 2022)



Domain labels within the EARTH SCIENCES superdomain showing GEOLOGY as domain and identifying its subdomains

```
<encodingDesc>
  <classDecl>
    <taxonomy xml:id="domain">
      <category xml:id="domain.earth_sciences">
        <catDesc xml:lang="en">Earth Sciences</catDesc>
        <catDesc xml:lang="pt">Ciências da Terra</catDesc>
        <catDesc xml:lang="es">Ciencias de la Tierra</catDesc>
        <catDesc xml:lang="fr">sciences de la Terre </catDesc>
        <category xml:id="domain.earth_sciences.geology">
          <catDesc xml:lang="en">Geology</catDesc>
          <catDesc xml:lang="pt">Geologia</catDesc>
          <catDesc xml:lang="es">Geología</catDesc>
          <catDesc xml:lang="fr">Geologie</catDesc>
          <category xml:id="domain.earth_sciences.geology.mineralogy">
            <catDesc xml:lang="en">Mineralogy</catDesc>
            <catDesc xml:lang="pt">Mineralogia</catDesc>
            <catDesc xml:lang="es">Mineralogía</catDesc>
            <catDesc xml:lang="fr">Mineralogie</catDesc>
          </category>
        </category>
      </category>
    </taxonomy>
  </classDecl>
</encodingDesc>
```

Multilingual hierarchy of EARTH SCIENCES superdomain



Improving definitions

Terminological definition

≠

Lexicographic definition

The **terminological definition** attempts to state a concept designated by a term and to characterise it by relation to other concepts within a concept system.

The **lexicographic definition** seeks to describe the (signified) meaning(s) of a lexical unit.



Improving definitions

The definition must simultaneously designate:

- 1) a **logical operation** at the level of abstraction in which the concept is delimited by the 'combination of characteristics' (ISO 1087, 2019, p. 3) established by differentiation;
- 2) the **production of a string of natural language**, where the term or the 'designation that represents a general concept by linguistic means' (ISO 1087, 2019, p. 7) is the *definiendum*.



Improving definitions

To write a definition in natural language corresponds to a linguistic operation that implies the activation of several levels of analysis:

- the concept to be defined;
- to consider the concept system to which the concept belongs;
- the term that designates the concept to be defined;
- the linguistic expertise in writing the definitional text according to a specific scheme.

Costa (2021, 04-04 July)



Improving definitions

Definitory Problems

[each of the great divisions of geological time, whose boundaries are marked by geological or paleontological changes and which span several periods]

era [ére]. *s. f.* (Do lat. *aera*). [...]

4. **Geol.** Cada uma das grandes divisões do tempo geológico, cujos limites estão marcados por mudanças geológicas ou paleontológicas e que abrange vários períodos. **era primária**, primeira grande divisão do éon fanerozóico que engloba os primeiros 345 milhões de anos. \approx PALEOZÓICO, PRIMÁRIO. **era quaternária**, última grande divisão do éon fanerozóico que se segue à era terciária e que engloba os dois últimos milhões de anos da história da Terra. \approx ANTROPOZÓICO, QUATERNÁRIO. **era secundária**, segunda grande divisão do éon fanerozóico, que se segue à era primária e que engloba cerca de 160 milhões de anos. \approx MESOZÓICO, SECUNDÁRIO. **era terciária**, terceira grande divisão do éon fanerozóico, que se segue à era secundária e que engloba cerca de 65 milhões de anos. \approx CENOZÓICO, TERCIÁRIO.



Fragment of the entry 'era' [era] in the DLPC (ACL)

[time span, in the stratigraphic divisions, which is relative to the formation of a series or set of terrains; period subdivision]

época [épukə]. *s. f.* (Do gr. *ἐποχή*). **1.** Período marcado por factos relevantes, doutrinas políticas, transformações culturais, económicas, sociais, estilo artístico específicos ou personalidades marcantes. \approx ERA. *A época dos Descobrimentos portugueses. A arte da época manuelina.* **2.** Momento em que um acontecimento ocorre, em que um facto [...]

