



NOVA FCSH

FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

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

Ana Maria de Castro Faria Salgado

PhD in Translation and Terminology

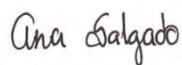
Specialisation in Terminology

December 2021

DECLARAÇÕES

Declaro que esta tese é o resultado da minha investigação pessoal e independente.
O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas
no texto, nas notas e na bibliografia.

A candidata,

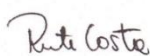


[Ana Salgado]

Lisboa, 14 de dezembro de 2021

Declaro que esta tese se encontra em condições de ser apreciada pelo júri a
designar.

A orientadora,

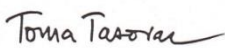


[Rute Costa]

Lisboa, 14 de dezembro de 2021

Declaro que esta tese se encontra em condições de ser apreciada pelo júri a
designar.

O coorientador,



[Toma Tasovac]

Lisboa, 14 de dezembro de 2021

Thesis submitted to fulfil the requirements for obtaining the doctorate degree in

Translation and Terminology

Specialisation in Terminology

Developed under the supervision of

Professor Rute Costa

and

Doctor Toma Tasovac

To Rico, who made me gaze at the stars, and for all he represents to me.
To Pedro and Jó, who trusted me and gave me the chance to follow this journey.
To Jojó, who always reminds me how nice it is to be sweet and spontaneous.
To Tomás and Francelina, my beloved parents.
To all those who have always believed.

ACKNOWLEDGEMENTS

Here I am at the end of a long journey, which has involved travelling day and night, not always in the best conditions. In my early childhood, my parents told me about a stubborn seagull with an immense desire to fly high and perform incredible acrobatics. That same will and the principles inherent in the seagull's story have guided me through life.

I express my sincere gratitude to my supervisors, Professor Rute Costa and Doctor Toma Tasovac. To Professor Rute, for introducing me to the theoretical foundations of terminology, for the exchange of experiences and different visions, for knowing how to guide me constructively, for the contagious energy and commitment, and above all, for opening my eyes to a whole new world. Thank you, Doctor Toma, for the many fruitful discussions, excellent feedback, and encouragement you gave me during this journey where we shared and exchanged so many ideas about our shared passion: dictionaries.

A special thank you to my mentor in geological sciences, Professor Lemos de Sousa, who was always available to give advice and impart excellent lessons. I also thank Professors Telmo Verdelho and Artur Anselmo for having challenged me to venture into a PhD programme, and to all the members and confreres of the Academia das Ciências de Lisboa, some of whom have already left, but whose consistent encouragement I always enjoyed. Special thanks to Alberto Simões, who has been my right-hand in the lexicographic adventure and has been relentless in solving the issues that emerged. I would also like to thank my fellows and researchers at NOVA CLUNL, Bruno, Margarida, Raquel and Sara; and Maria João Ferro, for carefully proofreading the draft. To Isabel Maria, who started this path with me and supported me in the very beginning.

I cannot forget the connections I made worldwide that have greatly advanced me as a researcher and as a humble human being, eager to have a chat. To ELEXIS, which provided me with a stay at the ILex of the Real Academia Española, where I met other professional lexicographers and shared lexicographic know-how, frustrations and many ambitions.

Last but not least, I would like to express my deepest gratitude to my family and friends who inspired and supported me through this work. To Celso, always attentive and available. To André, for sharing and constant interest. To Inês, for her comradeship during my stay in Lisbon. And to all those who took care of my parents so that I could finish this journey more peacefully. To my family and my brothers, of course, who have to listen to my eternal and tireless venting, and especially to Rico, Pedro, Maria Antónia, Jojó and Guiomar, who were there for me at different times.

Charles Bukowski said it much better than me, and may it serve as an inspiration to us all:

if you're going to try, go all the way. // otherwise, don't even start.
[...]
if you're going to try, // go all the way.
there is no other feeling like // that.
you will be alone with the gods // and the nights will flame with fire.
do it, do it, do it. // do it.
all the way // all the way.

TERMINOLOGICAL METHODS IN LEXICOGRAPHY: CONCEPTUALISING, ORGANISING AND ENCODING TERMS IN GENERAL LANGUAGE DICTIONARIES

Ana Maria de Castro Faria Salgado

RESUMO

Os dicionários de língua geral apresentam inconsistências de uniformização e cientificidade no tratamento do conteúdo lexicográfico especializado. Analisando a presença e o tratamento de termos em dicionários de língua geral, propomos um tratamento mais uniforme e cientificamente rigoroso desse conteúdo, considerando também a necessidade de compilar e alinhar futuros recursos lexicais em consonância com padrões interoperáveis. Partimos da premissa de que o tratamento dos itens lexicais, sejam unidades lexicais (palavras em geral) ou unidades terminológicas (termos ou palavras pertencentes a determinados domínios), deve ser diferenciado, e recorremos a métodos terminológicos para tratar os termos dicionarizados.

A nossa abordagem assume que a terminologia – na sua dupla dimensão linguística e conceptual – e a lexicografia, como domínios interdisciplinares, podem ser complementares. Assim, apresentamos objetivos teóricos (aperfeiçoamento da metalinguagem e descrição lexicográfica a partir de pressupostos terminológicos) e práticos (representação consistente de dados lexicográficos), que visam facilitar a organização, descrição e modelização consistente de componentes lexicográficos, nomeadamente a hierarquização das etiquetas de domínio, que são marcadores de identificação de léxico especializados. Queremos ainda facilitar a redação de definições, as quais podem ser otimizadas e elaboradas com maior precisão científica ao seguir uma abordagem terminológica no tratamento dos termos.

Analisámos os dicionários desenvolvidos por três instituições académicas distintas: a Academia das Ciências de Lisboa, a Real Academia Española e a Académie Française, que representam um valioso legado da tradição lexicográfica académica europeia. A análise inicial inclui um levantamento exaustivo e a comparação das etiquetas de domínio usadas, bem como um debate sobre as opções escolhidas e um estudo comparativo do tratamento dos termos. Elaborámos, depois, uma proposta metodológica para o tratamento de termos em dicionários de língua geral, tomando como exemplo dois domínios, GEOLOGIA e FUTEBOL, extraídos da edição de 2001 do dicionário da Academia das Ciências de Lisboa. Revimos os termos selecionados de acordo com os princípios terminológicos defendidos, dando origem a sentidos especializados revistos/novos para a primeira edição digital deste dicionário. Representamos e anotamos os dados usando as especificações da TEI Lex-O, uma extensão da TEI (Text Encoding Initiative), dedicada à codificação de dados lexicográficos. Destacamos também a importância de ter etiquetas de domínio hierárquicas em vez de uma lista simples de domínios, vantajosas para a organização dos dados, correspondência e possíveis futuros alinhamentos entre diferentes recursos lexicográficos.

A investigação revelou que a) os modelos estruturais dos recursos lexicais são complexos e contêm informação de natureza diversa; b) as etiquetas de domínio nos dicionários gerais da língua são planas, desequilibradas, inconsistentes e, muitas vezes, estão desatualizadas, havendo necessidade de as hierarquizar para organizar o conhecimento especializado; c) os critérios adotados para a marcação dos termos e as fórmulas utilizadas na definição são díspares; d) o tratamento dos termos é heterogéneo e formulado de diferentes formas, pelo que o recurso a métodos terminológicos podem ajudar os lexicógrafos a redigir definições; e) a aplicação de métodos terminológicos e lexicográficos interdisciplinares, e também de padrões, é vantajosa porque permite a construção de bases de dados lexicais estruturadas, concetualmente organizadas, apuradas do ponto de vista linguístico e interoperáveis. Em suma, procuramos contribuir para a questão urgente de resolver problemas que afetam a partilha, o alinhamento e vinculação de dados lexicográficos.

Palavras-chave: Academia, dicionário de língua geral, humanidades digitais, interoperabilidade, lexicografia, TEI Lex-O, termo, terminologia

TERMINOLOGICAL METHODS IN LEXICOGRAPHY: CONCEPTUALISING, ORGANISING AND ENCODING TERMS IN GENERAL LANGUAGE DICTIONARIES

Ana Maria de Castro Faria Salgado

ABSTRACT

General language dictionaries show inconsistencies in terms of uniformity and scientificity in the treatment of specialised lexicographic content. By analysing the presence and treatment of terms in general language dictionaries, we propose a more uniform and scientifically rigorous treatment of this content, considering the necessity of compiling and aligning future lexical resources according to interoperable standards. We begin from the premise that the treatment of lexical items, whether lexical units (words in general) or terminological units (terms or words belonging to particular subject fields), must be differentiated, and resort to terminological methods to treat dictionary terms.

Our approach assumes that terminology – in its dual dimension, both linguistic and conceptual – and lexicography, as interdisciplinary domains, can be complementary. Thus, we present theoretical (improvement of metalanguage and lexicographic description based on terminological assumptions) and practical (consistent representation of lexicographic data) objectives that aim to facilitate the organisation, description and consistent modelling of lexicographic components, namely the hierarchy of domain labels, as they are specialised lexicon identification markers. We also want to facilitate the drafting of definitions, which can be optimised and elaborated with greater scientific precision by following a terminological approach for the treatment of terms.

We analysed the dictionaries developed by three different academic institutions: the Academia das Ciências de Lisboa, the Real Academia Española and the Académie Française, which represent a valuable legacy of the European academic lexicographic tradition. The initial analysis includes an exhaustive survey and comparison of the domain labels used, as well as a debate on the chosen options and a comparative study of the treatment of the terms. We then developed a methodological proposal for the treatment of terms in general language dictionaries, exemplified using terms from two domains, GEOLOGY and FOOTBALL, taken from the 2001 edition of the dictionary of the Academia das Ciências de Lisboa. We revised the selected terms according to the defended terminological principles, giving rise to revised/new specialised meanings for the first digital edition of this dictionary. We represent and annotate the data using the TEI Lex-0 specifications, a TEI (Text Encoding Initiative) subset for encoding lexicographic data. We also highlight the importance of having hierarchical domain labels instead of a simple list of domains, which are beneficial to the data organisation itself, correspondence and possible future alignments between different lexicographic resources.

Our investigation revealed the following: a) structural models of lexical resources are complex and contain information of a different nature; b) domain labels in general language dictionaries are flat, unbalanced, inconsistent and often outdated, requiring the need to hierarchise them for organising specialised knowledge; c) the criteria adopted for marking terms and the formulae used in the definition are disparate; d) the treatment of terms is heterogeneous and formulated differently, whereby terminological methods can help lexicographers to draft definitions; e) the application of interdisciplinary terminological and lexicographic methods, and of standards, is advantageous because it allows the construction of structured, conceptually organised, linguistically accurate and interoperable lexical databases. In short, we seek to contribute to the urgent issue of solving problems that affect the sharing, alignment and linking of lexicographic data.

KEYWORDS: Academy, digital humanities, general language dictionary, interoperability, lexicography, TEI Lex-0, term, terminology

TERMINOLOGICAL METHODS IN LEXICOGRAPHY: CONCEPTUALISING, ORGANISING AND ENCODING TERMS IN GENERAL LANGUAGE DICTIONARIES

Ana Maria de Castro Faria Salgado

RESUMÉ

Les dictionnaires de langue générale présentent des incohérences en termes d'uniformité et de scientificité dans le traitement du contenu lexicographique spécialisé. En analysant la présence et le traitement des termes dans les dictionnaires de langue générale, nous proposons un traitement plus uniforme et scientifiquement rigoureux de ce contenu, compte tenu de la nécessité de compiler et d'aligner les futures ressources lexicales selon des normes interopérables. Nous partons du principe que le traitement des éléments lexicaux, qu'il s'agisse d'unités lexicales (mots en général) ou d'unités terminologiques (termes ou mots appartenant à des domaines particuliers), doit être différencié et recourir à des méthodes terminologiques pour traiter les termes du dictionnaire.

Notre approche suppose que la terminologie – dans sa double dimension linguistique et conceptuelle – et la lexicographie, en tant que domaines interdisciplinaires, peuvent être complémentaires. Ainsi, nous présentons des objectifs théoriques (amélioration du métalangage et description lexicographique basée sur des hypothèses terminologiques) et pratiques (représentation cohérente des données lexicographiques) qui visent à faciliter l'organisation, la description et la modélisation cohérente des composants lexicographiques, à savoir la hiérarchie des étiquettes de domaine, car ce sont des marqueurs d'identification du lexique spécialisé. Nous voulons également faciliter la rédaction de définitions, qui peuvent être optimisées et élaborées avec une plus grande précision scientifique en suivant une approche terminologique pour le traitement des termes.

À ce titre, nous avons analysé les dictionnaires développés par trois institutions académiques différentes : l'Academia das Ciências de Lisboa, la Real Academia Española et l'Académie Française, qui représentent un héritage précieux de la tradition lexicographique académique européenne. L'analyse initiale comprend une enquête exhaustive et une comparaison des étiquettes de domaine utilisées, ainsi qu'un débat sur les options choisies et une étude comparative du traitement des termes. Nous avons ensuite développé une proposition méthodologique pour le traitement des termes dans les dictionnaires de langue générale, illustrée à l'aide de termes de deux domaines, la GÉOLOGIE et le FOOTBALL, tirés de l'édition 2001 du dictionnaire de l'Academia das Ciências de Lisboa. Nous avons révisé les termes sélectionnés selon les principes terminologiques défendus, donnant lieu à des significations spécialisées révisées/nouvelles pour la première édition numérique de ce dictionnaire. Nous représentons et annotons les données en utilisant les spécifications TEI Lex-O, une extension TEI (Text Encoding Initiative) pour le codage des données lexicographiques. Nous soulignons également l'importance d'avoir des étiquettes de domaine hiérarchiques plutôt qu'une liste simple de domaines, qui sont bénéfiques pour l'organisation des données elle-même, la correspondance et les alignements futurs possibles entre différentes ressources lexicographiques.

Notre enquête a révélé ce qui suit : a) les modèles structurels des ressources lexicales sont complexes et contiennent des informations de natures différentes ; b) les étiquettes de domaine dans les dictionnaires de langues générales sont plates, déséquilibrées, incohérentes et souvent désuètes, ce qui nécessite de les hiérarchiser pour les connaissances organisées ; c) les critères appliqués pour marquer les termes et la formule et utilisés dans la définition sont absurdes ; d) le traitement des termes est hétérogène et formulé différemment, les méthodes terminologiques pouvant aider les lexicographes à rédiger des définitions ; e) l'application de méthodes terminologiques et lexicographiques interdisciplinaires ainsi que de normes est avantageuse parce qu'elle permet la construction de bases de données lexicales structurées, conceptuellement organisées, linguistiquement précises et interopérables. En bref, nous cherchons à contribuer à la question urgente de la résolution des problèmes qui affectent le partage, l'alignement et la liaison des données lexicographiques.

MOTS-CLÉS: Académie, dictionnaire de langue générale, humanités numériques, interopérabilité, lexicographie, TEI Lex-O, terme, terminologie

TERMINOLOGICAL METHODS IN LEXICOGRAPHY: CONCEPTUALISING, ORGANISING AND ENCODING TERMS IN GENERAL LANGUAGE DICTIONARIES

Ana Maria de Castro Faria Salgado

RESUMEN

Los diccionarios de lengua general presentan inconsistencias de uniformización y científicidad en el tratamiento del contenido lexicográfico. Analizando la presencia y tratamiento de términos en diccionarios de lengua general, proponemos un tratamiento más uniforme y científicamente más riguroso de ese contenido, considerando la necesidad de compilar y alinear futuros recursos lexicales en consonancia con modelos interoperables. Partimos de la premisa de que el tratamiento de los elementos lexicales, sean unidades lexicales (palabras en general) o unidades terminológicas (términos o palabras pertenecientes a determinados dominios), debe ser diferenciado, y recurrimos a métodos terminológicos para tratar los términos diccionarizados.

Nuestro abordaje asume que la terminología – en su doble dimensión lingüística y conceptual – y la lexicografía, como dominios interdisciplinarios, pueden ser complementarios. Así, presentamos objetivos teóricos (perfeccionamiento del metalenguaje y descripción lexicográfica a partir de presupuestos terminológicos) y prácticos (representación consistente de componentes de datos lexicográficos), que buscan facilitar la organización y modelización consistente de componentes lexicográficos, concretamente la jerarquización de las etiquetas de dominio, que son marcadores de identificación de léxico especializado. Asimismo, queremos facilitar la redacción de definiciones, las cuales pueden ser optimizadas y elaboradas con mayor precisión científica al seguir un abordaje terminológico para el tratamiento de los términos.

