Usage Labels in an Eighteenth-century Portuguese Dictionary:
the Case of Morais (1789)

Ana SALGADO, Rute COSTA, Toma TASOVAC, Laurent ROMARY, **Anas Fahad KHAN**,
Bruno ALMEIDA, Margarida RAMOS, Sara CARVALHO, Raquel SILVA, Mohamed KHEMAKHEM

MorDigital

CLUNL
CENTRO DE LINGUÍSTICA DA UNIVERSIDADE NOVA DE LISBOA

NOVAFCSH
FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

fct
Fundação
para a Ciência
e a Tecnologia

Outline

- Historical Context

- Morais, the Lexicographer

- *Diccionario da Lingua Portugueza* by Morais

- MORDigital Project

- Usage Labels

  Methodological Approach

- Conclusion and Future Work

Historical Context

The 18th century – the era known as the **Enlightenment** – was a crucial epoch in world history, one that (especially in Europe) is renowned for its atmosphere of intellectual fervour and for its transformative ideas.

A period that profoundly influenced various fields, including linguistics and lexicography.

Lexicographers of the 18th century found themselves obliged to document languages and provide materials to guide linguistic choices. Their work was inseparable from the process of the formation of national identities that was taking place at the time.

Initially published in 1789, Morais's dictionary can only be truly understood when viewed within the broader context established by earlier lexicographers. Among these predecessors, Rafael Bluteau (1638–1734) holds a prominent place, as well as a number of French lexicographers, who were active earlier during the 17th century.

Understanding this lineage becomes indispensable for comprehending the development of usage labeling in the Morais dictionary.

Morais, the Lexicographer

- ANTÓNIO DE MORAIS SILVA (1757?–1824) was a Brazilian lexicographer born in Rio de Janeiro.

- Graduated in Civil and Canon Law from the University of Coimbra in Portugal. He did not embark on a legal career, as he was persecuted by the Inquisition, which, having accused him of heresy, ordered his arrest.

- Fled to England and devoted himself to studying languages; it was here that he planned the structure of his dictionary.

- Morais moved to Brazil in 1794, where he entered the judiciary and held the position of judge in the Bahia Court of Appeal (from which he soon resigned).

- He subsequently moved to Muribeca, in Pernambuco, where he lived until his death in April 11, 1824.
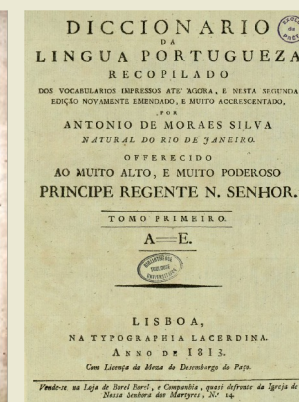
Diccionario da Lingua Portugueza by De Morais Silva

- Referred to by one of the surnames of its author (Morais), it was the first Portuguese monolingual dictionary, and played a key role in normalising the Portuguese language; it was the model for subsequent Portuguese dictionaries.

- It was influenced by other modern language dictionaries published in Europe during the 16th and 17th centuries.

- In its 1st ed., authorship is attributed to Rafael Bluteau, whose Portuguese-Latin Vocabulary (10 vols., 1712–1728) was the basis for the Morais dictionary.

- Morais directly oversaw the 2nd (1813) and 3rd (1823) editions in which the dictionary was greatly overhauled.
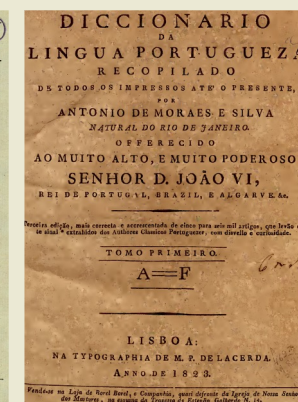
Frontispiece of Morais (1789, 1813, 1823)

Two volumes
A to K, 752 pp.
L to Z, 541 pp.

Two volumes
A to E, 889 pp.
F to Z, 886 pp.