9. *Desp.* Temporada desportiva. *O clube contratou um novo jogador para a próxima época.* **10.** **Geol.** Intervalo de tempo, nas divisões estratigráficas, que é relativo às formações de uma série ou conjunto de terrenos; subdivisão do período. **fazer época**, tornar-se importante, deixar memória de si. **marcar época**, o m. que *fazer época*.

Fragment of the entry 'época' [epoch] in the DLPC (ACL)

Fragment of the entry 'ère' [era] in the DAF (AF)

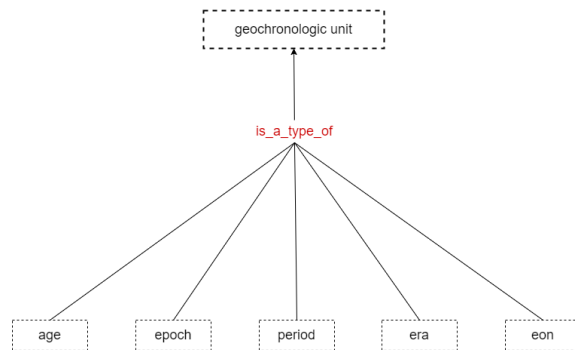


Improving definitions

Modelling concept systems

A **generic relation** exists between two concepts when the intension of the subordinate concept includes the intension of the superordinate concept plus at least one additional delimiting characteristic. (ISO 704, 2009, p. 9)

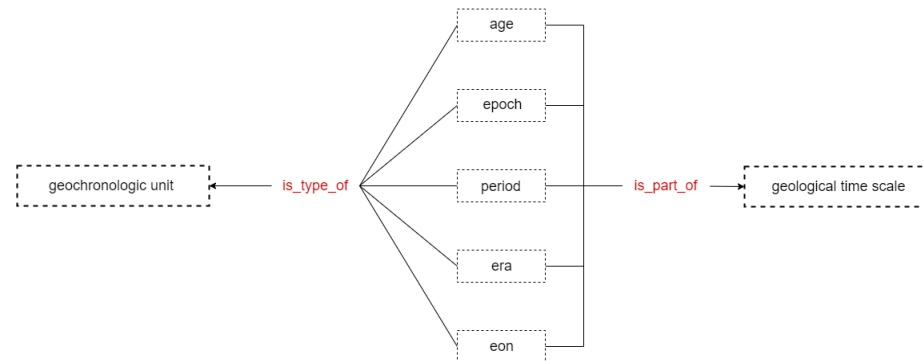
- *X is a type of A.*
- *X, Y, and Z are types of A.*



Representation of a generic relation using the concept of <GeochronologicUnit>

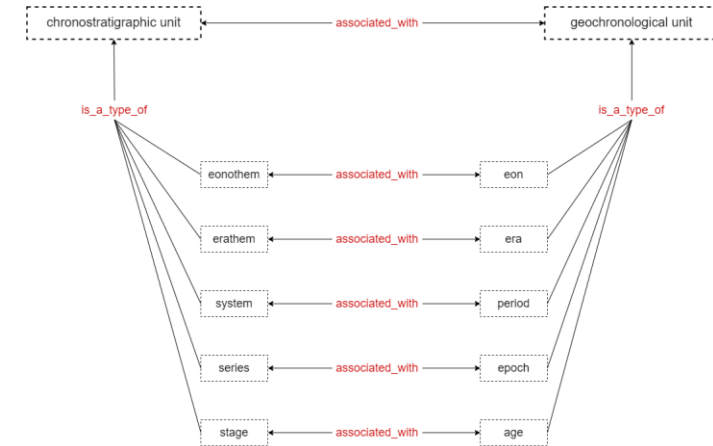
A **partitive relation** is said to exist when the superordinate concept represents a whole, while the subordinate concepts represent parts of that whole. (ISO 704, 2009, p. 13)