Analizamos los diccionarios desarrollados por tres instituciones académicas distintas: la Academia das Ciências de Lisboa, la Real Academia Española y la Académie Française, que representan un valioso legado de la tradición lexicográfica académica europea. El análisis inicial incluyó un rastreo exhaustivo y comparación de etiquetas de dominio usadas en estos diccionarios, así como un debate sobre las opciones escogidas y un análisis comparativo del tratamiento de los términos. Después, elaboramos una propuesta metodológica del tratamiento de términos en diccionarios de lengua general, tomando como ejemplo dos dominios, GEOLOGÍA y FÚTBOL, extraídos de la edición del 2001 del diccionario de la Academia das Ciências de Lisboa. Estos términos fueron revisados de acuerdo con los principios terminológicos que aquí defendemos, dando origen a sentidos especializados revisados/nuevos para la primera edición digital del diccionario académico portugués. Representamos y anotamos los datos usando las especificaciones de la TEI Lex-O, una extensión TEI (Text Encoding Initiative) restringida a la codificación de datos lexicográficos. Destacamos la importancia de tener etiquetas de dominio jerárquicas en vez de una lista simple de dominios, ventajosas para la organización de los datos, correspondencia y posibles futuros alineamientos entre diferentes recursos lexicográficos.

La investigación reveló que a) los modelos estructurales de los recursos lexicales son complejos y contienen información de naturaleza diversa; b) las etiquetas de dominio en los diccionarios de lengua general son planas, desequilibradas, inconsistentes y, muchas veces, están desactualizadas, habiendo necesidad de jerarquizarlas para organizar el conocimiento especializado; c) los criterios adoptados para la marcación de los términos y las fórmulas utilizadas en la definición son dispares; d) el tratamiento de los términos es heterogéneo y formulado de diferentes formas, por lo que recurriendo a métodos terminológicos pueden ayudar a los lexicógrafos a redactar definiciones; e) la aplicación de métodos terminológicos y lexicógrafos interdisciplinarios, y también de modelos, es ventajosa porque permite la construcción de bases de datos lexicales estructuradas, conceptualmente organizadas, precisas desde el punto de vista lingüístico, e interoperables. En suma, procuramos contribuir a la cuestión urgente de resolver problemas que afectan al intercambio, al alineamiento e vinculación de datos lexicográficos.

Palabras clave: Academia, diccionario de lengua general, humanidades digitales, interoperabilidad, lexicografía, TEI Lex-O, término, terminología

LIST OF ABBREVIATIONS

ACL: Academia das Ciências de Lisboa

AF: Académie Française

ASALE: Asociación de Academias de la Lengua Española

DA: *Diccionario de Autoridades*, RAE

DAF: *Dictionnaire de l'Académie Française*, AF

DLE: *Diccionario de la Lengua Española*, RAE

DLP: *Dicionário da Língua Portuguesa*, ACL (forthcoming new digital edition)

DLPC: *Dicionário da Língua Portuguesa Contemporânea*, ACL

ELEXIS: European Lexicographic Infrastructure

ERI: *Entorno de Redacción Integrado*, RAE

GDLP: *Grande Dicionário da Língua Portuguesa*, Porto Editora

HOUAISS: *Grande Dicionário Houaiss da Língua Portuguesa*, Círculo de Leitores

ICS: International Commission on Stratigraphy

ILex: Instituto de Lexicografía, RAE

ILLP: Instituto de Lexicologia e Lexicografia da Língua Portuguesa, ACL

INFOPÉDIA: *Dicionário Infopédia da Língua Portuguesa*

ISO: International Organisation for Standardisation

IUGS: International Union of Geological Sciences

Lemon: Lexicon Model for Ontologies

LLOD: Linguistic Linked Open Data

LOD: Linked Open Data

NLP: Natural Language Processing

NOVA CLUNL: Linguistics Research Centre of NOVA University Lisbon

OED: *Oxford English Dictionary*, Oxford University Press

OWL: Web Ontology Language

POS: Part-Of-Speech

PRIBERAM: *Dicionário Priberam da Língua Portuguesa*

RAE: Real Academia Española

RDF: Resource Description Framework

SKOS: Simple Knowledge Organisation System

TEI P5: *Guidelines for Electronic Text Encoding and Interchange*

TEI: Text Encoding Initiative

UML: Unified Modelling Language

W3C: World Wide Web Consortium

XML: Extensible Markup Language

TYPOGRAPHIC CONVENTIONS

For the sake of consistency, throughout this thesis, we have adopted some typographic conventions as exemplified below:

- Domain labels are written in small caps, e.g., GEOLOGY.
- Terms are written in quotation marks, e.g., “term”. The lemmas extracted from dictionaries are also in quotation marks when considered as terms.
- Concepts are written in angled brackets and with the first letter capitalised in a fixed-width (monospace) font, e.g., <Concept>.
- Characteristics are written with forward slashes, e.g., /characteristic/.
- Concept relation identifiers are written with an underscore between the forms in a fixed-width (monospace) font, e.g., has_relation.
- TEI P5 terms (element names, attribute names, attribute values, etc.) are written in a fixed-width (monospace) font and:
 - for individual element names, we surrounded the name of the element with angle brackets (<entry>);
 - for the names of nested elements, we used the XPath notation, e.g., (cit/quote/bibl);
 - for attribute names, we used the @ sign before the name of the attribute, e.g., @type;
 - for attribute values, we surrounded the string with quotation marks ("), e.g., "domain".

TABLE OF CONTENTS

INTRODUCTION.....	1
Motivation.....	1
Dictionaries as a Case Study	2
Background Issues	6
Problem Statement.....	8
Objectives.....	10
Research Questions	11
Research Methodology.....	11
Thesis Structure	14

PART I – FRAMEWORK ISSUES

CHAPTER 1 Theoretical Background	17
1.1 The Emergence of the Digital Humanities	17
1.2 A Walk Through the Lexicographic Universe	20
1.3 The Twofold Nature of Lexicography	28
1.4 Terminology as an Interdisciplinary Field	33
CHAPTER 2 Dictionaries	42
2.1 Dictionaries are Like Diamonds.....	42
2.1.1 <i>The Dictionary as a Text</i>	44
2.1.2 <i>The Dictionary as a Research Object</i>	45
2.1.3 <i>The Dictionary as a Cultural Artefact</i>	46
2.1.4 <i>The Dictionary as a Tool</i>	47
2.1.5 <i>The Dictionary as a Language Model</i>	48
2.2 Dictionary Classifications.....	50
2.2.1 <i>An Overview of Dictionary Classifications</i>	50
2.2.2 <i>Taxonomic Classification Proposal</i>	54
2.3 Dictionary Structure.....	59
2.4 Going Further: Modelling and Standardising Lexicographic Resources	62
CHAPTER 3 European Lexicographic Tradition	64
3.1 The Origins of Lexicography	65
3.2 The First Monolingual Dictionaries	67
3.3 The Rise of the Academy Tradition	69
3.3.1 <i>Académie Française</i>	72
3.3.1.1 <i>Dictionnaire de l'Académie</i>	74
3.3.1.2 <i>Le Dictionnaire de l'Académie française est en ligne</i>	78
3.3.2 <i>Real Academia Española</i>	79
3.3.2.1 <i>Diccionario de la Lengua Española</i>	82
3.3.2.2 <i>Diccionario de la Lengua Española en línea</i>	83
3.3.3 <i>Academia das Ciências de Lisboa</i>	84
3.3.3.1 <i>The First Attempts at Making a Dictionary</i>	86
3.3.3.2 <i>Dicionário da Língua Portuguesa Contemporânea</i>	90
3.3.3.3 <i>Dicionário da Língua Portuguesa</i>	92
3.4 Final Considerations	93
CHAPTER 4 Usage Labels in General Language Dictionaries.....	95
4.1 Labelling Practices	95
4.2 Labels: Definition and Practices	98

4.2.1 What Is a Label, Really?	98
4.2.2 What Does a Label Label?	99
4.2.3 Form and Position of Usage Labels	100
4.2.4 Purpose and Role of Usage Labels.....	103
4.3 Classifying Usage Labels: An Overview	104
4.3.1 Diachronic Marking	106
4.3.2 Diatopic Marking.....	107
4.3.3 Diaintegrative Marking	108
4.3.4 Diastratic/Diaphasic/Diatextual Marking	108
4.3.5 Diafrequentative Marking	108
4.3.6 Diaevaluative Marking	109
4.3.7 Dianormative Marking	109
4.3.8 Diasemantic Marking	110
4.3.9 Diatechnical Marking	111
4.4 The Domain Label	111
4.4.1 Types of Domain Labels.....	113
4.4.2 The Domain Label as a Challenging Lexicographic Issue	115
4.4.3 Organisation of Domain Labels	115
CHAPTER 5 Terms in General Language Dictionaries.....	117
5.1 Terms in General Dictionaries: To Include or Not To Include?	117
5.2 Research on the Inclusion of Terms in General Dictionaries	125
5.3 Dealing with Terms in General Dictionaries	129
5.3.1 Term and Concept	129
5.3.2 Term as a Polylexical Unit	132
5.3.3 Term and Domain.....	135
5.3.4 Term and Definition.....	136
PART II – DATA ANALYSIS AND PROCESSING	
CHAPTER 6 Coverage and Treatment of Terms in Academy Dictionaries	143
6.1 Lexicographic Data Analysis	143
6.1.1 Analysis of the Dictionaries' Front Matter.....	144
6.1.2 List of Abbreviations.....	146
6.1.3 Exploring Labelling Practices	149
6.1.4 Domain Lists	150
6.2 Comparison Between Results	155
6.2.1 Mapping Domains	159
6.2.2 Domain Organisation	163
6.3 Geology and Football Domains: Analysis of Lexicographic Articles	169
6.3.1 Geological Terms	169
6.3.2 Football Terms.....	178
6.4 Final Considerations.....	188
CHAPTER 7 A Terminological Approach for Lexicographic Purposes	191
7.1 Terminological Working Methods for Lexicographic Work	191
7.2. Establishing the Lexicographic Source Corpus (dictionary).....	197
7.3 Delimiting the Domain.....	198
7.3.1 The Geology Domain as a Case Study	199
7.3.2 The Football Domain as a Case Study	203
7.4 Organising the Domain.....	206
7.4.1 Comparing Classification Systems	206
7.4.2 Hierarchising domain labels	212
7.5 Extracting Terminological Data	221

7.6 Organising Terms	222
7.7 Validating Terminological Data	222
7.7.1 Domain organisation	222
7.7.2 Terms	222
7.8 Modelling Concept Systems	224
7.9 Editing Lexicographic Content	240
7.9.1 Identifying Definitory Problems	240
7.9.2 Reformulation Definitions and Notes	241
7.10 Validating Terminological Data	252
7.10.1 Concept Systems	253
7.10.2 Definitions and Notes	253
7.11 Encoding Terms	254
7.12 Publishing Terms	255

PART III – ENCODING AND MODELLING DICTIONARIES

CHAPTER 8 Standards for Structured Lexicographic Resources	259
8.1 ISO Standards for Lexicography	260
8.2 Simple Knowledge Organisation System	264
8.3 OntoLex-Lemon	265
8.4 Lexical Markup Framework	268
8.5 Text Encoding Initiative	269
8.5.1 The TEI Dictionary Module	273
8.5.2 The TEI Lex-0	274
CHAPTER 9 TEI Lex-0 in action	276
9.1 Different Views of Modelling	277
9.2 The DLPC and DLP as a TEI Dictionary Projects	279
9.2.1 Basic Structure of an Entry	281
9.2.2 Macrostructural Level	283
9.2.3 Microstructural Level	288
9.3 Encoding Terms	290
9.3.1 Encoding Domain Labels	290
9.3.2 Encoding Polylexical Terms	299
9.3.3 Encoding Semantic Relations	303
9.3.4 Encoding Other Components	306
CONCLUDING REMARKS	308
BIBLIOGRAPHY	316
LIST OF FIGURES	346
LIST OF TABLES	351
ANNEXES	352
ANNEX 1	353
ANNEX 2	357
ANNEX 3	361
ANNEX 4	362
ANNEX 5	364
ANNEX 6	368

INTRODUCTION

I know of no more enjoyable intellectual activity than working on a dictionary.

HULBERT (1955, p. 42)

Motivation

An old passion for lexicography and a more recent interest in terminology were instrumental in choosing a research subject that would combine these two separate but interconnected universes. Bearing this in mind, we chose a shared study object – the term.

At first glance, it may seem as though terminology science understood as a ‘science studying *terminologies*’ (ISO 1087, 2019, p. 2) does not fit within general language dictionaries. While a terminological dictionary only collects specialised lexical units that are related to a concept (and thus, each lemma is a term), general language dictionaries are the product of a discourse made by lexicographers, which includes as lemmas lexical units that can either belong to the general language or to a particular subject field. The practice of including terms in general language dictionaries is not new. Still, we argue that the lexicographic methodology would benefit significantly from terminological assumptions.

Right from the start, we wanted to analyse the inclusion and treatment of terms belonging to different subject fields in general language dictionaries. In other words, we aimed to study *terminologies* understood as the ‘set of *designations* and *concepts* belonging to one *domain* or *subject*’ (ISO 1087, 2019, p. 2). Here, the ‘set of *designations*’ points to terms, a ‘*designation* that represents a *general concept* by linguistic means’ (ISO 1087, 2019, p. 7).

We must also clarify that for the purpose of our research, the term is always understood as a specialised lexical unit and not as a general lexical unit, as one may assume by consulting some general language dictionaries (e.g., INFOPÉDIA; PRIBERAM;

DLE)¹. This work does not aim to reflect theoretically on what a term *is* but rather to establish how terms *should be* treated in general language dictionaries.

The title of this thesis highlights our belief that terminology as a *science* with its own methodology and multidisciplinary nature – drawing support from various disciplines, such as philosophy, epistemology, logic, information science, linguistics and translation studies, and intersecting with all other subject fields that provide material for terminological work (ISO 704, 2009) – can contribute to a practice-based rethinking of lexicographic work when a lexicographer has to deal with terms. We will demonstrate in these pages that terminological methods are advantageous for the process of lexicographic knowledge-building, making it possible to conceptualise and organise knowledge. We will dedicate our research to systematically studying the domain labelling system and guiding the drafting of definitions in general language dictionaries.

Dictionaries as a Case Study

Even if someone never looks up a word in a dictionary, they will still hold a copy of one – perhaps abandoned or forgotten – on one of their shelves at home. In a way, we can say that people know what a dictionary is. Thus, “dictionary” is a term that may seem very simple to define at first glance. Nevertheless, as we will explore more deeply in Chapter 2, although the usefulness of a dictionary is widely recognised, when someone starts researching into dictionaries, they realise the extreme complexity involved.

The concept of a dictionary as a repertoire is present in the very etymology² of the word. Although dictionaries have always been considered consultation objects *par excellence* and are not precisely intended to be read from cover to cover³, a dictionary

¹ INFOPÉDIA and PRIBERAM define ‘*termo*’ [term] as ‘*vocábulo; palavra*’ [vocabulary; word]. The DLE also defines it as ‘*palabra (|| unidad lingüística)*’ [word (linguistic unit)].

² The word ‘dictionary’ comes from the medieval Latin *dictionary*, which means repertoire of *diciones* (phrases or words), formed on the Latin *dictiō*, or ‘the action of saying’, plus the suffix *-arium*, which conveys the notion of collection.

³ Let us remember the words of D’Alembert, in his ‘*Discours Préliminaire*’ to the *Encyclopédie*: ‘*les Dictionnaires par leur forme même ne sont propres qu’à être consultés, & se refusent à toute lecture suivie*’ [Dictionaries due to their very own form are only suitable for consultation and cannot be read from end to end]. See <https://encyclopedie.uchicago.edu/node/88>.

can be many things simultaneously. Restricting the dictionary concept to its primary function, i.e., consultation, or stating that it is only a book that contains meanings, falls short of the truth.

A traditional dictionary definition usually indicates that it is ‘a book’ (OED) or a reference book (*‘obra de referência’*, INFOPÉDIA) that explains the meaning of a set of lexical units of a language according to an agreed order, ‘usually in alphabetical order’ (OED). These definitions, although still present in many contemporary dictionaries, are outdated. The mental image that most of us have of a dictionary is undoubtedly that of the book, which in itself indicates the cultural importance that these works have assumed throughout history. From the mid-20th century onwards, even if it were reasonable to define a dictionary as a lexical resource, for example, as ‘an electronic resource’ (another OED definition), there are still very few dictionaries that describe themselves as such.

The dictionary as a book is no longer successful, especially from a commercial point of view. However, there is another side to this coin. The irreversible transition to the digital environment has imposed on lexicography (and the humanities and social sciences in general) the challenge of adopting new methods concerning the traditional research methodology. It has led to the need to rethink certain topics in order to create strategies that will respond to better-quality data and sustainable, operational, accessible and long-term preservable practices. This paradigm shift requires a confluence of knowledge. And much has already been done. In this regard, see the number of existing articles that already account for this convergence, for example, ‘[science x] meets [science y]’, because synergies are more crucial now than ever. There is a crossover of various disciplines involving different specialists in any dictionary project. Several scholars have discussed the nature of interdisciplinarity in lexicography (e.g., Nielsen, 2018; Hartmann, 2005), and we argue that the work of a lexicographer and that of a terminologist should be complementary (Costa, 2013).