Two volumes
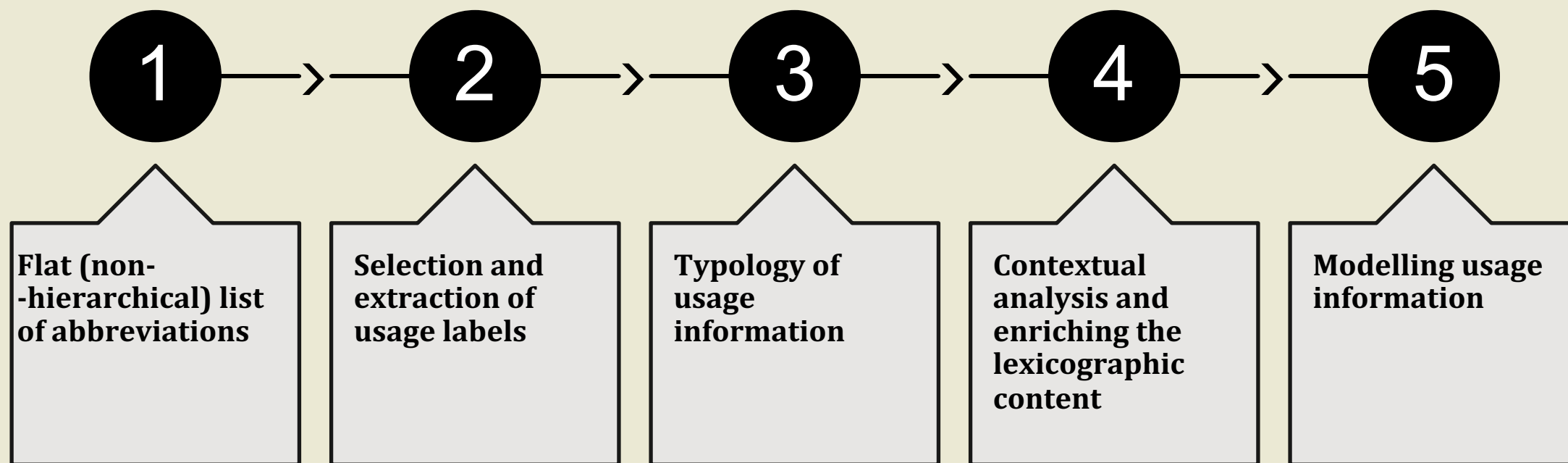A to F, 962 pp.
G to Z, 872 pp.

MORDigital Project

- The current work takes place within the context of **MORDigital** – Digitalização do Diccionario da Lingua Portugueza de António de Morais Silva [PTDC/LLT-LIN/6841/2020] is a project financed by the Portuguese National Funding agency through the FCT – Fundação para a Ciência e Tecnologia.

- Although it is a Portuguese national project, it also includes collaborators from all over Europe.

- The main aim of the project is to convert three editions (1789; 1813; 1823) of Morais into a structured lexical resource in both TEI-XML (based on the ISO LMF standard) and in RDF (based on the OntoLex-Lemon model and its recent extensions).

- These editions will also be made available via an online interface on the website (at the moment only PDFs are available).

- As part of this work we want to analyse usage labels and make the information contained in them more directly accessible to scholars.

Usage Labels

- Refers to the use of conventional labels in dictionaries to categorise and elucidate word meanings and usages. They are integral to the microstructure of dictionary articles, guiding users in their linguistic choices of lexical units. Indicate conventional usage (restrictions) in certain senses and instances where language usage departs from the norm (deviations) along various dimensions of linguistic variation.

- Evolved from earlier classification systems, initially serving as 'jugements de valeur' (Rey, 1990, p. 19), where their primary role was to assess the value or appropriateness of a lexical unit.

- Our analysis of usage labels in Morais is conducted within the framework of contemporary metalexicography. In particular we explore the specific types of usage labels used and their significance within the linguistic landscape of the 18th century.

- This approach enables us to distinguish between
  - **usage labelling**, which deals with established labels
  - **diasystematic marking**, which explores how language varies across different dimensions (such as geography, context, social factors)

- These labels can help us to understand not only what words mean but how their meanings and usages evolve across the context of its time.

Methodological Approach (MA)

1	2	3	4	5
Flat (non-hierarchical) list of abbreviations	Selection and extraction of usage labels	Typology of usage information	Contextual analysis and enriching the lexicographic content	Modelling usage information

1. MA: Flat (non-hierarchical) list of abbreviations

- The usage labels were retrieved from the list of abbreviations in the front matter of Morais.

- Each table of abbreviations has two distinct columns: one with the abbreviation and the other with the complete designations.

- The usage labels follow specific conventions (abbreviated form).

- The list follows alphabetical ordering without any concern for the relationships that can be established between those types of labels.

EXPLICAÇÃO DAS ABBREVIATURAS USADAS NESTE DICCIONARIO.