- *X is a constituent part of Y.*
- *X, Y, and Z are constituent parts of A.*



Representation of a partitive relation using the concepts of <GeochronologicUnit> and <GeologicalTimeScale>

An **associative relation** exists when a thematic connection can be established between concepts. (ISO 704, 2009, p. 17)



Representation of an associative relationship with the concepts of <ChronostratigraphicUnit> and <GeochronologicUnit> with generic and partitive relations – a mixed concept system

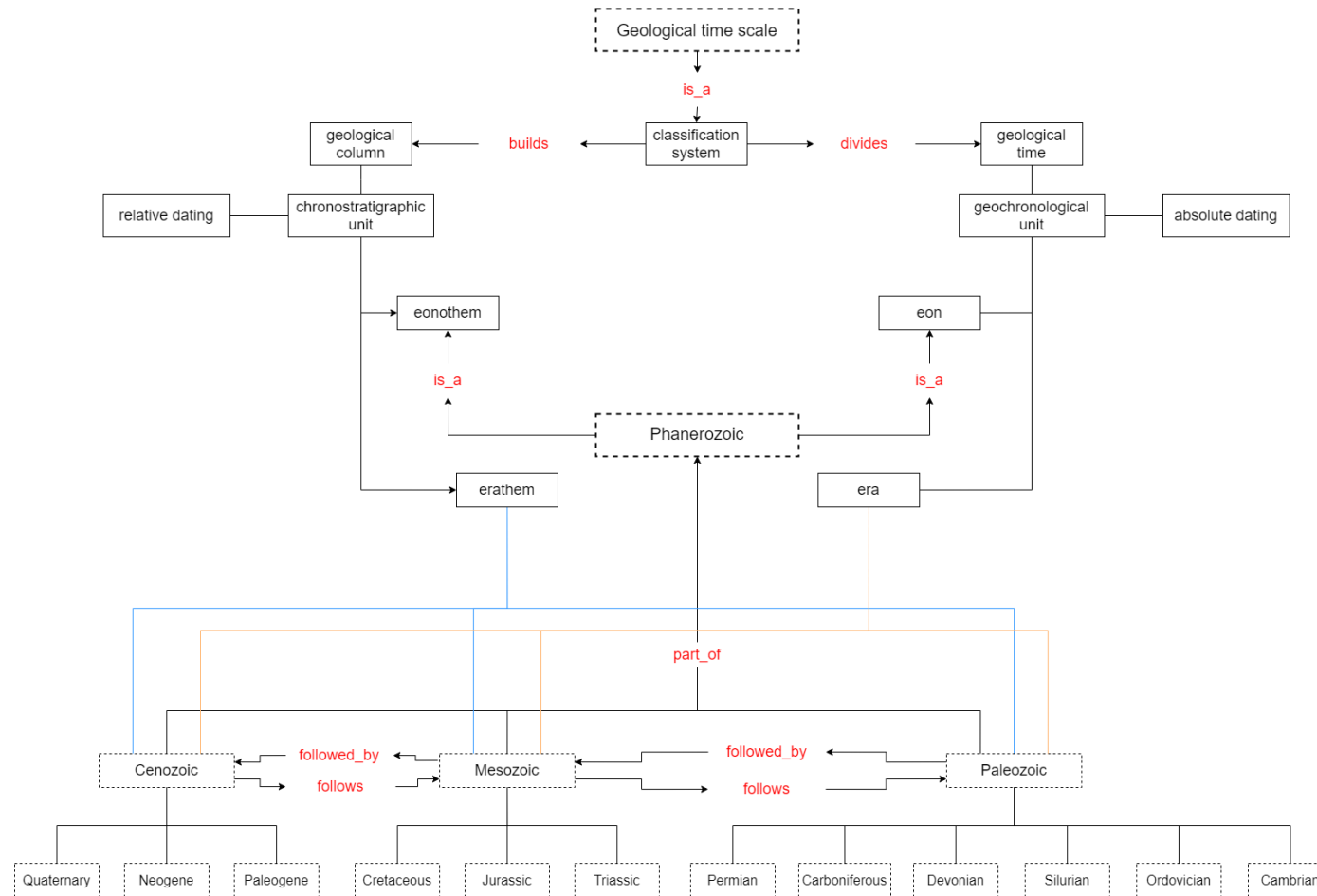
ISO 704 (2009), Terminology work – Principles and methods
UML notation