Considering that lexicographic resources constitute a valuable linguistic and cultural heritage in our multilingual society, this research aims to underline the importance of general language dictionaries and to emphasise the need to apply

consistent and well-explained linguistic methods and standards to ensure their necessary scientific accuracy, preservation, interoperability and reusability.

We chose general language dictionaries because they are repositories that aim to make a complete inventory of a language, ideally recording every lexical unit that can be found in a particular language. This type of lexicographic work assembles and describes the lexicon of a particular language. In a well-structured way, as referred to above, these information repositories contain units belonging to the general lexicon and others from specialised knowledge fields. Under some conditions, the latter can also be integrated into the so-called general lexicon. However, this type of dictionary not only gathers or provides the meaning or evolution of lexical units through time but also puts together pronunciation, syllabification, etymology, and information about the usage of certain items in the communication system conveyed by specific labels, to give just a few examples. Therefore, the value and importance of these works for the communities of speakers is unquestionable since they are instrumental as a learning resource and a cultural work for the affirmation of the language and the nation.

As the first digital edition of the Academia das Ciências de Lisboa's dictionary⁴ (DLP) is being coordinated by Salgado, we decided to use this dictionary as a starting point and to take a contrastive turn in our work, investing efforts in a broader multilingual view within the European lexicographic scenario. So as not to restrict our research to the national level, we selected other dictionaries produced by academies as our objects of study. Thus, our research project is based on three main lexicographic works:

- The dictionary of the Academia das Ciências de Lisboa (*Dicionário da Língua Portuguesa Contemporânea*, henceforth, DLPC);
- The dictionary of the Real Academia Española (*Diccionario de la lengua española*, hereinafter, DLE);
- The dictionary of the Académie Française (*Dictionnaire de l'Académie Française*, hereinafter, DAF).

⁴ DLP or *Dicionário da Língua Portuguesa* is the title of the new dictionary project that stems from the DLPC and is being updated and revised.

Historically speaking, all three dictionaries were created based on the so-called ‘academy principle’⁵ (Considine, 2014; see Chapter 3, note 42, p. 68), i.e., the established need to conserve and perfect the language, regulating its usage, vocabulary and grammar. Nevertheless, these dictionaries are authoritative⁶ in their respective languages because they were produced by regulatory bodies, i.e., the academies, issuing recommendations and guidelines regarding the use of each language. Each of the chosen dictionaries is a general language monolingual dictionary of a Romance language (Portuguese, Spanish and French), covers a wide range of terms and addresses a vast potential audience of speakers on multiple continents. All three dictionaries started as print dictionaries, and now each one has an online version that is currently being updated.⁷ At the same time, these dictionaries have a heterogeneous structure in terms of lexical data representation. With their ‘pursuit of completeness concerning the entries relevant to subject matters’ (Kinable, 2015), academy dictionaries present detailed lexicographic information and elaborate microstructure, which can more often than not pose challenges in terms of consistent data modelling.

The relevance of doing comparative work in monolingual lexicography is magnified by the technical and scientific development of a globalised society, where well-documented and structured data and knowledge must be shared. Globalisation implies a constant interaction between individuals from different countries and cultures, where language is the medium that conveys the specific culture of each country. Comparing the various monolingual lexicographic resources developed by different countries is a crucial task, as there is a need to interconnect their respective datasets and achieve data interoperability. While, on the one hand, the heterogeneity of these resources is evident, and somehow it will have to be maintained so as not to lose the specifics of each of these works, on the other hand, it is necessary to work on the homogenisation of these data using agreed-upon standardised works in machine-readable formats.

⁵ We want to note that, throughout this work, when referring to the dictionaries produced by these institutions, we will use the term ‘academy dictionaries’, obviously inspired by the reference work by Considine (2014).

⁶ The question of authority is relative and its influence varies from country to country.

⁷ In the case of the Portuguese academy dictionary, the online version is still private.

The need to create and make available structured, organised and interoperable lexical resources has led us to follow a path in which the application of standards and best practices for representing and modelling all the components that constitute a lexicographic article are fundamental requirements. So, the author of this thesis invested much time in various courses, summer schools, and specialised training, which must be highlighted since they impacted the present research. We begin by referring to the highly specialised training in terminology at conferences such as the TOTH International Conference; the courses 'Terminology and Lexicography' and 'From Print to Screen: The Theory and Practice of Digitising Dictionaries' in the scope of the Lisbon Summer School in Linguistics, 2018 edition; and participation in the Lexical Data Masterclass in December 2018 that took place at the Berlin-Brandenburg Academy of Sciences. Subsequently, the idea of associating the analysis and treatment of lexicographic data with its encoding and modelling emerged after we started to make some contributions to the Digital Research Infrastructure for the Arts and Humanities (DARIAH) Working Group on Lexical Resources.

It is also worth mentioning that this work has benefited a lot from a lexicographic project currently underway: the European Lexicographic Infrastructure (ELEXIS). A scholarship granted for this project allowed Salgado a four-week stay at the Instituto de Lexicografía (ILex) of the Real Academia Española, and she has participated actively in the ELEXIS project under this scholarship since 2020. The stay in Spain allowed exploration of the DLE, getting to know the work methodology and discussing and sharing ideas with the team of lexicographers while collecting important data for this research. We also have to thank the Académie Française for sharing the list of domains included in DAF, which was essential to conducting this research.

Background Issues

Addressing the issue of how terms are dealt with in monolingual general dictionaries requires an early examination of the theoretical framework in which the present investigation has been developed.

Due to the technological and scientific boom, terms are exceptional sources of lexical renewal and enrichment of the language systems, and their registration in dictionaries is no exception. That is why the inclusion of *terminologies* in general language dictionaries has increased. Although terms included in dictionaries may have gone through a process of determinologisation (Costa et al., 2021b, p. 128) – a concept that will be explored in Chapter 5 –, our methodology reveals that terminological principles contribute to a better organisation of data regarding, for example, the hierarchy of domains, as well as contributing to a better description of lexicographic articles, namely by adding accuracy to the lexicographic definition in which the conceptual dimension helps the writing process.

When focusing on the portion of the lemma list that is made up of terms, our viewpoint will have to aim at the markers that restrict and identify the specialised knowledge field of a given lexical item. Such markers are known as domain labels. Analysing, integrating and combining high-quality lexicographic data from different sources and between different languages requires, among other things, a clear understanding of the mutual (in)compatibility of the labels used in different dictionaries throughout the world, primarily when these dictionaries rarely communicate their classification criteria or the details of their underlying decision-making process.

Thus, one of this thesis's main contributions is to analyse, confront and discuss the different domain labels used in academy dictionaries and show how the currently recommended TEI practice for representing domain labels as flat values is not robust enough to deal with more complex, hierarchical domain structures.

We believe that these new methodological perspectives are necessary to increase the quality of the organisational and structural model of dictionaries and lexicographic descriptions, as well as to take advantage of the digital environment. We aim to invest, above all, in a qualitative improvement of lexical data and how they are modelled, i.e., we argue that a good organisation of knowledge and an accurate linguistic analysis of the components of a lexicographic article will make it easier for users to navigate the dictionary and locate the specific information they are looking for. Nevertheless, let us step back a little to explain and justify why we chose this research topic.

Problem Statement

Although the digital revolution has unquestionably transformed the concept of dictionary, much of the lexicographer's basic work remains – hunting for new words, describing them, updating and completing existing records. However, everything is implemented differently, starting with many current post-editing methods and the necessity to deal with a significant amount of lemmas or meanings belonging to different fields of knowledge in which the lexicographer is not an expert. This corresponds to one of the great difficulties in a lexicographer's daily work.

Since lexicography and terminology have different theoretical and methodological assumptions, we start from the premise that the treatment of lexicographic units, depending on whether they are lexical (words in general) or terminological (terms), must be divorced from the postulation that lexicography and terminology are two disciplines with different theoretical and methodological assumptions and whose final products aim to respond to different social needs. However, since general language dictionaries also include terms, we advocate adopting a holistic approach that breaks down barriers between lexicography and terminology, and even other disciplines, as Leroyer and Simonsen (2020) argued when they recently proposed a reconceptualisation of lexicography.

General language and terminological dictionaries are different reference objects regardless of how the dictionary content is represented and made available. The language dictionary functions as a repository that integrates the set of lexical units of a given linguistic system, presenting information related to the meanings used in specific contexts of each lexical item. In turn, the terminological dictionary contains terminological units and describes/defines the concepts or objects of one or several subject fields for a more restricted target audience. These two types of dictionaries present different information because they respond to different social needs. But a general language dictionary actually also contains lexical items that are considered terms insofar as they designate concepts that are part of concept systems of general knowledge. The difficulty of establishing boundaries between linguistic knowledge and conceptual knowledge makes it impossible to separate the material collected in a general language dictionary from what is found in a terminological dictionary (Iriarte

Sanromán, 2001, p. 231). Because of the differences between terminological and lexicographic dictionaries, which we have outlined in the previous paragraph, we believe it is crucial to understand how these lexical items are included and treated in this type of lexical resource.

Based on the analysis of the lexicographic and traditional dictionary's theoretical and methodological principles, we conclude that the methodology or criteria adopted are never appropriately explicit. The front matter of the dictionaries under study, as we will demonstrate, does not include the criteria for inclusion and treatment of specialised senses. The scientific community recognises that there is '*uma espécie de lexicografia anómala*' [a kind of anomalous lexicography] (Verdelho, 1998, p. 27), which has been carried out in '*modo artesanal*' [a crude way] (*ibidem*), because it is based more on the lexicographer's intuition (Correia, 2008, p. 9) than on a '*clasificación científica de tecnolectos*' [scientific classification of technolects] (Haensch et al., 1982, p. 497). On the other hand, and because the organisation of domains is fundamental to a good structuring and conceptualising of knowledge and, consequently, to proper lexicographic treatment, it is crucial to fill this gap that has already been identified by Guilbert (1973), who stressed that terms establish relationships with each other and that this fact has been neglected in most dictionaries. As it is easier to highlight these relationships in the digital domain, we will take this opportunity to conceptualise and organise the domains found in the dictionaries under study.

Thus, this research project aims to debate certain decisions traditionally taken by lexicographers. In our view, the methodology usually adopted needs to be reformulated, especially regarding the use of domain labels – which seem to be more the result of a lexicographic heritage than of a scientific domain questioning or an accurate proposal for taxonomic classification.

In addition to the problem related to the labelling system, which, as we will see, differs between the various dictionaries, we pay special attention to the description (or lexicographic definition) of the meaning of terms in the lexicographic article. As stated by Iriarte Sanromán (2001), '*a diferença entre um dicionário terminológico e um dicionário de língua não estará tanto no tipo de unidades utilizadas – o que na prática corresponderá à seleção de entradas (nomenclatura ou macroestrutura [...] – como no*

tipo de definição utilizada’ [the difference between a terminological dictionary and a language dictionary does not lie so much in the type of units used – which in practice will correspond to the selection of entries (nomenclature or macrostructure [...] – as in the type of definition used] (p. 226). In this sense, we discuss the consistency of the current definitions and which formulae refer to specialised contexts, and we propose optimising the terms’ definitional wording. We argue that definitions of terms, even in general language dictionaries, must be ‘the linguistic description of a concept, based on the listing of a number of characteristics, which conveys the meaning of the concept’ (Sager, 1990, p. 39). When dealing with terms, the lexicographer must write a definition that fixes the *intension* (ISO, 704, 2009, p. 6) of the concept, i.e., first identifying the characteristics that make up the concept. Thereafter, the concept must be analysed in relation to others in the same concept system.

Objectives

In short, we aim to meet the following objectives:

- (1) Examine the presence of terms in academy dictionaries.
- (2) Propose a more uniform and consistent use of domain labelling in academy dictionaries to promote interoperability.
- (3) Identify, organise and describe some of the different levels of linguistic knowledge in dictionary articles, focusing on domain labelling and the definition of terms.
- (4) Show how consistent lexicographic data encoding – in this case, the use of TEI Lex-0 – can help us to rethink the theoretical and methodological assumptions of the treatment of terms in the lexicographic tradition, and discuss its applicability in the representation of lexicographic data.
- (5) Create and develop a mixed methodology that can be replicated when dealing with terms of other domains.
- (6) Propose the best practices for harmonising and encoding terms in TEI Lex-0.

Research Questions

The points identified above have led us to ask the following questions:

- (i) Might principles and methods of terminology work contribute to lexicographic work?
- (ii) How are terms treated in general language dictionaries, namely in academy dictionaries?
- (iii) What domains are currently represented in these works? Are those domains conceptually organised?
- (iv) What is the role or function of the domain label in academy dictionaries?
- (v) Is it possible to map the domain labels between the different academy lexicographic resources?
- (vi) If we organise the domains, identify the concepts and the relations between them, model concept systems and then search for the terms linked to the identified concepts, will it improve the definitions of terms?
- (vii) Do the TEI Lex-O's specifications meet the identified requirements to represent terms?

Research Methodology

This research project is governed by the premise that terminology, as an interdisciplinary domain, has a double dimension (Costa, 2013; Santos & Costa, 2015; Roche, 2015). As we will see, the linguistic dimension, focused on terms, and the conceptual dimension, focused on concepts, are not antagonistic. The complementarity between these two systems is achieved by iteratively following two different approaches: the semasiological and the onomasiological. In this context, we will describe the method we apply to treat terms in general language dictionaries, mainly backed by ISO 704 (2009) and ISO 1087 (2019).

Terminologists and lexicographers have different perspectives. Even if both start with an existing collection of terms, terminologists concentrate their activity primarily on the structure of knowledge, privileging an onomasiological approach. In contrast, a lexicographer starts from the collected lexical units to identify their meaning, pushing

the concept to a secondary level, or ultimately disregarding it. According to Sager (1990), the lexicographer ‘collects *all* the words of a language to sort them in various ways. Once he has collected these words, he proceeds to differentiate them by their meanings’ (p. 55). In turn, the terminologist ‘starts out from a much narrower position; he is only interested in subsets of the lexicon, which constitute the vocabulary (or lexicon) of special languages’ (*ibidem*).

While the lexicographic methodology follows a semasiological path, in the sense that it begins from an existing corpus of lexical units to explore their semantic values, the terminological methods first try to identify the concepts and subsequently order the terms found by reference to a concept system, following an onomasiological approach and resorting to the construction of conceptual representations of the domains under analysis. These different approaches should not be seen as antagonistic; in fact, they are quite the contrary: ‘*la perspective linguistique, plutôt sémasiologique et la perspective conceptuelle, plutôt onomasiologique, [...] ne s’excluent pas mutuellement, mais se complètent*’ [the linguistic perspective, which is more semasiological, and the conceptual perspective, which is more onomasiological, [...] are not mutually exclusive; they are complementary] (Costa, 2006a, p. 85). In this process, the consultation with a subject field specialist plays a fundamental role in the validation stages.

By conceiving the language dictionary as a *repository of meanings* and the terminological dictionary as a *repository of terms*, we can establish a continuum and a complementarity between lexicographic and terminological work (Costa, 2013, p. 29; Iriarte Sanromán, p. 91). Thus, in light of the double dimension of terminology, we stress the relevance of the systematisation of concept designations – a lexical network based on the lexical-semantic relations established between terms – assuming that a concept systematisation underlies the systematisation of terms and their respective relations in the two domains selected for this purpose.

We anticipate that dealing with *terminologies* is a very ambitious task, which is why we decided to restrict our research to two domain labels and related fields: GEOLOGY and FOOTBALL. The former is a highly specialised domain, and the latter is familiar to most speakers. Thus, we assume that this may influence our methodology. The selection of two domains that are distant from each other is intentional, as it allows us to test the

proposed scenario. In this research, interaction with specialists and professionals from the areas under analysis plays a fundamental role in clarifying doubts or ambiguities that may arise, contributing to a good understanding of the domains and the lexicographic treatment assigned. In turn, the semasiological analysis of lexicographic articles will take place after the organisation of the domain knowledge, and the definitions may be analysed for onomasiological purposes.

Combining the lexicographic methodology with terminological assumptions will be an advantage when planning the macrostructure and microstructure of a dictionary, i.e., for the organisation and description of the lexicographic articles so that dictionaries become more scientifically accurate and guarantee greater scientific exactness, both for lexicographers when editing lexicographic articles with specialised senses and for end users. The need for terminological research in lexicographic work arises when organising knowledge or analysing a subset of terms is necessary. Lexicographers and terminologists together can guarantee better and more accurate solutions.

The ultimate goal of this methodology is to propose strategies that can help lexicographers write definitions. Meeting this need, we will address one of the most problematic tasks for any lexicographer – how to feel more secure when defining terms of subject fields that they have not mastered.

Concerning the representation of lexicographic data, having been aware of the development of a new specific TEI format for encoding dictionaries led us to experiment TEI Lex-0, although we have been following the *TEI Guidelines for Electronic Text Encoding and Interchange* (TEI P5) in our lexicographic work in the last past years. This new format is used in the context of the ELEXIS. We also adopted this scheme in the DLP, where we have been experimenting with the best way to represent a hierarchical proposal of the domain labels under study, which will be presented here. This simplified format involves a critical analysis of the guidelines as applied to dictionaries, and we have collaborated and discussed some recommendations with the DARIAH Working Group on Lexical Resources. The constraints of this new format are potentially advantageous for data sharing and future dictionary alignment (e.g., Ahmadi et al., 2021; Martelli et al., 2021; Salgado et al., 2020).