Abbreviation	Designation
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.	ufado na India Portug.
Aftrol.	Aftrologico.
Aftron.	Aftronomico.
At.	Verbo ativo.
Aument.	aumentativo.
Botan.	Botanico.
Braf.	do Brasão.
(C. ou (Cap.	Capitulo.
Chim.	Chimico.
Cirurg.	Cirurgico.
Com.	Commum de dois.
Compar.	Comparativo.
Conj.	Conjuncção.
(Ch. (Chul.	t. Chulo.

2. MA: Selection and extraction of usage labels

- We selected and collected all the usage labels in an Excel sheet.

- To facilitate analysis and to focus on interoperability issues, we generate a metalabel, i.e., a tag that identifies the English equivalent of the corresponding label.

Abbreviation	Full form	Metalabel (English)	Abbreviation	Full form	Metalabel (English)
Agric.	Agricultura	Agriculture	Jurid.	Juridico	Law
Anat.	Anatomia, Anatomico	Anatomy	Jurisp.	da Jurisprudencia	Jurisprudence
Ant. ou antiq.	antiquado	old	Lat.	Latino	Latino
Archit.	d'Architectura	Architecture	Log.	Logico	Logic
Arithm.	Arithmetico	Arithmetic	Manej.	Termo do Manejo dos cavallos	Equestrianism
Artelh.	d'Artelharia	Artillery	Mathem.	Mathematico	Mathematics
Asiat.	usado na India Portug.	Asianism	Med.	Medico	Medicine
Astrol.	Astrologico	Astrology	Milit.	Militar	Military
Astron.	Astronómico	Astronomy	Mus.	Musico	Music
Botan.	Botanico	Botany	Naut.	da Nautica	Seamanship
Bras.	do Brasão	Heraldry	Opt.	Optico	Optics
Ch., Chul.	t. Chulo	coarse	Ortogr.	Ortografico	Orthography
Chim.	Chimico	Chemistry	P. us.	Pouco usado	Pouco usado
Cirurg.	Cirugico	Surgery	Persp.	da Perspectiva	Perspective
Escult.	Escultura	Sculpture	Pharmac.	Pharmaceutico	Pharmacy
Fam.	familiar	familiar	Pint.	da Pintura	Painting
Filos.	Filosofico	Philosophy	Poet.	Poetico	Poetic
Fisic.	da Fisica	Physics	Rhet.	Rhetorico	Rhetoric
Fortif.	da Fortificação	Fortification	Theol.	Theologico	Theology
Geogr.	Geografico	Geography	Volat.	Volateria	Falconry
Geometr.	Geometrico	Geometry	Vulg.	Vulgar	Vulgar
Grammat.	da Grammatica	Grammar			

3. MA: Typology of usage information

- There are several typologies of usage information, mostly adapted from Hausmann (1989), which organise it according to several criteria (e.g. time, degree of speciality).

- The list analysis allowed us to develop a typology of usage information in the Morais dictionary: diatechnical, diatextual, diastratic, among many others.

Almeida et al. (2022). Adapted from: Salgado, Costa & Tasovac (2019)

Hausmann (1989) adopted by others	Milroy, J. and Milroy, L. (1990)	Landau (2001)	Jackson (2002)	Atkins and Rundell (2008)	TEI Lex-0 usage type	Criterion	Examples
diachronic	temporal	currency/ temporality	history	time	temporal	TIME	archaic, old
diatopic	geographical	regional/ geographic variation	dialect	region dialect	geographic	PLACE	AmE., dial.
diaintegrative	—	—	—	—	hint	NATIONALITY	Latin, English
diamedial	—	style, functional variety/register	—	—	hint	MEDIUM	spoken
diastratic	—	restricted or taboo sexual scatological usage and slang	status	slang and jargon offensive terms	socioCultural	SOCIO-CULTURAL	slang, vulgar, formal
diaphasic	register	style, functional variety/register	formality	register	socioCultural	FORMALITY	slang, vulgar, formal
diatextual	—	style, functional variety/register	—	style	textType	TEXT TYPE	bibl., poet., admin., journalese
diatechnical	field	technical or specialized terminology	topic or field	domains	domain	SPECIALITY	Med., Biol., Phys.
diafrequential	frequency	—	—	—	frequency	FREQUENCY	rare, occas.
diaevaluative	—	insult style, functional variety/register	effect	attitude	attitude	ATTITUDE	derog., euph.
dianormative	—	status or cultural level	disputed usage	—	normativity	NORMATIVITY	non-standard, incorrect
—	—	—	—	—	meaningType	MEANING	fig. (=figurative), lit. (= literal)

3. MA: Typology of usage information

Diachronic (temporal): the historical development of lexical units and their meanings over time (e.g. Ant.)