ISO 1087 (2019), Terminology Work - Vocabulary - Part 1: Theory and Application



Improving definitions

Modelling concept systems

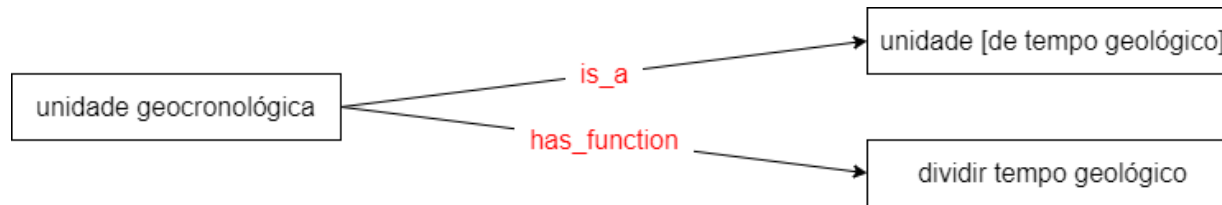


Conceptualising <Phanerozoic> (Salgado, 2021)



Improving definitions

Defining concepts



Representation of the relation the conceptual markers `is_a` and `has_function` established from <GeochronologicUnit>

[unit that divides geological time; geological time subdivision]



unidade geocronológica

GEOLOGIA

unidade que divide o tempo geológico; subdivisão do tempo geológico



Improving definitions

Defining concepts

era [éɾɐ]. *s. f.* (Do lat. *æra*). [...]

4. *Geol.* Cada uma das grandes divisões do tempo geológico, cujos limites estão marcados por mudanças geológicas ou paleontológicas e que abrange vários períodos. **era primária**, primeira grande divisão do éon fanerozóico que engloba os primeiros 345 milhões de anos. ≈ PALEOZÓICO, PRIMÁRIO. **era quaternária**, última grande divisão do éon fanerozóico que se segue à era terciária e que engloba os dois últimos milhões de anos da história da Terra. ≈ ANTROPOZÓICO, QUATERNÁRIO. **era secundária**, segunda grande divisão do éon fanerozóico, que se segue à era primária e que engloba cerca de 160 milhões de anos. ≈ MESOZÓICO, SECUNDÁRIO. **era terciária**, terceira grande divisão do éon fanerozóico, que se segue à era secundária e que engloba cerca de 65 milhões de anos. ≈ CENOZÓICO, TERCÁRIO.

era [ˈɛɾɐ]

nome feminino

1. período de tempo que serve de base a um sistema cronológico e principia com uma data memorável, a partir do qual se contam os anos, em sucessão contínua

① era cristã

a que se conta a partir do nascimento de Jesus Cristo

① era de Cristo

o mesmo que **era cristã**

① era hispânica

a que inicia 38 anos da era cristã

① era muçulmana

a que inicia o calendário islâmico e que corresponde ao ano 622 da era cristã, assinalando Maomé de Meca para Medina (Hégira)

2. início de uma nova ordem de coisas

✎ Entrámos numa nova era de maior justiça e liberdade.

① era atômica era digital; era do fogo; era dos computadores; era espacial

3. qualquer período cronológico demarcado

SINÓNIMOS época ; tempo

① era dos afonsinhos

tempos antiquados, desatualizados

SINÓNIMOS tempo dos afonsinhos

✎ Por alusão irónica aos tempos em que reinaram os primeiros Afonsos de Portugal.

