MorDigital

The teiHeader and standardising dictionary metadata

Ana Salgado, Toma Tasovac, Laurent Romary, Boris Lehečka, Rute Costa, Anas Fahad Khan, Margarida Ramos, Bruno Almeida, Sara Carvalho, Mohamed Khemakhem, Raquel Silva

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

MORDigital project

- 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.
- MORDigital aims to make a historic Portuguese-language dictionary, the Morais dictionary, available as a digital resource.
- This dictionary will also be made available via an online interface on the website (at the moment only PDFs are available).

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

MORDigital project

The *Diccionario da Lingua Portugueza*

- Our aim is to convert Morais dictionary into a structured lexical resource in both TEI-XML (the lexicographic content is structured in TEI Lex-0) and in RDF (based on the OntoLex-Lemon model and its recent extensions) RDF by means of the W3C recommendation for publishing lexicons as Linked Data.

Frontispiece of *Morais* dictionaries (1789, 1813, 1823)

Two volumes
A to K, 752 pp.
L to Z, 541 pp.
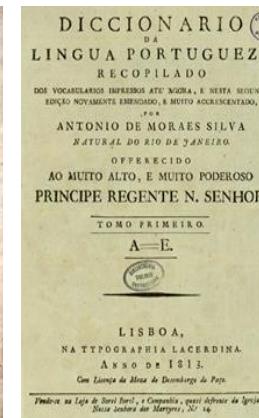
Two volumes
A to E, 889 pp.
F to Z, 886 pp.
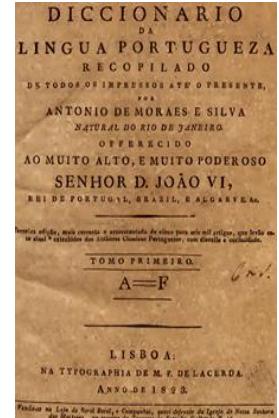
Two volumes
A to K, 952 pp.
L to Z, 872 pp.

3

MORDigital project

```xml
<teiHeader>
    <fileDesc>
        <titleStmt>
            <title type="full" xml:lang="en">Morais Dictionary (1st ed., 1789): digital
                edition</title>
        </titleStmt>
        <editionStmt>
            <edition>XML edition using GroBID-dictionaries</edition>
            <principal>
                <persName>
                    <forename>Rute</forename>
                    <surname>Costa</surname>r </persName>
            </principal>
            <editor role="Researcher">
                <persName>
                    <forename>Sara</forename>
                    <surname>Carvalho</surname>
                </persName>
                <affiliation n="1">NOVA CLUNL, Centro de Linguística da Universidade NOVA de
                    Lisboa</affiliation>
                <affiliation n="2">CLLC, Centro de Línguas, Literaturas e Culturas da Universidade de
                    Aveiro</affiliation>
            </editor>
            <editor role="Researcher">
                <persName xml:id="Salgado">
                    <forename>Ana</forename>
                    <surname>Salgado</surname>
                </persName>
                <affiliation n="1">NOVA CLUNL, Centro de Linguística da Universidade NOVA de
                    Lisboa</affiliation>
                <affiliation n="2">Academia das Ciências de Lisboa, Instituto de Lexicologia e
                    Lexicografia da Língua Portuguesa</affiliation>
            </editor>
```

```xml
<sourceDesc>
    <listBibl type="dictionaries">
        <biblStruct type="book" xml:id="Morais1789">
            <monogr>
                <title level="m" type="main">Diccionario da lingua portugueza composto pelo
                    padre D. Rafael Bluteau, reformado, e accrescentado por Antonio de Moraes
                    Silva, natural do Rio de Janeiro</title>
                <title level="m" type="sub">A - K</title>
                <author>
                    <persName>
                        <forename>António de</forename>
                        <surname>Morais Silva</surname>
                    </persName>
                </author>
                <imprint>
                    <pubPlace>Lisboa</pubPlace>
                    <publisher>Officina de Simão Thaddeo Ferreira</publisher>
                    <pubPlace>Lisboa</pubPlace>
                    <date when="1789">1789</date>
                    <note>Com Licença da Real Meza da Comissão Geral, sobre o Exame, e Censura
                        dos Livros.</note>
                    <!-- This sentence describes the store where the dictionary was for sale. We decided to use note (04.02.2023) -->
                    <note>Vende-ſe na loja de Borel Borel, e Companhia, quaſi defronte da Igreja
                        nova de Noſſa Senhora dos Martyres, na eſquina.</note>
                </imprint>
                <extent>Tomo primeiro</extent>
                <extent>752 pp.</extent>
            </monogr>
        </biblStruct>
</sourceDesc>

<encodingDesc>
    <projectDesc>
        <!-- In progress. -->
        <p>Prepared within MorDigital Project (PTDC/LLT-LIN/6841/2020).</p>
        <p>We already had access to OCR'ed versions of the dictionary editions at the beginning
            of the project. These files needed to be post-corrected. For this, we decided to use
            ABBYY FineReader, version 12.</p>
    </projectDesc>
    <editorialDecl>
        <p>Original spelling and typography is retained.</p>
    </editorialDecl>

        <profileDesc>
            <langUsage>
                <language role="objectLanguage" ident="pt">Portuguese</language>
                <language role="workingLanguage" ident="pt">Portuguese</language>
            </langUsage>
        </profileDesc>
```