Diatopic (geographic): regional or geographical variations (e.g. Asiat.)

Diaintegrative (hint): integration of loanwords from other languages into Portuguese (e.g. Lat.)

Diastratic (socioCultural): restricted to different social or cultural groups (e.g. Vulg.)

Diaphasic (socioCultural): based on different situational contexts (e.g. Fam.)

Diatextual (textType): restricted to different types of discourse (e.g. Poet.)

Diafrequential (frequency): based on the frequency, whether common or rare (e.g. P. us.)

Diaevaluative (attitude): reflecting the speaker's attitude (e.g. t. chulo)

Diatechnical (domain): restricted to subject domains, activities, or occupations, often accompanied by the textual marker *Termo de*… [term of] (e.g. Med.)

4. MA: Contextual analysis and enriching the lexicographic content

CADEIXAS , ſ. m. Beir. bacamarte, livro velho.

TOME (old book)

CUNCA, ſ. f. tigella, ou ſopeira de páo no Minho ,, huma cunca de berças.

BOWL

BAJULO , ſ. m. mariola , homem , que vive de fazer carretos. Vieira. p. uſ.

SCOUNDREL

BATECU, ſ. m. pleb. golpe que ſe dá com o aſſento do corpo , cahindo.

PRATFALL

CHANCA , ſ. f. vulgar , pé grande: (cangoa em Vaſconſo coixa): ,, Shank ,, Inglez o meſmo , ſoa chank.

BIG FOOT

PATACHOCA , ſ. m. vulg. o ſervente da ſacriſtia.

A boy who serves in the sacristy

COÇA , ſ. f. ch. coça de pancadas , tunda.

THRASHING

4. MA: Contextual analysis and enriching the lexicographic content

BEXANO, f. m. famil. gato novo.

PUSSY CAT

ORTO, f. m. Aftron. nafcimento, ou apparição do aftro no horizonte v. g. ,, orta vefpertino, ou matutino.

RISING

VECTOR, adj. *raio vector*, he a recta terminada no centro da Orbita, e no planeta, a qual fe concebe como levando o planeta do centro á fua Orbita. t. Aftronom.

VECTOR

TELESCOPIO, f. m. inftrumento optico de Aftronomia que ferve de obfervar na terra, ou no Ceo os objectos remotos, por meio da reflexão, ou refracção da luz.

TELESCOPE

5. MA: Modelling usage information

ORDENADA , f. f. Mathem. linha recta ti-
rada perpendicularmente do ponto da curva a feu
eixo.