This research deals with general language dictionaries as working tools and reference works that are widely used to broaden knowledge and presents theoretical (improvement of metalanguage and lexicographic description using terminological assumptions) and practical (consistent representation of lexicographic data) objectives, investing in the quality of lexicographic products, which are governed by theoretical and methodological principles that enable the desired interoperability – a key concept in the digital age, and consequently an improvement in the users' linguistic skills. The results of this research will be directly applied to the new digital version of the Portuguese academy dictionary (DLP). Taking advantage of the practical experience in lexicography, we want to prove that some individual ordinary observations in this field are obsolete – e.g., Kilgarriff's (1997) statement regarding how 'lexicographers write dictionaries rather than writing about writing dictionaries' (p. 102).

Thesis Structure

The content of this thesis comprises three main parts, which are divided into nine chapters, followed by the necessary concluding remarks.

The first part, 'Framework Issues', which comprises chapters 1 to 5, is dedicated to the theoretical background upon which this research rests, consisting of a review of state-of-the-art general language dictionaries, namely academy dictionaries, and an approach to the inclusion of terms therein. Chapter 1, 'Theoretical Background' (pp. 17–40), addresses the theoretical framework. Chapter 2, 'Dictionaries' (pp. 41–62), tries to deal synthetically with the dictionary concept's complexity. After an overview, Chapter 3, 'European Lexicographic Tradition' (pp. 63–93), introduces the institutions and our objects of study, i.e., academy dictionaries. Chapter 4, 'Usage Labels in General Language Dictionaries' (pp. 94–115), discusses the labelling system by analysing labelling practices in the three selected dictionaries. Chapter 5, 'Terms in General Language Dictionaries' (pp. 116–141), introduces the discussion circulating in the field around the presence of terms in general language dictionaries and then briefly addresses some of the key concepts of terminological work.

The second part, ‘Data Analysis and Processing’, consisting of chapters 6 and 7, sets out the practical work carried out. Chapter 6, ‘Coverage and Treatment of Terms in Academy Dictionaries’ (pp. 142–189), is entirely dedicated to covering and treating terms in the dictionaries under study. Chapter 7, ‘A Terminological Approach for Lexicographic Purposes’ (pp. 190–256), discusses in detail the mixed methodology applied in the Portuguese academy project (DLP). Since the comprehensive treatment of terms would be an excessively time-consuming task for the purposes of this research, our methodology has been applied to two domains only: GEOLOGY, in the more general scope of EARTH SCIENCES, and FOOTBALL, which falls within the general domain of SPORTS. We describe the domains under focus, grounding our choice and proposing an organisation for each of them. We also concentrate on the question of the term’s definitions, showing how terminological methods can improve lexicographic work.

The third and final part, ‘Encoding and Modelling Dictionaries’, consists of chapters 8 and 9 and points to the issues involved in representing and publishing the analysed lexicographic data. Chapter 8, ‘Standards for Structured Lexicographic Resources’ (pp. 258–274), discusses the formal representations and standardised models that are best known and most widely used within the lexicographic universe, focusing on the encoding of dictionaries in TEI. Chapter 9, ‘TEI Lex-0 in action’ (pp. 275–306), describes the application of this TEI subformat to the Portuguese academy dictionary’s new edition and highlights the importance of hierarchical domain labels.

PART I

FRAMEWORK ISSUES

CHAPTER 1

Theoretical Background

Terminology should provide an opportunity for progress in lexicography.

REY (1995, p. 123)

The theoretical framework of the thesis takes as its starting point the digital humanities, a field in which different branches of knowledge intersect. Among them is lexicography, which we do not see as a subdiscipline of linguistics or lexicology, but as a discipline *per se*, with its own object of study, and therefore we claim its scientificity, which comprises two components, one of a practical nature (practical lexicography) and the other theoretical (theoretical lexicography; dictionary research or metalexicography). The convergence of this discipline with others is a necessity – one would even say an imposition – and here we establish a bridge between lexicography and terminology, which we regard as a primarily interdisciplinary field and whose methodological assumptions, we argue, can be put to work in service of lexicography. We review some of the theoretical and descriptive works and the most important initiatives in the emergence and development of these two areas of the language sciences. We consulted and analysed the “lexicography” and “terminology” lexicographic articles in different general language dictionaries in order to observe how these terms are currently described/defined by lexicographers. Taking into account the elaboration of theoretical and practical principles that materialise the production of lexicographic works, we also present our theoretical position on the dual dimension of terminology, namely the conceptual and the linguistic, in which we advocate their complementarity.

1.1 The Emergence of the Digital Humanities

In the past two decades, the humanities, as an academic branch, have undergone a profound turnaround with the global rise of networked technology and especially the explosion of the so-called Web 2.0 (O’Reilly, 2005) – the second generation of web-based communities and services that have made the online environment more dynamic.

User-generated content, interoperable formats, and the possibility of crowdsourcing are now widespread. The next move forward is the much-heralded major evolution in connecting information, Web 3.0 (Markoff, 2006), an artificially intelligent web or the third generation of internet-based services, aka the Semantic Web. These changes in the technological infrastructure of our culture have led to the emergence of a new buzzword whose field is expanding and changing: digital humanities.

The term “digital humanities” was coined by Schreibman, Siemens and Unsworth (2004) with the publication of their book *A Companion to Digital Humanities* and appeared as an alternative to an array of previous designations, such as ‘humanities computing’ (Terras, Nyhan & Vahouette, 2013). Although Schreibman, Siemens & Unsworth (2004) consider the digital humanities ‘a discipline in its own right’ (p. xxiii), its status and definition are far from consensual and have become a matter of heated debate (see Gold & Klein, 2016; Alves, 2016). The struggle in defining the term arises ‘from its disciplinary and institutional diversity, and its multiple modes of engagement with information technology’ (Svensson, 2009). Within digital humanities, it is possible to find a wide variety of works from different branches of knowledge within the scope of social and human sciences, characterised by the digital use of tools, methods and standards.

A definition from *The Digital Humanities Manifesto 2.0* (2009) – the result of nine seminars held as part of the University of California, Los Angeles (UCLA) Mellon Seminar in 2008/2009 – proposes:

Digital humanities is not a unified field but an array of convergent practices [*emphasis added*] that explore a universe in which: a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences.

In turn, the signatories of the French manifesto, *L’Affiche du Manifeste* (2010), circulated at a THATCamp in Paris in May 2010, emphasise the multi and transdisciplinary nature of digital humanities:

The digital humanities designate a ‘transdiscipline’ [*emphasis added*], embodying all the methods, systems and heuristic perspectives linked to the digital within the fields of humanities and the social sciences.

This transdisciplinary nature enables digital humanities to act as a centripetal force around a set of humanistic and computational disciplines, as well as other knowledge branches, encompassing a wide range of methods and practices.

Beyond discussing whether digital humanities are a discipline⁸ in their own right (Schreibman, Siemens & Unsworth, 2004), an ‘empty buzzword’ (Fish, 2018) used for fundraising, a ‘movement’ (Holm, Jarrick & Scott, 2015) or a ‘cross-disciplinary endeavour’ (McCarty, 2015) that brings digital information technology to existing humanities disciplines, we acknowledge that it is a broad field of research and scholarly activity, which implies a new modality of research and data sharing that has particularly brought in significant epistemological and methodological challenges (Gonçalves & Banza, 2013, p. 5) as well as expanded the use of sophisticated computing techniques and digital methods, concerning the way data is produced, researched and preserved.

Currently, we are encountering a new way of conceiving the traditional field of the humanities. According to Berry and Fagerjord (2017), this reconceptualisation could be carried out by what they call the ‘digital humanities stack’ (Figure 1), which was designed to facilitate the project of critical digital humanities.

⁸ Luhmann and Burghardt (2021), analysing the role and position of digital humanities in the academic landscape, compared articles published over the past three decades in three established English-language digital humanities journals. They concluded that, in fact, digital humanities already constitute their own cluster but, at the same time, the cross-disciplinary endeavour is evident.

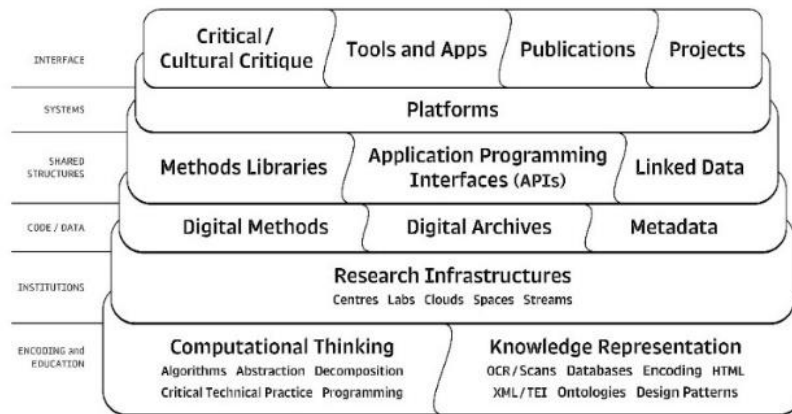


Figure 1: *The Digital Humanities Stack (Berry & Fagerjord, 2017)*

At the base of the diagram, we detect the elements of ‘computational thinking’ and ‘knowledge representation’ that are essential to our investigation as well. Berry and Fagerjord (2017) argue that ‘this type of diagram is common in computation and computer science to show how technologies are *stacked* on top of each other in increasing levels of abstraction’ (p. 28.). With this illustration, they intend to demonstrate the range of activities, practices, skills, technologies and structures that purportedly compose digital humanities, with the aim of yielding a high-level map.

Like many humanities disciplines – including literature, philosophy, history, law and musicology, among many others –, lexicography has been transformed by technological change (Wooldridge, 2004) and requires digital humanities to reformulate the access to its products – dictionaries themselves – ‘not as an object, but a service’, as Tasovac (2010) stated when arguing that ‘dictionaries do not [yet] come to us’ when we consult them from a website. The field must endeavour to achieve this reformulation within the near future.

1.2 A Walk Through the Lexicographic Universe

Traditionally, lexicography has been understood as the art and craft of compiling dictionaries or the practice of dictionary making. Despite this essentially practical strand, the discipline presents another strand, of a theoretical nature, that develops and formulates theoretical models and methodologies for compiling lexicographic works and

solving problems related to the creation of dictionaries. It is well known that lexicographic practice is much older than lexicographic theory (Gouws, 2005). Even if we can trace the origin of ‘dictionaries’⁹ back to antiquity, the truth is that it was only in the 20th century, beginning in the 1940s, that the first actual theoretical contributions to the development of lexicography emerged (Rey & Delesalle, 1979, pp. 4–5). As stated by Lino (1992), ‘*assistimos à mudança de estatuto da lexicografia que deixou de ser a arte de fazer dicionários, para designar a ciência*’ [we have seen a change in the status of lexicography that ceased to be the art of making dictionaries to designate the science] (p. 2). Eventually, this science will be recognised ‘as a field in its own right’ (Granger, 2012, p. 1).

Lexicography must be looked upon as a global phenomenon with a detailed account of lexicographic works across the world (e.g., from China: Yong & Peng, 2008; Xue, 1982; India: Vogel, 1979; Arabia: Al-Kasimi, 2019; Romania: Burada & Sinu, 2020; among others), even if this would be beyond the scope of this thesis. Nevertheless, we will outline some of the key moments in the theoretical and methodological development of the discipline, taking into account the following points: (a) the theoretical lexicographic frameworks focused on two primary, different approaches, viz. a structural and a functional approach; (b) the synthesis and relevant works displaying a certain maturity level of lexicography as a scholarly field are mentioned; (c) the advent of digital lexicography; (d) the increased disciplinary professionalisation of lexicography (conferences, journals, associations) along with the references to a few of the most recent lexicographic projects.

Concerning theoretical lexicographic frameworks, we aim to stress that reflections on the nature, structure and role of dictionaries existed even in the pre-theoretical era, i.e., before the 20th century when lexicography had no disciplinary status yet. The prefaces or introductions to legacy dictionaries – e.g., and we cite only two examples among many, *The Plan of a Dictionary of the English Language* (Johnson, 1747), Samuel Johnson’s famous lexicographic work, or *Planta para se formar o*

⁹ We use quotation marks because we are referring to the dictionary in a very wide sense; we mean Sumerian clays or Egyptian papyri, for example. There are also those who prefer to use the prefix *proto-*, that is, ‘protodictionaries’ and ‘paleolexicography’, given the great lexicographic activity in ancient civilisations.

Dicionário da língua portuguesa (ACL, 1793), the introduction of the first Portuguese dictionary of the Academia das Ciências de Lisboa – were very extensive and contained some theoretical reflections on lexicographic issues, which makes it possible for us to speak of incipient metalexicographic discourses.

To summarise this literature review, we decided to establish two major divisions, i.e., two fundamentally different ways of approaching dictionaries as research objects: between scholars who devoted themselves to structural questions about dictionaries, referring to the essential components of lexicographic works that compound their structure, and, on the other hand, those who dedicate their study more to functional issues, typologies and focusing on user needs.

In what can be considered the first steps towards the constitution of lexicographic theoretical foundations, the initial topic was a reflection on dictionary content, as well as an attempt to classify the different types of existing dictionaries. We begin by referring to the work of Lev Vladimirovich Shcherba (1880–1944), a Soviet linguist and lexicographer, whose work¹⁰ contributed abundantly to establishing lexicology and lexicography as distinct scientific disciplines; his work will be mentioned later in the section dedicated to dictionary typologies (Chapter 2) for his ground-breaking effort to classify dictionaries. In the subsequent phase, theoretical lexicographic studies focused on the identification and discussion of dictionary structures. At the time, Josette Rey-Debove (1929–2005), lexicographer, introduced the concepts of macrostructure and microstructure (Rey-Debove, 1971, p. 21). With his pioneering studies on dictionary structure, the French lexicographer Jean Dubois (1920–2015) argued that the dictionary could be approached as a communicative text or discourse (Dubois, 1962). Thus, the initial notions of macrostructure and microstructure gave rise to other metalexicographic distinctions related to the different dictionary components and structures (Hausmann & Wiegand, 1989; Wiegand, 1989a; 1989b; Bergenholtz & Tarp, 2003), including the access structure, data distribution structure,

¹⁰ We refer to *Opyt obshchei teorii leksikografii* (Shcherba, 1940/1995), a speech given in an academic session in 1939 and published in the magazine of the Russian Academy of Sciences in 1940.

frame structure, macrostructure, microstructure, mediostructure and addressing structure.¹¹

Concerning a functional approach, the Aarhus School of Business (Aarhus University), in Denmark, formulated what they called the ‘theory of lexicographical functions’. Henning Bergenholtz and Sven Tarp contended that more than describing the lexicon of languages, lexicography aims to solve specific types of information needs detected in society. They proposed a new theory, which is still prevalent today (Bergenholtz, Nielsen & Tarp, 2009), focusing on dictionary functions, i.e., those related to communication (such as text reception, text production, proofreading, text editing and translation, all of which are dependent on the text) and those related to knowledge or cognition (how to obtain general knowledge).

Moving forward, we want to highlight some syntheses and relevant works that illustrate a certain level of maturity of lexicography as a scholarly field, additionally focusing on the discourse of some of its main proponents – respected references in lexicographic circles today – who have been approaching the discipline from a theoretical or methodological perspective.

Van Sterkenburg (2003) considers Ladislav Zgusta (1924–2007), Czech-American historical linguist and lexicographer who published the first international lexicography textbook in 1971, ‘the twentieth-century godfather of lexicography’ (p. 4). According to him, Zgusta dominated the field of lexicography in the 1970s and 1980s.

Sidney Landau (1933–present) is, in turn, the great authority on American lexicography. His book *Dictionaries: The Art and Craft of Lexicography* (Landau, 2001), first published in 1984, offers a comprehensive overview of English lexicography. Hartmann (2003), for example, states that this book has been a *vademecum* for himself and his students for many years. The second edition, published in 2001, is still available on the market today besides being a subject of research. This was to be followed by another textbook, still frequently referenced today: Bo Svensén’s *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making* (Svensén, 2009), whose

¹¹ We will discuss these concepts in Chapter 2.

first edition was published in Swedish in 1987 and was subsequently translated into English in 1993.

At the dawn of the 21st century, new introductory handbooks and charts appeared on the desk of many lexicographers worldwide, as is the case of B. T. Sue Atkins and Michael Rundell's *The Oxford Guide to Practical Lexicography* (Atkins & Rundell, 2008), which details how commercial dictionaries for monolingual and bilingual learners were compiled in the 2000s.