4. **GEOLOGIA** intervalo de tempo geológico (**unidade geocronológica**) durante o qual se formou um cronostratigráfica)

1) Na escala do tempo geológico, a era é hierarquicamente superior ao período e inferior ao éon. 2) A era int

① era primária

1. **GEOLOGIA**, OBSOLETO designação da era paleozoica

SINÓNIMOS primário

2. **GEOLOGIA** o mesmo que paleozoico

① era quaternária

1. **GEOLOGIA**, OBSOLETO denominação da era antropozoica

SINÓNIMOS antropozoico ; quaternário

2. **GEOLOGIA** o mesmo que cenozoico

① era secundária

1. **GEOLOGIA**, OBSOLETO designação da era mesozoica

SINÓNIMOS secundário

2. **GEOLOGIA** o mesmo que mesozoico

① era terciária

1. **GEOLOGIA**, OBSOLETO designação da era cenozoica

SINÓNIMOS terciário

2. **GEOLOGIA** o mesmo que cenozoico

5. numeração que indica um ano ou um certo número de anos

✎ a era de 74 (1974); a era dos anos 90 (1990)

6. **BRASIL** idade; número de anos

✎ Estes minutos são da mesma era.

ETIMOLOGIA Do latim *æra*

4. **GEOLOGIA** intervalo de tempo geológico (**unidade geocronológica**) durante o qual se formou um eretema (**unidade cronostratigráfica**)

1) Na escala do tempo geológico, a era é hierarquicamente superior ao período e inferior ao éon. 2) A era integra vários períodos.

era primária

1. **GEOLOGIA**, OBSOLETO designação da era paleozoica

SINÓNIMOS primário

2. **GEOLOGIA** o mesmo que paleozoico

era quaternária

1. **GEOLOGIA**, OBSOLETO designação da antiga era antropozoica

SINÓNIMOS antropozoico; quaternário

2. **GEOLOGIA** o mesmo que quaternário (atualmente 'sistema/período')

era secundária

1. **GEOLOGIA**, OBSOLETO designação da era mesozoica

SINÓNIMOS secundário

2. **GEOLOGIA** o mesmo que mesozoico

era terciária

1. **GEOLOGIA**, OBSOLETO designação da era cenozoica

SINÓNIMOS terciário

2. **GEOLOGIA** o mesmo que cenozoico



Improving definitions

Definitions accuracy and quality

HEADWORD	DLPC (2001)	DLP (2021)
éon [eon]	Geol. longo período de tempo geológico que abarca duas ou mais eras	intervalo de tempo geológico (<i>unidade geocronológica</i>) durante o qual se formou um eonotema (<i>unidade cronostratigráfica</i>) Notas: 1) Na escala do tempo geológico, o éon é a categoria hierárquica mais elevada. 2) O éon integra várias eras.
era [era]	Geol. cada uma das grandes divisões do tempo geológico, cujos limites estão marcados por mudanças geológicas ou paleontológicas e que abrange vários períodos	intervalo de tempo geológico (<i>unidade geocronológica</i>) durante o qual se formou um eratema (<i>unidade cronostratigráfica</i>) Notas: 1) Na escala do tempo geológico, a era é hierarquicamente superior ao período e inferior ao éon. 2) A era integra vários períodos.
período [period]	—	intervalo de tempo geológico (<i>unidade geocronológica</i>) durante o qual se formou um sistema (<i>unidade cronostratigráfica</i>) Notas: 1) Na escala do tempo geológico, o período é hierarquicamente superior à época e inferior à era. 2) Na escala do tempo geológico, o período integra várias épocas.
época [epoch]	Geol. intervalo de tempo, nas divisões estratigráficas, que é relativo às formações de uma série ou conjunto de terrenos; subdivisão do período	intervalo de tempo geológico (<i>unidade geocronológica</i>) durante o qual se depositou uma série (<i>unidade cronostratigráfica</i>) Notas: 1) Na escala do tempo geológico, uma época é hierarquicamente superior à idade e inferior ao período. 2) Uma época integra várias idades.
idade [age]	—	intervalo de tempo geológico (<i>unidade geocronológica</i>) durante o qual se formou um andar (<i>unidade cronostratigráfica</i>) Notas: 1) A idade é a unidade básica da hierarquia do tempo geológico. 2) Quando necessário, a idade pode ser dividida em unidades geocronológicas de categoria inferior designadas por <i>crono</i> .