TELESCOPIO , f. m. inftrumento optico de
Aftronomia que ferve de obfervar na terra , ou
no Ceo os objectos remotos , por meio da re-
flexáo , ou refracçáo da luz.

```
<entry xmlns="http://www.tei-c.org/ns/1.0"
xml:id="MORAIS.DLP.1.ORDENADA" type="mainEntry" xml:lang="pt">
    <form type="lemma">
        <orth>ORDENADA</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.ORDENADA.s.1">
        <usg type="domain"
valueDatcat="#domain.mathematics">Mathem.</usg>
        <def>linha recta tirada perpendicularmente do ponto da
curva a feu eixo</def>
    </sense>
    <metamark function="senseDelimiter">.</metamark>
</entry>
```

```
<entry xmlns="http://www.tei-c.org/ns/1.0"
xml:id="MORAIS.DLP.1.TELESCOPIO" type="mainEntry" xml:lang="pt">
    <form type="lemma">
        <orth>TELESCOPIO</orth>
    </form>
    <metamark function="lemmaDelimiter">,</metamark>
    <gramGrp>
        <gram type="pos" norm="NOUN">f.</gram>
        <gram type="gen">m.</gram>
    </gramGrp>
    <sense xml:id="MORAIS.DLP.1.TELESCOPIO.s.1">
        <usg type="domain" valueDatcat="#domain.astrology"
resp="#Salgado"/>
        <def>inftrumento óptico de
            Aftronomia que ferve de obfervar na terra , ou
            no Ceo os objectos remotos, por meio da reflexão
        , ou refracçáo da luz</def>
    </sense>
    <metamark function="senseDelimiter">.</metamark>
</entry>
```
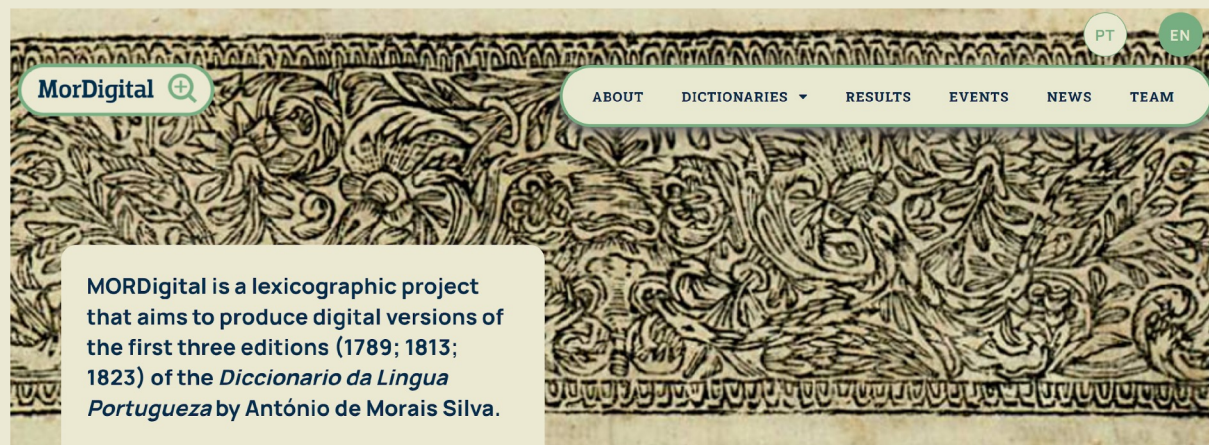
Conclusions

- Our analysis has revealed a rich variation in the distribution of usage labels, encompassing diachronic, diatopic, diaphasic dimensions, among others.

- Usage information is usually represented through abbreviated labels, but sometimes also included within definitions (challenge for machine-learning models).

- These labels are not mere linguistic artifacts but offer broader cultural insights into how language was perceived, used, and shaped during the 18 the century.

- Knowledge organisation is crucial to improve the labeling system in dictionaries.

- The contemporary metalexicographic framework used within the scope of the MorDigital project has facilitated a deeper understanding of the usage labeling system and to organise knowledge.
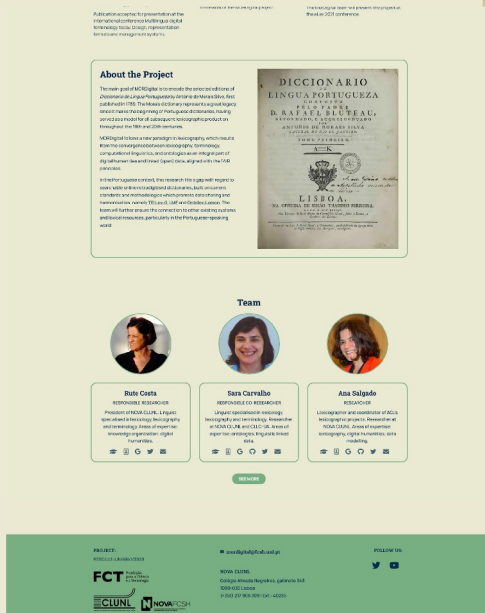
Future Work

- TEI Lex-0 enables the XML encoding of the Morais dictionary in an interoperable way, providing the basis for visualising, querying and mining dictionary data in the MorDigital platform.

- 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.

- We have already worked on the representation of domain labels in a linguistic linked data framework using OntoLex-Lemon. We have produced a set of guidelines for doing this.

  See http://bit.ly/domain_labels_rdf

If you want to know more about our project...

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

Grazie!

Ana Salgado: anasalgado@fcsh.unl.pt

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

Toma Tasovac: ttasovac@humanistika.org

Laurent Romary: laurent.romary@inria.fr

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

Bruno Almeida: brunoalmeida@fcsh.unl.pt

Margarida Ramos: mvramos@fcsh.unl.pt

Sara Carvalho: sara.carvalho@ua.pt

Raquel Silva: raq.silva@fcsh.unl.pt

Mohamed Khemakhem: medkhemakhemfsegs@gmail.com