Also worthy of mention are the *Dictionnaires: An International Encyclopedia of Lexicography* (Hausmann et al., 1989–1991), published in three volumes, and the *Dictionary of Lexicography* (Hartmann & James, 1998/2002). In the 21st century, Gouws et al. (2014) published a supplementary volume to the *Encyclopedia* publication to account for recent developments, focusing on electronic and computational lexicography, and a new volume of the *Dictionary of Lexicography and Dictionary Research* (Wiegand et al., 2020) was launched.

Finally, and of great interest to the topic of this thesis, the work of John Considine, especially the 2014 publication, *Academy Dictionaries 1600–1800* (Considine, 2014), which traces the history of lexicography on a European scale, discusses the numerous dictionaries compiled by various national academies in the 17th and 18th centuries. In particular, for each of the case studies in this thesis, we can also quote the volume *Le Dictionnaire de l'Académie française: Langue, littérature, société* (Carrère d'Encausse et al., 2017), *La Real Academia Española – Vida e historia* (García de la Concha, 2014) and in the Portuguese case, the academy works, for example, of Casteleiro (1981) and Verdelho (2007).



As our research topic revolves around three dictionaries of different languages, we briefly inspect some of the lexicographic studies developed in these three countries, namely France, Spain and Portugal.

Quemada (1926–2018), one of the pioneers of French lexicography in the 20th century, made a profound mark on lexicological research and lexicography worldwide. His thesis, *Les dictionnaires du français moderne, 1539–1863: Étude sur leur histoire*,

leurs types et leurs méthodes (Quemada, 1968), revolutionised the understanding of lexicography. He was the director of the *Trésor de la langue française*, published in 16 volumes, and the director of the publication *Cahiers de lexicologie*, started in 1959. Quemada (1987, p. 229) also introduced a new concept, referring to the dictionary as an object, that of *dictionarique*, which is used to designate the field of the production of dictionaries, while lexicography would entail the collection activity and study of lexical data. The works of Quemada and Jean Pruvost (1949–present), which are dedicated to the prefaces of the first eight editions of the DAF (Quemada, 1997), were also invaluable to this research. Apart from his work, Pruvost is known for being the organiser and creator of *Journée des dictionnaires*.¹²

Later on, in the 1990s, the contributions of Collinot and Mazière (1997), in *Un prêt à parler: le dictionnaire*, with their works in discourse analysis, are also referred to in this thesis. Last but not least, Alain Rey (1928–2020), ‘Monsieur Dictionnaire’, was the editor-in-chief at the French dictionary publisher Dictionnaires Le Robert and enjoyed the status of a French media personality, where he presented an entertaining examination of French vocabulary. Many of his works (Rey, 1970; 1979; 1983; 1985; 1989; 1995; 2008) will be referred to throughout this research.

In Spain, one of the first metalexicographic works is Casares’ *Introducción a la lexicografía moderna* (Casares, 1982), which captures our interest chiefly due to how it addresses the academic dictionary. Another reference work includes Günther Haensch’s *Los diccionarios del español en el umbral del siglo XXI* (Haensch, 1997). Additionally, we insert a reference to Porto Dapena’s (2002) book *Manual de Técnica Lexicográfica*.

In Portugal, recent scientific activity around lexicographic work have been presented by Costa, Salgado et al. (2021), Villalva and Williams (2019), Salgado, Costa and Tasovac (2019), Salgado and Costa (2019a), Lino (2018), Silvestre (2008; 2016), Iriarte Sanromán (2015, 2001), Gonçalves and Banza (2013), Correia (2008; 2009), and Verdelho (1994; 1998; 2002; 2007), to cite a few.



¹² <https://www.jeanpruvost.com/journ%C3%A9e-des-dictionnaires>

Concerning the advent of digital lexicography, although many dictionaries were still published on paper in the 2000s, the scenario has changed dramatically in the last decade with the definitive transition to digital platforms. In the first decade of this century, the first publications entirely devoted to this topic or seeking to make it one of their main focuses began to appear (Fuertes-Olivera & Bergenholtz, 2011; Fuertes-Olivera & Tarp, 2008; Gouws, 2011). Although computerised lexicography took its first steps in the late 1950s and early 1960s (Granger, 2012), the computers' capabilities at the time did not allow the complete compilation and editing of an entire lexicographic work. However, they were (are) undoubtedly invaluable for any lexicographer tasked with the compilation, systematisation and control of data. The lexicography landscape has changed, and technological advances have been dictating new strategies and directions. Space restrictions are no longer a concern (Lew, 2011), and the integration of corpora (Rundell, 2019) and development of various dictionary writing systems (Abel, 2012) became a requirement in the daily life of a lexicographer.

The 21st century is witnessing a profound shift in the territory of lexicography. First, the introduction of big data (available electronic corpora) with a lot of relevant lexicographic data has bloated the printed dictionaries 'almost to the point of impracticality' (Rundell, 2010, p. 170). Second, as the availability of free digital versions of dictionaries started to increase, dictionary sales declined significantly, which has led, among other things, to a reduction in the number of hired lexicographers and the downfall – or, at least, changes to the business models – of many renowned publishers (Rundell 2010, p. 170).¹³

Dictionaries have become 'digital assistants', as Nielsen (2013), who sees dictionaries as information tools to satisfy specific types of user needs, suggests. Although the terms "electronic" and "e-dictionary" continue to be used copiously by the

¹³ In Portugal, for example, the children's dictionary for school-age groups is one of the few dictionaries that continue to be published on paper, given the need for consultation in the classroom. Apart from this, paper-based dictionary releases have been very sporadic (see, for example, the new edition of *Dicionário da Língua Portuguesa – Léxico, Gramática e Prontuário* by Aldina Vaza and Emília Amor, published by Texto in 2018, or the *Dicionário Gramatical de Verbos do Português* by Jorge Baptista and Nuno J. Mamede, published in 2020 by the Universidade do Algarve).

lexicographic community¹⁴, we fail to make any distinction between these terms and “digital dictionary”, particularly because electronic dictionaries are no longer published.



The collective will and effort to create a scientific forum for discussion and foster the exchange and sharing of interdisciplinary knowledge has borne much fruit. Moreover, the numerous interdisciplinary conferences, initiatives, actions and projects on lexicography must be mentioned.

In 1957, the first congress on lexicography was held in Strasbourg (*Lexicologie et lexicographie françaises et romanes*). Another example is the biennial eLex conference¹⁵, which opened in 2009 in Louvain, Belgium, and the Dictionary Society of North America, which also acts as editor of the journal. In the late 1980s, the *International Journal of Lexicography*¹⁶ was launched by the European Association for Lexicography (EURALEX) under the initial direction of Robert Ilson and the current direction of Robert Lew.



A final list of the projects that propelled lexicography to prominence within the humanities includes the H2020 ELEXIS EU funded project¹⁷, already mentioned in the Introduction, and in which NOVA CLUNL (Linguistics Research Centre of NOVA University Lisbon) is actively participating; the European Network of Lexicography¹⁸; DARIAH, namely the Working Group on Lexical Resources¹⁹, to which we have contributed to the definition of the TEI Lex-0; the COST NexusLinguarum²⁰; and a series of projects, some

¹⁴ Perhaps due to a professional bias, we have always associated ‘electronic dictionaries’ with the publication of dictionaries in the CD-ROM or pen-drive version.

¹⁵ <https://elex.link/>

¹⁶ <https://academic.oup.com/ijl>

¹⁷ <https://www.elex.is>

¹⁸ <https://www.elexicography.eu>

¹⁹ <https://www.dariah.eu/activities/working-groups/lexical-resources/>

²⁰ <https://nexuslinguarum.eu/>

finished and some in progress, including BASNUM²¹, Nénufar²², ARTFL²³, VICAV²⁴, and MORDigital²⁵.

1.3 The Twofold Nature of Lexicography

Wiegand et al. (2020) quite recently proposed a broader definition of lexicography: ‘total of all activities directed at the preparation of a lexicographic reference work’ (p. 224). It is assumed that these activities, related to the elaboration of a wide variety of resources – dictionaries, vocabularies, glossaries, encyclopaedias, etc. –, necessarily possess a theoretical and practical component, a point that the entire lexicographic community seems to agree on.

The field of lexicography has a twofold nature: (1) a practical element, called *practical lexicography*, which refers to the planning and compilation of actual dictionaries; and (2) a theoretical element, called *theoretical lexicography* or *dictionary research* (Hartmann, 1998/2002) or *metalexicography* (Rey-Debove, 1971; Wooldridge, 1977; Rey & Delesalle, 1979), which deals with the theoretical discussion of the content of dictionaries and can be descriptive, critical or historical. Metalexicography also examines existing dictionaries, focusing predominantly on complex topics, such as the definition of a typology, including a pragmatic dimension concerning usage. Simply put, a lexicographer is someone who produces dictionaries; when speaking and writing about them, that someone becomes a metalexicographer. In any case, although the term metalexicography was only coined in the 1970s, it should be noted that ‘*existiu sempre uma certa tradição teórica, mais em forma de análise ou apreciação crítica de um produto terminado*’ [there has always been a certain theoretical tradition, more in the

²¹ <https://anr.fr/Project-ANR-18-CE38-0003>

²² <https://nenufar.huma-num.fr/>

²³ <https://artfl-project.uchicago.edu/>

²⁴ <https://www.oeaw.ac.at/acdh/projects/vicav>

²⁵ MORDigital – Digitisation of *Dicionário da Língua Portuguesa* by António de Morais Silva [PTDC/LLT-LIN/6841/2020]. The intention is to make these dictionaries available in both TEI-XML and linked data. We advocate a holistic approach in which the field of lexicography intersects with terminology and many other disciplines, such as information science (Costa et al., 2021b). Recently, regarding another project, *Digital Edition of the Vocabulário Ortográfico da Língua Portuguesa* (VOLP-1940), we wrote a book chapter in which we mentioned the advantages of interdisciplinarity between information science and lexicography; see Costa, Salgado & Almeida, 2021a.

form of analysis or critical appreciation of a finished product] (Iriarte Sanromán, 2001, p. 51).

Considerations about whether lexicography is a science have been widely debated (Shcherba, 1940/1995; Zgusta, 1971; Wiegand, 1984; Hausmann & Wiegand, 1989; Lew, 2007; Tarp, 2008; Bogaards, 2010; Bergenholtz & Gouws, 2012; Ilson, 2012; Rundell, 2012; Piotrowski, 2013; Adamska-Sałaciak, 2019). Ilson (2012) presents the problematic question as follows:

Between them, the academics, professional lexicographers, and computerniks provided a round view of lexicography as a whole. The problem was, however, that each group had on its own a limited view of the subject. The academics had their Ideas; the computerniks, their Algorithms. But too often, alas, they seemed to lack detailed knowledge of what dictionaries are actually like and how dictionaries are actually produced. On the other hand, the professional lexicographers seemed often to lack detailed knowledge of linguistics; and their superbly detailed knowledge of Really Existing Dictionaries seemed often to be limited to those they had actually worked on... but lexicographers have scant time or incentive to contribute to learned journals: after all, they have dictionary deadlines to meet.

Furthermore, when examining the relationship between lexicography and linguistics, Béjoint (2000, pp. 169–208) draws attention to the same fact that many lexicographers have little training in linguistics and little knowledge of how dictionaries are compiled. We recognise that in many situations, this is what happens. However, as lexicographers, we argue that the lexicographic practice obeys scientifically rigorous methodology and principles (Margalitadze, 2018), and a prior theoretical linguistic reflection on the criteria must be made, not solely based on the lexicographer's 'intuition' (Correia, 2008, p. 9). For his part, Rundell (2012) fears that theoretical lexicography in its present form is unlikely to offer a perspective on what a dictionary does, while Piotrowski (2013) argues – in our view, justifiably – that we need new appropriate theoretical perspectives to determine how to deal with the current situation in which dictionaries undergo radical changes, becoming abstract objects in virtual space – 'the dictionary of the future will not be perceived as an object at all, it will work like a background process' (Piotrowski, 2013, p. 317).

Given the different points of view on the status of lexicography, we need to take a stand. Some argue that lexicography is a branch of applied linguistics (Rey, 1995, p. 113; Meier, 1969; Villers, 2006), while others consider it an independent discipline (Wiegand, 1984; Granger, 2012). For almost the entirety of the 20th century, linguistics has believed lexicography to be the art or craft of making dictionaries, questioning its controversial scientific status. In fact, some scholars insist that lexicographic theory does not exist (Béjoint, 2000, p. 381; Atkins & Rundell, 2008, p. 4). Leroyer (2011), in turn, defines lexicography as part of the social and information sciences that is mainly concerned with the development, planning and publication of electronic reference tools. However – and although lexicography also involves data, information, knowledge and ‘there are a number of commonalities between information science and lexicography’ (Bothma, 2017, p. 198) – we do not consider lexicography to be a subdiscipline of the information sciences, despite the intersection being very advantageous (Costa, Salgado & Almeida, 2021a).

As lexicography is concerned with the development of theoretical and practical principles and the production of lexicographic tools, several disciplines are involved in any dictionary project (Nielsen, 2018). In short, we argue that lexicography should be seen as a discipline in and of itself, with its own object of study: the dictionary.

Alongside lexicography is lexicology, and opinions have always differed regarding the relationship between these two disciplines. We understand lexicology as the science that analyses the lexicon of a specific language – including formation, spelling, origin, usage, semantic relations and definition. Lexicography also studies the lexicon as lexicology does but ‘whereas lexicology concentrates more on general properties and features that can be viewed as systematic, lexicography typically has the so to say individuality of each lexical unit in the focus of its interest’ (Zgusta 1971, p. 14). Corresponding to Wiegand (1984, pp. 13–15), we see lexicology as an autonomous discipline because although it deals with the lexicon’s study, both disciplines have different methods and purposes. While a lexicographer is concerned strictly with the inclusion and treatment of lexical units in dictionaries, a lexicologist is concerned with diachronic aspects – such as the etymology of the words or morphological features – and synchronic aspects – for example, their present meaning and usage. Ideally, all

lexicographers are lexicologists but not the other way round. While lexicology investigates the lexicon as a research object *per se*, lexicography pursues a much more practical aim: to represent the meaning of words in order to compile dictionaries.



We have just presented how the paths that define lexicography are intricate. In the article ‘What is Lexicography?’, Bergenholtz and Gouws (2012, pp. 32–35) attest to the different interpretations of what is meant by lexicography, collecting and analysing definitions extracted from different lexicographic works, whether general or specialised language dictionaries and scientific publications. Assuming the information conveyed by the general dictionaries is relevant, we resolved to conduct the same exercise, enlisting the academic lexicographic corpus of the present thesis. We therefore decided to consult the “lexicography” article in each of the three previously referenced academic resources (DAF, DLE, DLPC), as seen in Figures 2, 3 and 4.



Figure 2: Definition 1 – Entry ‘lexicographie’ [lexicography] in the DAF (AF)

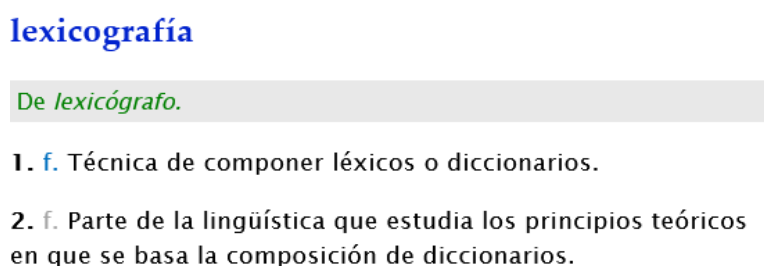


Figure 3: Definition 2 – Entry ‘lexicografía’ [lexicography] in the DLE (RAE)

lexicografia [lɛksikugɾɐfjɐ]. *s. f.* (De *léxico* + *-grafia*). *Ling.* Ramo da linguística que se ocupa dos aspectos teóricos e práticos que têm em vista a elaboração de dicionários, vocabulários, glossários. *Hoje a lexicografia recorre com frequência às tecnologias informáticas.*

Figure 4: Entry 'lexicografia' [lexicography] in the DLPC (ACL)

At first glance, we can see that all the entries focus on different points but none turn out to be satisfactory. In Definition 1 (Figure 2), lexicography is understood as a '*science et technique*' [science and technique]. In Definition 2 (Figure 3), as a '*técnica*' [technique], which seems to deny the status of science, although there seems to be a clear intention to distinguish the practical lexicographic component from the theoretical one, with the division in two senses. However, sense 2 of Definition 2, which refers to the more theoretical character of lexicography, such as sense 1 of Definition 3 (Figure 4), sees lexicography as a branch of linguistics. All of the definitions above are reductive, limited to composition and elaboration, without any reference to the function, structure or use of dictionaries.

This small exercise leads us to the conclusion that, in fact, the concept of lexicography is controversial and somewhat confusing, as it seems that lexicographers themselves interpret it differently. What may surprise us most is that while we are aware that there are defenders of different theories, the theoretical and practical components of lexicography have been universally recognised and it has gained its independence from linguistic fields, which somehow has not been reflected in the consulted definitions.