MORDigital project

```xml
<!-- Hierarchical domain labels -->
<taxonomy xml:id="domains">
    <category xml:id="domain.agriculture">
        <!-- In MORAIS, Agric. -->
        <catDesc xml:lang="en">Agriculture</catDesc>
        <catDesc xml:lang="pt">Agricultura</catDesc>
    </category>
    <!-- [...] -->
    <category xml:id="domain.astronomy">
        <!-- In MORAIS, Astron. -->
        <catDesc xml:lang="en">Astronomy</catDesc>
        <catDesc xml:lang="pt">Astronomia</catDesc>
    </category>
    <category xml:id="domain.botany">
        <!-- In MORAIS, Bot. -->
        <catDesc xml:lang="en">Botany</catDesc>
        <catDesc xml:lang="pt">Botânica</catDesc>
    </category>
    <!--<usg type="domain" valueDatacat="domain.medical_and_health_sciences.medicine">Med.</usg>-->
    <category xml:id="domain.mathematical_sciences">
        <catDesc xml:lang="en">Mathematical Sciences</catDesc>
        <catDesc xml:lang="pt">Ciências Matemáticas</catDesc>
        <category xml:id="domain.mathematical_sciences.mathematics">
            <catDesc xml:lang="en">Mathematics</catDesc>
            <catDesc xml:lang="pt">Matemática</catDesc>
            <category xml:id="domain.mathematical_sciences.mathematics.arithmetics">
                <catDesc xml:lang="en">Arithmetics</catDesc>
                <catDesc xml:lang="pt">Aritmética</catDesc>
            </category>
            <category xml:id="domain.mathematical_sciences.mathematics.geometry">
                <catDesc xml:lang="en">Geometry</catDesc>
                <catDesc xml:lang="pt">Geometria</catDesc>
            </category>
        </category>
    </category>
```

Salgado et al. (2023). Lexical Resources Summit, 18th April 2023

MORDigital project

```xml
<div type="section" n="5">
    <head>ABREVIATURAS DAS CITAÇÕES DOS LIVROS PORTUGUEZES COM QUE SE AUTHORISA O USO DAS
        PALAVRAS.</head>
    <list rend="simple" xml:lang="pt">
        <item n="0.1.1">
            <abbr type="title" rend="italic">ABeced. Real</abbr>
            <pc>.</pc>
            <expan>Abecedario Real do P. João dos Prazeres</expan>
            <pc>.</pc>
        </item>
        <item xml:id="Morais_0.1.2">
            <abbr type="title" rend="italic" corresp="#Academ.sing">Academ. ſing.</abbr>
            <expan>Academia dos ſingulares de Lisboa</expan>
            <pc>.</pc>
        </item>
        <item n="0.1.3">
            <abbr type="title" rend="italic">Acções Epiſc.</abbr>
            <!-- Lucas de Andrade not Andrada -->
            <expan>Acções Epiſcopaes de Lucas de <choice><sic>Andrada</sic><corr
resp="#Salgado">Andrade</corr></choice></expan>
            <pc>.</pc>
        </item>
```