Comparison of definitions ‘éon’, ‘era’, ‘período’, ‘época’, ‘idade’ in DLPC (2001) and DLP (2021)

HEADWORD	DLPC (2001)	DLP (2021)
eonotema [eonothem]	—	conjunto de rochas (<i>unidade cronostratigráfica</i>) formadas durante um éon (<i>unidade geocronológica</i>) Nota: Na escala cronostratigráfica, o eonotema é a categoria hierárquica mais elevada.
eratema [erathem]	—	conjunto de rochas (<i>unidade cronostratigráfica</i>) formadas durante uma era geológica (<i>unidade geocronológica</i>) Nota: Na escala cronostratigráfica, o eratema é hierarquicamente superior ao sistema e inferior ao eonotema.
sistema [system]	Geol. período geológico que se caracteriza pela fauna, flora e mutações próprias	conjunto de rochas (<i>unidade cronostratigráfica</i>) formadas durante um período geológico (<i>unidade geocronológica</i>) Nota: Na escala cronostratigráfica, o sistema é hierarquicamente superior à série e inferior ao eratema.
série [serie]	—	conjunto de rochas (<i>unidade cronostratigráfica</i>) formadas durante uma época geológica (<i>unidade geocronológica</i>) Nota: Na escala cronostratigráfica, a série é hierarquicamente superior ao andar e inferior ao sistema.
andar [stage]	Geol. conjunto dos terrenos ou das camadas geológicas correspondentes a uma idade. O andar é definido pelos seus fósseis característicos.	conjunto de rochas (<i>unidade cronostratigráfica</i>) formadas durante uma idade geológica (<i>unidade geocronológica</i>) Notas: 1) Embora o conceito de estratótipo se possa, em princípio, aplicar a todas as unidades estratigráficas, considera-se particularmente importante em relação ao andar, uma vez que corresponde ao conjunto das características descritivas que permitem individualizar cada andar, e a sua base, como formação geológica padrão no registo estratigráfico, equivalente, no tempo geológico, a uma idade. 2) Na escala cronostratigráfica, o andar é a unidade básica da hierarquia. 3) Quando necessário, o andar pode ser subdividido em unidades cronostratigráficas de categoria inferior designadas por <i>subandar</i> e <i>cronozona</i> .

Comparison of definitions ‘eonothem’, ‘erathem’, ‘system’, ‘series’, ‘stage’ in the DLPC (2001) and the DLP (2021)



Concluding remarks

- Following a terminological-based approach improve the quality of the lexicographic product, combining conceptual and linguistic dimensions
- A well-organised domain hierarchy will help lexicographers and end-users better understand the relations between concepts
- To fix the boundaries of the concept implies finding the distinctive characteristics that differentiate them within a concept system
- The conceptual identifiers and linguistic markers may help lexicographers draft definitions
- The solutions presented for the Portuguese academy dictionary (DLP) can be replicated



Future work and research

- To publish online the *Dicionário da Língua Portuguesa*
- To continue to study Academy dictionaries in more detail
- To test the proposed methodology on terms from other fields
- To solve the problematic issues found (domains with multiple labels, unlabelled equivalent headwords, outdated definitions, etc.)
- To establish an agreement between academies/institutions to systematise and optimise a new type of lexicography that can better represent the entire European lexicographic heritage
- We will continue to invest in approaches that combine terminological and lexicographic assumptions, and even other disciplines, placing best practice standards at the core of our research



Obrigada.

Danke schön.

Thank you.

Ana Salgado: anasalgado@fcsh.unl.pt

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

Toma Tasovac: ttasovac@humanistika.org

This talk is supported by the Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia, the MORDigital and Elexis projects.