In summary, the theoretical and practical components of lexicography could be represented in the following scheme (Figure 5) that was inspired by and adapted from Hartmann and James (1998/2002, p. 86) and Bergenholtz and Gouws (2012, p. 40).

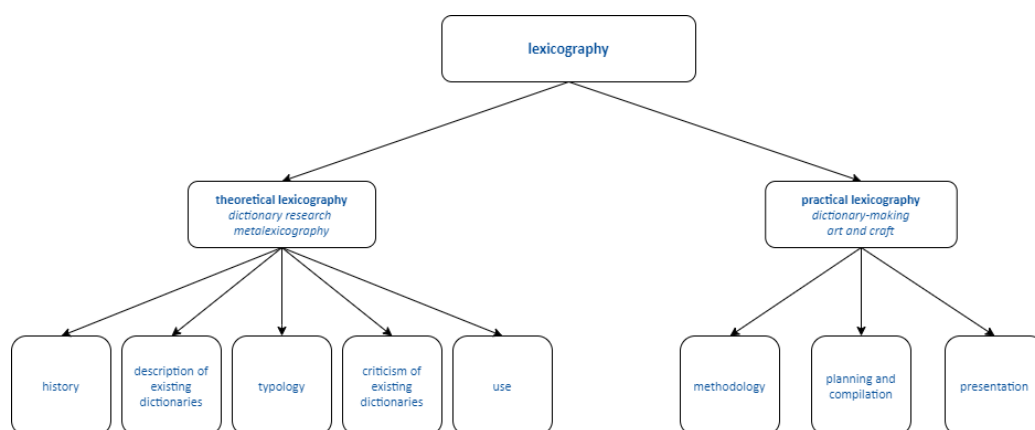


Figure 5: *The Theoretical and Practical Components of Lexicography*

In this sense, regarding dictionaries in general, and recalling the definitions from the exercise above, it will be necessary to consider that the two lexicographic components should ideally be acted upon, combining the two aspects.²⁶

1.4 Terminology as an Interdisciplinary Field

Concerning terminology, we recognise its statute of autonomous science and aim to emphasise its interdisciplinary and transdisciplinary nature (Felber, 1987, p. 1). As terminology is a polysemic term, we decided to perform the same exercise as we did in the previous section and consulted the “terminology” lexicographic article in each of the academy resources (DAF, DLE, DLPC) to verify how lexicographers have defined this term. The searches are presented in Figures 6, 7 and 8.

8^e ÉDITION **TERMINOLOGIE.** n. f.

- T. didactique. Ensemble des termes techniques d’une science ou d’un art.

Figure 6: *Definition 1 – Entry ‘terminologie’ [terminology] in the DAF (AF)*

²⁶ Since the old editorial deadlines (which were short because they were strictly commercial in nature and often prevented best practices of work planning) no longer make sense, as dictionaries are no longer a commercial investment, today, this alliance between theory and practice seems more achievable.

terminología

1. f. Conjunto de términos o vocablos propios de determinada profesión, ciencia o materia.
2. f. Estudio de la **terminología**.

Figure 7: Definition 2 – Entry ‘terminología’ [terminology] in the DLE (RAE)

terminologia [tɨrminuluʒiɐ]. *s. f.* (Do lat. *terminus* ‘termo’ + suf. *-logia*). **1.** Estudo dos termos técnicos de uma ciência, de uma arte, de uma disciplina. *Criou-se uma secção de terminologia, imprescindível ao trabalho lexicográfico.* **2.** Conjunto organizado desses termos. *Terminologia gramatical. Terminologia informática.* **3.** Conjunto dos termos peculiares a um escritor, a um autor. *«uma ou outra pessoa, por motivos políticos, ficou de tal maneira sensibilizada que, utilizando um determinado tipo de terminologia, passou a adoptar atitudes de agressividade e de menos respeito por ele mesmo» (DN, 22.10.2000).*

Figure 8: Definition 3 – Entry ‘terminologia’ [terminology] in the DLPC (ACL)

The first caveat we must raise is the DAF consultation (Figure 6). The consulted article still corresponds to the eighth edition since the last one is only available up to the letter ‘s’. What strikes us the most about this entry is the label ‘T. didactique’. What do lexicographers want to mark with this label? What should be understood as a didactic term? Without explanatory introductions, we can only question its employment.

After comparing the three definitions collected, a point that concerns us is that none of the dictionaries define terminology as a science or a domain of interdisciplinary knowledge. DLE and DLPC seem to approach this sense when referring to ‘*Estudio de la terminología*’ [terminology study] (Figure 7) and ‘*Estudo dos termos técnicos*’ [study of technical terms] (Figure 8), but the descriptions are too vague to draw accurate conclusions. The three dictionaries coincide in defining terminology as a set of terms, the meaning we referred to in the Introduction and one that leads us to speak of terminologies.

Having been unsuccessful in obtaining a satisfactory answer, we decided to consult an English dictionary. For this, we chose the *Oxford English Dictionary* (OED).²⁷

The screenshot shows the OED entry for 'terminology, n.'. At the top, the word is in a large red font. To the right, there's a 'Text size' control with 'A' and 'A' buttons. Below the word, there are links for 'View as: Outline | Full entry', 'Quotations: Show all | Hide all', and 'Keywords: On | Off'. The entry includes a 'Pronunciation' section with British and American phonetic transcriptions and audio icons. A 'Frequency (in current use)' section shows a series of colored dots. The 'Origin' section states it's a borrowing from French. The 'Etymology' section details its roots in French and Latin. A paragraph explains its apparent etymon and its post-classical Latin sense. A 'Compare' section mentions German 'Terminologie'. A 'Thesaurus' link is present. The definition is split into two parts: 'The system of terms...' and 'Also: the scientific study...'. At the bottom, the years '1770–2008' and a '(Show quotations)' link are visible.

terminology, n. Text size: A A

View as: Outline | [Full entry](#) Quotations: Show all | [Hide all](#) Keywords: On | [Off](#)

Pronunciation: ² Brit. [▶](#) /təːmɪˈnɒlədʒi/, U.S. [▶](#) /tərməˈnələdʒi/

Frequency (in current use): ●●●●●●●●

Origin: A borrowing from French. **Etymons:** French *terminologie*, Latin *terminologia*.

Etymology: < (i) French *terminologie* doctrine or scientific study of terms (1720), technical terms collectively (1755),

and its apparent etymon (ii) post-classical Latin *terminologia* (1702) < classical Latin *terminus* **TERMINUS** *n.* in its post-classical Latin sense ‘word, expression’ (see **TERM** *n.*) + *-logia* **-LOGY** *comb. form.*

Compare German *Terminologie* (1715 or earlier). ([Show Less](#))

The system of terms belonging to any science or subject; technical terms collectively; nomenclature. Also: the scientific study of the proper use of terms.

1770–2008 ([Show quotations](#)) [Thesaurus »](#)

Figure 9: Definition 4 – Entry ‘terminology’ in the OED, Oxford University Press

In Figure 9, we can see that the OED defines terminology as ‘the system of terms’ but also adds another meaning: ‘the scientific study of the proper use of terms’. Compared to the three academy dictionaries, the OED adds ‘scientific’, but there still seems to be some hesitation in accepting terminology as science.

Unfortunately, as we saw for the case of lexicography, these entries also require a revision to include a reference to three different meanings: terminology as (1) a theory or a science that explains the relationships between concepts and terms; (2) the vocabulary of a particular subject field; and (3) also as the set of practices and methods concerned with the collection, description, processing and presentation of terms.

²⁷ <https://www.oed.com/view/Entry/199439?redirectedFrom=terminology>

Terminology has established itself as a science – with Eugen Wüster (1899–1977) as we shall see next – but its tradition is already long. According to Rey (1995, pp. 17–22), the development of terminology spans over three distinct periods:

- 1) 17th and 18th centuries, the classical period in Western Europe, which is essentially characterised by reflections on knowledge, a new awareness of technical progress and a universal pedagogical attitude;
- 2) 19th century, characterised by how technical-scientific development, linguistic interventionism in socio-linguistic terms and the need for new designations were multiplied due to the advancement in the fields of science;
- 3) 19th and 20th centuries, characterised by profound transformations at economic, social and political levels with a big impact on knowledge, demanding more effective responses from terminology.

Completing Rey's proposal, Cabré (1999, p. 5) establishes four periods inherent to the development of modern terminology:

- 1) between 1930 and 1960, which corresponds to the origins;
- 2) between 1960 and 1975, concerning the structuring of the terminological field and the definition of theoretical knowledge assumptions;
- 3) between 1975 and 1985, the period of prosperity;
- 4) from 1985 to the present, which marks its expansion.

Wüster, an Austrian engineer, is considered the founder of the Vienna School and the General Theory of Terminology. Intending to eliminate ambiguity in technical and scientific discourses and transform them into an effective instrument, Wüster ended up being a pioneer in defining the concept of standardisation and, most notably the Technical Committee 37 'Terminology' of the International Organisation for Standardisation (ISO)²⁸ and Infoterm²⁹ (cf. Sager, 2004, p. 298). Its methodology was, in fact, revolutionary. For instance, in his dictionary, *The Machine Tool* (Wüster, 1968), terms representing concepts are organised according to the Universal Decimal Classification – he follows an onomasiological approach starting with the concept. For

²⁸ <https://www.iso.org/home.html>

²⁹ http://www.infoterm.info/about_us/history_of_infoterm.php

Wüster (1979/1998), all the concepts of a specific subject field should be organised into a hierarchical concept system.

The Vienna School perceives terminology as an autonomous field at the service of other disciplines, such as linguistics, logic, ontology, information science, computer science or philosophy (cf. Cabré, 1999; Sager, 1990). Indeed, in terminology, different theoretical and methodological perspectives coexist due to multiple factors.

There are several contradicting theoretical perspectives that emerged in an attempt to fill the gaps³⁰ of the Wüsterian theory. This includes the general theory of terminology, which underlies a prescriptive and onomasiological perspective concerning the relationship between concept and term (Wüster, 1979/1998). In the recognition of its interdisciplinary nature and multidimensionality, Sager (1990) identifies three dimensions that are crucial for terminology: the linguistic, cognitive and communicative dimensions. Another perspective is the communicative theory of terminology (Cabré, 1999, 2003), which focuses on a semasiological approach to terminological units. Other theoretical and methodological perspectives include socioterminology (Gaudin, 1990, 2007), a socio-cognitive model (Temmerman, 2000), and lexico-semantic theory and textual terminology (L'Homme, 2004), in which terms are studied in their linguistic environment to identify their lexical properties and behaviours, particularly in relation to other lexical items with which they co-occur in corpora. The terminology based on frame semantics (Faber, 2015) and ontoterminology (Roche et al., 2009; Roche, 2012) are among others perspectives that advocate syncretic approaches (Costa, 2013; Santos & Costa, 2015).

In summary, some terminologists follow a conceptual approach based on Wüster's doctrine. By analysing how the scientific community behaves in discourse, they conceptually model the domain and subsequently identify the terms that refer to previously defined concepts. Conversely, in a linguistic approach, the starting point is the term. The linguistic-communicative proposal (Cabré, 1999) fits into the perspective of studying the terms from a linguistic point of view, viewing them as lexical units that serve specialised communication. Cabré (1995) criticised Wüster's theory, calling it an

³⁰ A summary of the main criticisms made by several scholars concerning the classical theory of Wüster can be consulted in Santos, 2010, pp. 79–80.

‘idealised theory of terms’ (p. 14). According to Cabré (2003, p. 186), the terminological units represent units of knowledge, language and communication – the so-called ‘Theory of Doors’. She establishes three different doors (dimensions) – the cognitive (the concept), linguistic (the term) and sociocommunicative (the situation) aspects – to gain access to the terminological unit.

The interdisciplinarity of terminology, distinguished by its ‘plurality of theoretical approaches’ (Costa, 2006b), enables the establishment of a strong synergy between what is conceptual and what is linguistic.

Throughout this research project, we analysed the terms anchored in the double dimension (Costa, 2013; Santos & Costa, 2015; Roche, 2015) of terminology. We aimed to articulate the conceptual perspective (the knowledge organisation focused on the identification of concepts of specific subject fields and on the relations drawn between them) with the linguistic perspective (focusing on the terms themselves to better describe them). We hence foresee and will demonstrate (Chapter 7) that there are advantages to working on the relationship between concept and term in lexicography when the topic deals with terminologies.



We now move on to establish a bridge between lexicography and terminology. Lexicography is conceived as a field that deals chiefly with lexical units (words) but also with specialised lexical units (terms). Although lexicography and terminology are two different scientific disciplines with distinct theoretical-epistemological backgrounds, they have in common the fact that both collect data about the lexicon of a language and deal with terms, however, more often than not, with different aims. This means that working in terminology and lexicography requires individual approaches since the social, cultural or economic purposes are not the same.

Lexicographers follow mostly a semasiological perspective (from words to meanings), and terminologists (mostly concept-oriented) combine conceptual organisation and linguistic analysis where the definition of the concept is central to the view of reducing linguistic ambiguities. The difference between semasiology and onomasiology is in the perspective from which the relationship between a lexical unit

and its meaning is examined (Cabré, 1999, pp. 7–8; Sager, 1990, p. 56; Temmerman, 2000, pp. 4–5). Rey (1995, pp. 119–120) presents the question as follows:

The relationship between terminology and lexicography is, thus, obvious and very old because the objects of description are largely analogous or identical. But the designatory system of a field of knowledge or activities, i.e., the conceptual domain implied by the designatory system, is the specific object of terminology, whereas lexicography concerns itself with the functions and the behaviour of words in society, which is quite another matter.

Thus, the object of study is the same, but the angles differ. Rey (1995) follows this by declaring that both disciplines interact with each other. Costa (2013) states that terminology and lexicography should be seen as complementary regarding the methods they use. Bowker (2017), arguing for the relation between these fields, finds advantages in the fact that ‘lexicographers and terminologists continue to work together to tackle new challenges and embrace new opportunities’ (p. 149).

Lexicographers do not systematically organise specialised knowledge, generally obeying only criteria such as alphabetical ordering and the linguistic uses of lexical units in society. The lexicographer ‘collects *all* the lexical units of a language in order to sort them in various ways. Once he has collected these units, he proceeds to differentiate them by their meanings’. In turn, the terminologist ‘starts out from a much narrower position; he is only interested in subsets of the lexicon, which constitute the vocabulary (or lexicon) of special languages’ (Sager, 1990, p. 55).

Getting to know the domain and subsequently organising it are two requisite activities for a rapid and systematic identification of the basic concepts, which will result in a better description of the terminologies. Bearing in mind that we are working with specialised areas of knowledge, the intervention of the expert is necessary to aid in the task of categorising knowledge and validating the descriptions and definitions of terms. This facilitates a more accurate encoding by allowing a tidier classification of the data depending on each element.

The following schema (Figure 10) seeks to systematise the perspective adopted in this thesis and sums up what we consider to be the main specificities of lexicography versus terminology.

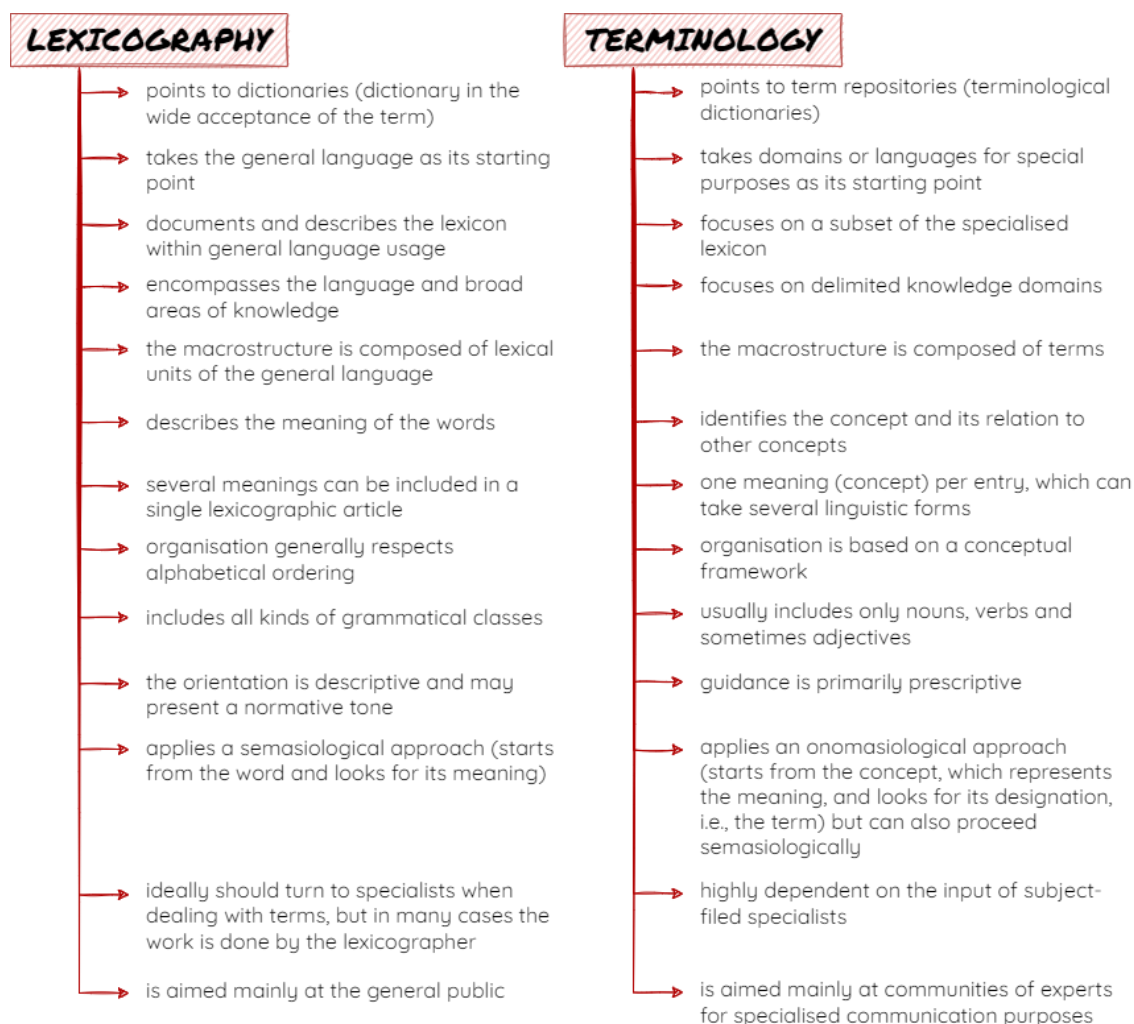


Figure 10: *Lexicography vs Terminology*

The establishment of these differences does not mean that lexicography and terminology are in opposition. Contrarily, we consider the two disciplines to complement each other as we intend to prove in the following chapters.