MORDigital project

HABITAÇÃO , ſ. f. lugar de morada , ou vivenda.

```xml
<entry xmlns="http://www.tei-c.org/ns/1.0" xml:id="MOR1.DLP.ABACO" type="mainEntry" xml:lang="pt">
    <form type="lemma">
        <orth>ABACO</orth>
    </form>
    <metamark function="lemmaDelimiter">,</metamark>
    <gramGrp>
        <gram type="pos" norm="NOUN">ſ.</gram>
        <gram type="gen">m.</gram>
    </gramGrp>
    <sense xml:id="MOR1.DLP.ABACO.s.1">
        <!-- SEE usa-se na Architect. = domain  -->
        <def>Peça superior do capitel da columna , ſerve como de coberta ao ceſto de flores , que nelle se representa</def>
        <metamark function="domainDelimiter">;</metamark>
        <usg type="domain">usa-se na <hi rend="italics">Architect.</hi></usg>
    </sense>
    <metamark function="senseDelimiter">§</metamark>
    <sense xml:id="MOR1.DLP.ABACO.s.2">
        <usg type="domain" rend="italics">t. arithm.</usg>
        <def>a taboada de Pythagoras</def>
    </sense>
    <pc>.</pc>
</entry>
```

MORDigital project

ABACO, f. m. Peça superior do capitel da columna, ferve como de coberta ao cefto de flores, que nelle fe reprefenta; ufa-fe na *Architect.* § t. *arithm.* a taboada de Pythagoras.

```
<entry xmlns="http://www.tei-
c.org/ns/1.0" xml:id="MOR1.DLP.ABACO" type="mainEntry" xml:lang="pt">
    <form type="lemma">
        <orth>ABACO</orth>
    </form>
    <metamark function="lemmaDelimiter">,</metamark>
    <gramGrp>
        <gram type="pos" norm="NOUN">f.</gram>
        <gram type="gen">m.</gram>
    </gramGrp>
    <sense xml:id="MOR1.DLP.ABACO.s.1">
        <!-- SEE usa-se na Architect. = domain  -->
        <def>Peça superior do capitel da columna , ferve como de coberta ao
cefto de flores , que nelle se representa</def>
        <metamark function="domainDelimiter">;</metamark>
        <usg type="domain">usa-se na <hi rend="italics">Architect.</hi></usg>
    </sense>
    <metamark function="senseDelimiter">§</metamark>
    <sense xml:id="MOR1.DLP.ABACO.s.2">
        <usg type="domain" rend="italics">t. arithm.</usg>
        <def>a taboada de Pythagoras</def>
    </sense>
    <pc>.</pc>
</entry>
```

MORDigital project

ASTROLABIO, f. m. inftrumento Aftrono-
mico, de que fe ufa para fe tomarem a altura dos
aftros. f. Paiva Serm. I. 54. ℣. ,, Porque não vos
governará por effe voffo aftrolabio.

```xml
<entry xmlns="http://www.tei-
c.org/ns/1.0" xml:id="MOR1.DLP.ASTROLABIO" type="mainEntry"
    xml:lang="pt">
    <form type="lemma">
        <orth>ASTROLABIO</orth>
    </form>
    <metamark function="lemmaDelimiter">,</metamark>
    <gramGrp>
        <gram type="pos" norm="NOUN">f.</gram>
        <gram type="gen">m.</gram>
    </gramGrp>
    <sense xml:id="MOR1.DLP.ASTROLABIO.s.1">
        <usg type="domain" corresp="#domain.astronomy" resp="#Salgado"/>
        <def>inftrumento Aftronomico, de que fe ufa para fe tomarem a altura
dos aftros</def>
        <pc>.</pc>
        <cit type="example" xml:lang="pt">
            <usg type="socioCultural">f.</usg>
            <bibl type="attestation" source="#PS.">
                <title>Paiva Serm.</title>
                <citedRange unit="volume">I</citedRange>
                <pc>.</pc>
                <citedRange unit="folium">f. 54</citedRange>
            </bibl>
            <lbl>v.</lbl>
            <metamark function="exampleDelimiter">,,</metamark>
            <quote>Porque não vos governará por effe voffo aftrolabio</quote>
            <pc>.</pc>
        </cit>
    </sense>
</entry>
```

… discussion
… questions

Ana Salgado: anasalgado@fcsh.unl.pt

Laurent Romary: laurent.romary@inria.fr

Toma Tasovac: ttasovac@humanistika.org

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

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

Margarida Ramos: mvramos@fcsh.unl.pt

Bruno Almeida: brunoalmeida@fcsh.unl.pt

Sara Carvalho: sara.carvalho@ua.pt

Mohamed Khemakhem: medkhemakhemfsegs@gmail.com

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

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