On the other hand, in an era where the computational component becomes a requirement in the curriculum of any lexicographer, it is time to create synergies between the linguistic and computational communities, putting an end to '*uma espécie de Guerra Santa*' [a kind of Holy War] (Simões, 2014, p. 359) that seems to exist between

their different members. We are aware of the importance of computational methods and argue that a prior and rigorous linguistic analysis of all lexicographic components is desirable. It is not acceptable for the humanities to be in the background when they are the central object under analysis. We must find a balance between the humanities and computing. The perspective that we propose here presupposes a rethinking of the lexicographic tradition's methodologies concerning the treatment of terms, perceiving lexicography and terminology as autonomous disciplines that can be found in the broad field of digital humanities.

CHAPTER 2

Dictionaries

Dictionary is a powerful word.

LANDAU (2001, p. 6)

This chapter introduces the object of study: dictionaries. The status and concept of dictionary have significantly evolved over the past few years in the way information is produced, researched, published, disseminated, preserved and shared. While digitisation has led to a paradigm shift, the spread of the Web gave shape to new frontiers in lexicography. These changes have impacted traditional dictionaries as well. We explore the concept of dictionary in its various facets: as a text, research object, cultural artefact, tool, language model. Then, based on the literature review, the next subchapter describes some of the attempts to build organised classifications since there is no standard, agreed-upon way to classify the existing types of dictionaries. We delimit the category that falls within the scope of our research and we end with the presentation of a classification proposal. Finally, we present and elucidate the basic operational concepts (macrostructure, microstructure, megastructure) essential to the development of subsequent discussions. The integration of lexicography and digital humanities should facilitate the creation of common standards for the harmonisation of policies and practices that improve the interoperability³¹ between a wide range of resources.

2.1 Dictionaries are Like Diamonds

Dictionaries have a multifaceted and undefined nature (Béjoint, 2000, p. 32). Classic visions of them, in comparison with encyclopaedias – such as that of Landau (2001), who states that a ‘dictionary is a text that describes the meanings of words, often

³¹ The ISO/IEC 2382 (2015) standard defines interoperability as the ‘capability to communicate, execute programs or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units’, <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-1:ed-3:v1:en>.

illustrates how they are used in context and usually indicates how they are pronounced’ (p. 6) – are reductive. As previously mentioned in the Introduction, any dictionary is more than a simple work, book or lexical resource containing a list of words that users look up to discover their meaning(s).

A dictionary can be several things simultaneously, and hence, a direct correlation can be drawn between the concept of a dictionary and the notion of multifunctionality. Humbley (2002) emphasises the evolution of the relationship between the dictionary and its users, not only because it constitutes a technical evolution but also because it results from a change in perspective. The author adopts a comprehensive definition of what a dictionary is, even though *‘il nous semble que cet abus de langage apparent est le prix à payer pour l’innovation, car non seulement le dictionnaire de demain ne ressemblera pas à celui d’hier, mais en plus il sera multiforme’* [this apparent abuse of language seems to be the price to pay for innovation because not only will the dictionary of tomorrow not resemble that of yesterday, it will also be multifaceted] (Humbley, 2002, p. 95).

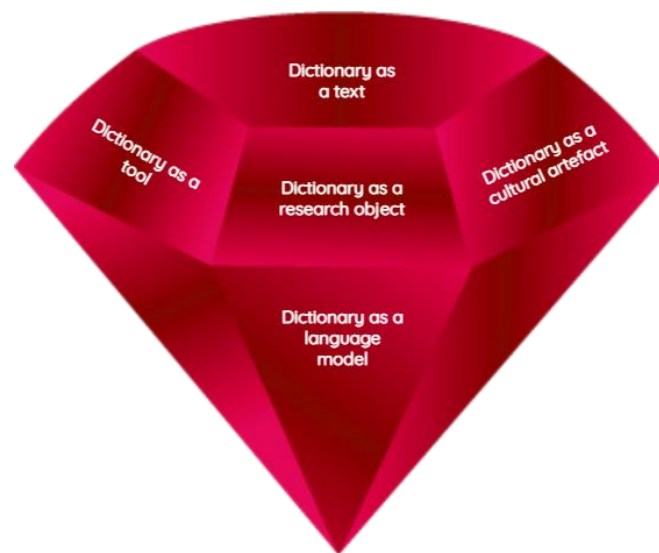


Figure 11: Dictionary seen as a diamond with multiple facets

Dictionaries as reference works can be considered diamonds. Like a diamond – one of the most precious gems – dictionaries are precious and multifaceted (Figure 11), possessing several distinct facets or features.

The following section explores the various facets of the dictionary, which must be repeatedly polished to shine. As such, we will now look at the dictionary as a text, research object, cultural artefact, tool, and finally as a language model.

2.1.1 The Dictionary as a Text

The first lexicographic works originated as clay tablets in cuneiform writing evolved over time into printed books of finite dimensions. The notion of the dictionary as an object can be associated with its concept as a *text* in the sense of an ordered set of written words. Even today, it is often considered a book, as we will see below.

In the last print edition of the DLE published by the RAE, we only find the term “diccionario” [dictionary] defined as: *‘libro en el que se recogen y explican de forma ordenada voces de una o más lenguas, de una ciencia o de una materia determinada’* [book in which entries from one or more languages, from a science or from a specific subject, are collected and explained in an orderly manner] (DLE, 2014). However, an online search of the same dictionary proves that there has been a recent update. The term “diccionario” is now defined as a *‘Repertorio en forma de libro o en soporte electrónico’* [Repertoire in book form or on electronic support] (DLE)³². Interestingly, the notion of a dictionary as a book (*‘libro’*) was maintained, and that is the instant association most of us will make. Somehow, this notion is rooted in our unconscious. In all likelihood, we imagine a structured list of words – the compilation of lexical items that make up the inventory of a given language – that form the lexicographic article as a whole. As stated by Dubois (1970), *‘Le dictionnaire n’est pas seulement un objet, un produit de consommation, défini par des besoins socio-culturels, c’est aussi et surtout un texte, un discours continu et clos’* [The dictionary is not only an object, a consumer product, defined by socio-cultural needs, it is also and above all a text, a continuous and closed discourse] (p. 35).

From the very beginning, the question of space restrictions was highly relevant to lexicographic issues in addition to being a significant concern for any lexicographer. The fact that a printed dictionary is a book with finite dimensions led to the development

³² <https://dle.rae.es/diccionario?m=form>

of a number of strategies and certain conventions that characterise it as a text today. Undoubtedly, the typographic technique was a determining condition for the diffusion of dictionaries and served multiple purposes: (1) to save space (e.g., space-saving devices such as abbreviated forms, especially in print dictionaries, or the use of swung dashes; cross-referencing to avoid duplicating information already available in another entry; highly concise mode of expression overall and also, for instance, pocket dictionaries that favour definitions by synonym if possible); (2) to reflect and facilitate the access structure (e.g., bold typefaces to signal the lemma or headword in a dictionary article are easier to find; the numbering of senses and use of different typefaces for different elements in the hierarchy); (3) labels to inform the user about certain restrictions of the entry (e.g., usage labels, such as a ‘colloquial’ register label, generally abbreviated to ‘col.’ or ‘coloq.’).

A few years ago, Rundell (2015) had already remarked on the use of these lexicographic conventions in a digital environment, arguing that they had to be rethought and new policies identified to replace them. Nevertheless, even though dictionaries are currently published on the web, a surprising number maintain these typographic conventions even in their digital versions – their display continues to reflect the configuration of the paper format.

We have also mentioned cross-referencing as an example of a convention; this is associated with the notion of hypertextuality. A print dictionary is never a sequential type of text. For instance, when looking up a word in a dictionary, one will probably not read it linearly since many lexicographic articles are linked with others.

In short, there is relevance in claiming that the dictionary, even in digital format, never ceases to be a properly structured type of text (Frawley, 1989) and can be classified as a textual genre (Bakhtin, 1992) due to its more or less stable format and functional structural aspects (Pereira & Nadin, 2019).

2.1.2 The Dictionary as a Research Object

A dictionary is a *research object* and one can conduct research on it (e.g., typology of dictionaries, behaviour of users, needs analysis); therefore, people explore

dictionaries for various reasons and interests. Insofar as a dictionary records the use of the language or provides guidance regarding its use, it could be an object of research according to the different topics it comprises. We quote some examples of works that demonstrate the diversity of topics we have found: studying a specific language over a period of time, for instance, sexism in dictionaries (Gershuny, 1974; Rodríguez Barcia, 2016); discussing the original meanings of a given lexical unit (Silvestre, Villalva & Pacheco, 2014; Alves, 1997); investigating the lexicographic tradition (Baalbaki, 2014; Kallas et al., 2019); examining a specific type of dictionary and tracing its story (Considine, 2014); scrutinising the content structure of lexicographic works (Amsler, 1980); inspecting dictionaries as a mirror of society (Iamartino, 2020); analysing diachronic and synchronic markup (Williams, 2019).

Many institutions are now involved in mass digitisation projects to make historical documents available online. These retrodigitised dictionaries should not merely reproduce paper versions. Instead, all the components must be appropriately structured to enhance search engines in the future and impart new analytical data on the evolution of lexicography but also of the language *per se*. We cite again as an example the MORDigital³³ project already referred, recently financed by the Fundação para a Ciência e a Tecnologia (FCT), whose main objective is to make the Moraes dictionary – the first modern dictionary of Portuguese lexicography – available online.

2.1.3 The Dictionary as a Cultural Artefact

Earlier, we mentioned that dictionaries from previous periods are gold mines of information on different scientific fields. However, the dictionary can also be considered a *cultural artefact*, reflecting the social, cultural and ideological values of the time it was created, representing some sort of cultural lexical collection. Pruvost (2006) claims that dictionaries are tools of a specific language and culture, portraying the evolution of vocabulary and constituting a historical source. The content of their definitions can grant the end user an idea of the society whose language is described. In the 1980s, Beaujot

³³ MORDigital – Digitisation of *Dicionário da Língua Portuguesa* by António de Moraes Silva [PTDC/LLT-LIN/6841/2020].

(1989, pp. 79–80) described the dictionary as a mirror of the ideology of the culture in which it is produced. Although impartiality is fundamental to a lexicographer, the truth is that when we look up a word in a dictionary, we may find particular ideological assumptions, judgments and prejudices reflecting the way society viewed certain topics at a given time. Homosexuality [*‘homossexualismo’*], for example, was once defined as a *‘inversão sexual’* [sexual inversion] (PE, 1956, p. 795), which is unthinkable today. The word *‘mulher’* [woman] (PE, 1956, p. 1018) was defined as *‘pessoa do sexo feminino pertencente à classe inferior’* [female person belonging to the lower class], also once synonymous with the *‘sexo fraco ou frágil’* [weak or fragile sex] (PE, 1956, p. 1369). In summary, this happens because, over time, most dictionaries tend to reflect the dominant culture established in a society by the group of individuals who direct the ruling ideas, values, and beliefs that become the dominant worldview of a given society.

2.1.4 The Dictionary as a Tool

A dictionary has always served a practical purpose, functioning as a kind of guide. We are used to looking at dictionaries as *tools* designed to respond to certain linguistic and specific user needs. Evidently, people use a dictionary as a tool, considering that no one will read it from A to Z. According to Tasovac (2020, p. 41), ‘the *toolness* of the dictionary is both functional and ideological’; ‘functional’ because it responds to specific user needs and ‘ideological’ because it plays an essential normative role in the codification and maintenance of standard language varieties. More than describing the lexicon of languages, lexicography’s objective is to respond to ‘specific types of information needs detected in society’ (Trap-Jensen, 2018, p. 22). In fact, it has been argued that it broadly aims to produce information tools (Bergenholtz & Gouws, 2012, p. 40), i.e., reference works whose primary function is the improved recovery of information. To this function, we also add the following:

- Facilitating the understanding of written vocabulary, including words whose meaning is unknown or of which we are not sure;
- Facilitating communication;
- Assisting in the study and understanding of a foreign language;
- Defining meanings and establishing the spelling of words;

- Informing about the etymology of words, providing explanations about their origin;
- Specifying the grammatical category or the gender of a lexical unit;
- Contributing to standardising and maintaining the unity of the language; and
- Imparting knowledge.

In summary, nowadays dictionaries are used for understanding texts (reception), for writing in a clear, comprehensible way (production), and for translating different languages.

Concerning the acquisition of knowledge, we disagree with Tarp (2008), who considers this function to be ‘quite simply a bonus’ (p. 87). Many people consult dictionaries to employ, for example, more erudite terms. They do so for etymological purposes, discovering a word’s origin, whether out of curiosity or for writing purposes. Additionally, the search for synonyms can imply the acquisition of knowledge, i.e., a stronger vocabulary. Initiatives such as ‘word of the day’ aim to cater to this need, encouraging dictionary consultation when taking the edited product to the potential audience.

2.1.5 The Dictionary as a Language Model

A dictionary is a linguistic product and somehow it is seen as a ‘judge’ (Mugglestone, 2011, p. 12).³⁴ For a non-specialist audience, what is in the dictionary is undisputed and authoritative (Harris & Hutton, 2007; Beaujot, 1989) and legitimises the use of words. This explains the prevalence of certain *vox populi* statements, such as ‘if

³⁴ In this regard, allow Ana Salgado to narrate a personal episode. In 2004, when Ana was still working as a lexicographer, she was interviewed by the Portuguese newspaper *Expresso*. In the pleasant and fun conversation she had with the journalist, she told her about her passion for words and how she enjoyed working on dictionaries. At one point, Ana said something along the lines of ‘Dictionaries, for me, *were* a Bible’. In using such a phrase, she simply meant that, until that moment, she had seen dictionaries as great sources of reference, somehow untouchable (Ana was far from imagining that she would have the great responsibility of updating dictionaries one day). The journalist (apparently) liked what she said when she made her statement the headline of the news story: ‘Dictionaries are a Bible’. Ana was shocked. It is a dangerous sentence, especially when uttered by someone who says she is a lexicographer. It was horrifying. But why this statement? The authority of the dictionary is unchallenged in society, and a lexicographer is well aware that a dictionary is not a Bible.

the word *x* is not in the dictionary, then it does not exist' or 'if the dictionary says it is so, it must be so', among others. Dictionaries thus stand as models, or as anchors, for a given language.

It should be noted that many words have never been registered in a dictionary, predominantly due to the material constraints of the printed editions. Fortunately, this limitation has been overcome in digital versions. However, a matter of great interest to this research is deciding which terms – words belonging to specialised fields – should be included in a general language dictionary. We are aware that it is precisely these specialised units that increase the number of entries in the dictionary daily. Take, for instance, the current case of the COVID-19 pandemic and the number of epidemiological terms being added to our dictionaries, especially to respond to the users' needs to clarify them. What comes in or goes out is also the concern of any lexicographer; adding or removing words are choices shaped by the actual conditions of writing the dictionary. On the other hand, registering new words and meanings can often turn into ethical issues and challenges related to society's norms and policies.

Concerning the dictionary as a language model – whether normative or descriptive –, descriptive guidance has gradually become more common, a process facilitated by the fact that lexicographers can access increasing amounts of corpora to support their descriptions. However, as Ten Hacken (2018) points out, languages are not 'empirical entities' (p. 838); new words, meanings and usage patterns are being proposed constantly. Therefore, it is wrong to assume that any dictionary can completely contain all the units of a particular language. It is probably more useful to consider dictionaries as problem-solving tools.

As we have seen so far, a dictionary must be seen as a kind of diamond with several facets. It is simultaneously a text, a research object of both digital humanities and digital heritage, a cultural artefact, a tool and a language model constantly mirroring current norms and epochal ideologies.

2.2 Dictionary Classifications

This section presents and describes some of the main dictionary classifications proposed by lexicographers and researchers. Following this comparison, we lay out a taxonomic classification proposal that will serve as the background to introducing the chosen object of study, i.e., academy dictionaries.

2.2.1 An Overview of Dictionary Classifications

There is no standardised and consensual dictionary taxonomy, and there probably never will be. The topic is so complex that Béjoint (2000) mentions various typologies and concludes that ‘dictionaries come in more varieties than can ever be classified in a simple taxonomy’ (p. 37), and for Rey (2003), the typology of the dictionaries ‘is as complex as that of leguminous plants or arthropods, still awaits its Linnaeus or its Cuvier’ (p. 89).

However, in the history of lexicography, it is possible to find some attempts to build organised schemes to classify existing dictionaries, where each author proposes their point of view. One of the first classifications was determined by the Soviet linguist Shcherba (1940/1995), and much of the terminology used by this author was reused in later classifications. The most recent ones assign greater weight to the lexicographic function as a criterion: the objective with which the dictionary is used. Gouws (2020), in turn, states that the decisions regarding the typology of dictionaries to be compiled must be based on the analysis of the target user and the lexicographic needs.

A detailed review of these dictionary classifications goes beyond the scope of this research, but it must be noted that this classification differs across various authors (e.g., Shcherba, 1940/1995; Sebeok, 1962; Malkiel, 1962, 1976; Rey, 1970; Zgusta, 1971; Haensch et al., 1982; Geeraerts, 1984; Arnold, 1986; Hausmann, 1989; Svensén, 1993; Landau, 2001; Hartmann & James, 1998/2002; Porto Dapena, 2002; Tekorienė & Maskeliūnienė, 2004; Devapala, 2004; Gouws & Prinsloo, 2005; Atkins & Rundell, 2008; Engelberg & Lemnitzer, 2009). Meanwhile, a group of researchers have already analysed, readapted, and criticised many of the existing classifications (e.g., Gapporov,

Vositov & Ibragimova, 2020), with some even stressing their limitations (e.g., Yong & Peng, 2007; Smit, 1996).

The construction of classifications is a crucial topic in lexicographic research, for which we can claim two main reasons:

- (1) the need to categorise the dictionaries themselves within the lexicographic universe, serving as a guide for those who make dictionaries;
- (2) from the user's perspective, this categorisation can enable users to clarify doubts when they need to consult dictionaries.

We will only highlight the most significant points. Above all, we will focus on the cases that overlap, i.e., the ones employing the same categories. Exclusive classifications for bilingual dictionaries will not be mentioned here as they do not fit the current research topic.

Concerning the theoretical basis for the classification of lexicographic works, we can distinguish two kinds of models that follow the opposition between taxonomy and typology. Taxonomy is a classification according to a system of predefined criteria that aims to separate elements of a group (*taxon*) into subgroups (*taxa*), which are mutually exclusive and unambiguous. On the other hand, a typology corresponds to a classification that gathers a density of entities that share a more prominent or characteristic feature. This property can be identified as a prototype. By executing the necessary methodological transpositions, a lexicographic taxonomy corresponds to a classification by descending dichotomous criteria. Conversely, a lexicographic typology corresponds to a classification according to a centripetal principle, insofar as in the presence of several features one of them stands out and becomes the highlighted feature of an entity, which has other features that are less dominant. Hausmann (1989) recalls that 'a typology is a classification that is guided by prototypes' (p. 969). In this conception, a prototype corresponds to a type of dictionary that represents the most typical exponent. The others that are less typical are in a more peripheral position than the centre of a category. The lexicographic exponent has a more 'salient' or 'dominant' trait. However, it should be noted that the designations 'taxonomy' and 'typology' are often used interchangeably.

Comparing the different approaches adopted, dictionaries are generally typologically classified into categories – what Atkins and Rundell (2008) calls ‘properties of dictionaries’ (p. 24) – which also vary widely, depending on the scope, perspective and presentation. In the literature, we found the following distinctive categories:

- a) size (from Lilliputian to large);
- b) coverage (from general to specialised);
- c) number of languages (monolingual, bilingual or multilingual);
- d) ordering (from alphabetical to thematic); medium (printed, electronic or digital);
- e) number of entries (very debatable because it is directly related to a given lexicographic tradition and the language it reflects);
- f) functionality; predominance of categorical information (dictionary, encyclopaedia, etc.);
- g) and target user (student, translator, etc.).

The most traditional classification considers two major categories:

- (a) language dictionaries;
- (b) encyclopaedic dictionaries, combining linguistic and extralinguistic information (e.g., Arnold, 1986; Zgusta, 1971).

The first (a) concern words and are designated as ‘books of words’; the second (b) focus on ‘things’ (*realia* or *denotata*), the encyclopaedias par excellence or ‘books of things’. Dubois & Dubois (1971) try to clarify:

Le dictionnaire de mots est le dictionnaire de langue; le dictionnaire de choses est le dictionnaire encyclopédique. Ils se différencient par la place qu’ils donnent à l’usage linguistique ou au contenu auxquels les mots renvoient. [The dictionary of words is the language dictionary; the dictionary of things is the encyclopaedic dictionary. They differ based on the emphasis they place on the language in use or the content to which the words refer.] (Dubois & Dubois, 1971, p. 13)

Hartmann & James (1998/2002, pp. 147–148) differentiate between general and specialised dictionaries, where the distinguishing factor is the presence of linguistic or factual information. These categories are intertwined; thus, according to the authors,

we can find language dictionaries with complementary information of an encyclopaedic nature and others that are more focused on linguistic descriptions.

Some classifications are both governed by categorical and factorial principles (Zgusta, 1971), using classic oppositions (language dictionaries vs encyclopaedic dictionaries) as descriptors and simultaneously quantitative descriptors (such as size). For instance, Malkiel (1962; 1976) employs three criteria to distinguish dictionaries: scope, perspective and presentation. Landau (2001, p. 8) detects advantages in this classification, considering that virtually every type of dictionary can be analysed based on these three distinctive characteristics: the scope refers to the size, extent of the lexicon covered, number of languages and concentration on lexical data, while the perspective refers to the approach of lexicographic work. This category distinguishes, for example, the length of time covered by the dictionary, i.e., diachronic (covering an extended period) or synchronous (limited to a period of time). It also refers to the conventional organisation of the presented information (alphabetically, by concept, etc.) and the tone (prescriptive vs normative; didactic vs playful). And, finally, the presentation refers to the content and presentation of each dictionary entry's information, such as usage information, examples and illustrations.

Among the various proposals, there are huge overlaps and, at times, some inconsistency. Many proposals are incomplete (Zgusta, 1971), whereas others are too theoretical, rendering their applicability vastly reduced (Rey, 1970). Hausmann (1989, p. 972) summarises Rey's classification by pointing out that he had covered the entire range of dictionaries despite lacking some precision. Additionally, he considers that these typological models correspond to the decisions made by the lexicographer regarding linguistic data, lexicographic units, lexical quantities, data ordering, non-semantic information and examples.

The function of the dictionary guides functional classifications. In this scenario, Engelberg & Lemnitzer's (2009) model is usually considered the best example. However, it is pertinent to direct attention to the fact that in this proposal, the division of dictionaries by the criterion of *Benutzergruppenorientiertes Wörterbuch* [dictionaries oriented by user groups] is only one classification criterion among others. It is also necessary to recognise that, to date, dictionaries listed based on this criterion are

restricted to the scope of teaching and learning, both in one's mother tongue and foreign languages.

It needs to be said that certain definitions of typologies of dictionaries emphasise the shared characteristics – 'the classification of dictionaries based on shared properties' (Van Sterkenburg, 2003, p. 459) – although we believe that it should be precisely the opposite, i.e., the focus should be on the contrasting characteristics (Devapala, 2004). We agree with Geeraerts (1984) when he states: 'an adequate typology of dictionaries should specify *the features concerning which dictionaries can differ*' (p. 38) [emphasis added]. In this path, we will now present the taxonomic classification adopted in the thesis, which is actually a revision of the proposal of Geeraerts & Janssens (1982).

2.2.2 Taxonomic Classification Proposal

Taking into account the scenario described above, we consider the following statements for the purpose of this thesis:

- (1) We recognise that it is impossible to delimit dictionary types in a rigid structure. Developing a universal classification that represents all the complexities surrounding the dictionary concept as a lexicographic product is hardly feasible.
- (2) As a categorisation system, we start by choosing a taxonomic classification and subsequently a typological classification.
- (3) The criteria can be linguistic and functional and will not take into account quantitative criteria. In the digital age, we believe it makes no sense to classify dictionaries by their formal characteristics, such as size, which were formerly very useful for publishers to define their range of dictionaries.
- (4) We argue that the criteria should be classified based on linguistic and extralinguistic features. We distinguish resources – for instance, considering the number of languages (linguistic) but also a semasiological, onomasiological or mixed approach concerning the organisation of knowledge (extralinguistic).

Our proposal can be viewed in the following diagram (Figure 12):

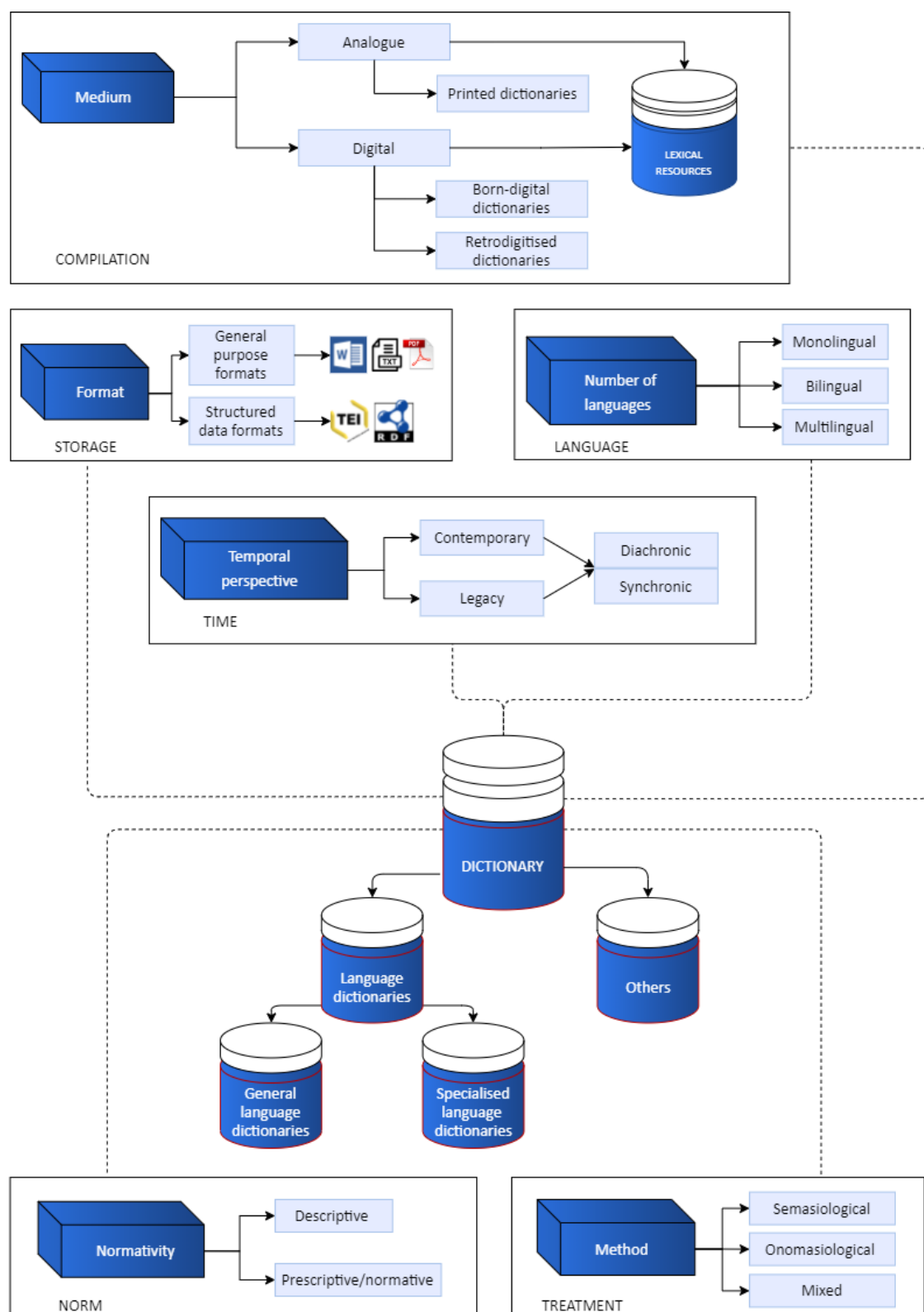


Figure 12: Categories of a Dictionary's Taxonomic Classification

In Figure 12, we consider two significant distinctions, LANGUAGE DICTIONARIES and OTHERS, to accommodate all the other works that do not fall under the first category, such as encyclopaedias, glossaries and terminological dictionaries, which will not be analysed here. In turn, LANGUAGE DICTIONARIES can be subdivided into GENERAL LANGUAGE DICTIONARIES, which assemble, preserve and describe (monolingual), or translate (bilingual) the lexicon of a given language in addition to being characterised by their syncretic nature (Silvestre, 2016, p. 204), and SPECIALISED LANGUAGE DICTIONARIES, i.e., dictionaries whose object is a specific element of the linguistic description, be it a specific portion of the lexicon or a thematic area; for example, orthographic dictionaries and etymological dictionaries, among others.

In this proposal, we also identify the main categories, which are described as follows:

Medium. A dictionary can be compiled and used on different media:

- *analogue*, which refers to all non-digital documentation media, whose example *par excellence* is paper, i.e., *printed dictionaries* but also includes, for instance, Sumerian clay tablets;
- *digital* refers to those dictionaries currently available on the web or in mobile apps, but it can also denote a print dictionary since it is possible to envisage a scenario where we use a dictionary-writing software and still distribute the dictionary as a book. Additionally, in this category, we include all dictionaries in electronic media that are no longer commercialised (e.g., floppy disk, CD, DVD, pen drive). In this case, it is still important to distinguish *born-digital dictionaries*, created as machine-readable, from *retrodigitised dictionaries*, which were converted from an analogue (paper) or digital (e.g., PDF) medium to a computer-readable format, using optical character recognition systems and involving the encoding step of the scanned version. As such, a dictionary generated with a word processor, such as *Microsoft Word*, can be described as born-digital. Further included in this category are any resources compiled using a computer.

Digital dictionaries and retrodigitised dictionaries are usually compiled into databases, giving rise to the so-called *lexical resources*³⁵.

Format. Dictionaries are modelled and encoded in multiple diverse *formats*, indicating that the information is organised and stored in files of a different nature, hindering the path of sustainability and imposing constraints due to interoperability issues. The formats can thus refer to different types of files:

- *general purpose formats*, such as plain text, *Microsoft Word* (e.g., doc or docx, xls) or PDFs; and
- *structured data formats*, such as Text Encoding Initiative (TEI), Lexical Markup Framework (LMF), Resource Description Framework (RDF).

Since general dictionaries are our object of study, we restricted our proposal to those we consider the most relevant distinctive properties for lexicographic research and work. When compiling general dictionaries, we also took into account the following attributes:

Number of languages. Depending on the number of languages described, dictionaries can be classified as *monolingual*, *bilingual* or *multilingual*. According to Svensén (1993), ‘The monolingual dictionary describes a language by means of that language itself: it gives the meanings of words by means of definitions or explanatory paraphrases’ (p. 20). Bilingual dictionaries routinely distinguish between the source and target languages. Svensén (1993) also stated that ‘The bilingual dictionary shows how words and expressions in one language (the source language) can be reproduced in another language (the target language). This is done by showing the expression in the source language, followed by one or more equivalents in the target language’ (pp. 20–21). Multilingual dictionaries, as we can infer from the name itself, include several

³⁵ Lexical resource is a ‘language resource in the form of a database consisting of one or more lexicons’ (cf. ISO/FDIS 24613-1, 2019).

languages; they are closely related to bilingual dictionaries, but the equivalent information for a lexical unit is given in several languages.

Temporal perspective. The time axis here is relevant. Dictionaries can be *contemporary*, i.e., they can be subject to constant updating, but we also intend to work with *legacy* dictionaries, i.e., dictionaries of great linguistic, historical and cultural interest. Both can be subdivided into *diachronic* (historical evolution of each word's form and meaning) and *synchronic* (the language in a specific period of its evolution) dictionaries.

Normativity. Dictionaries can be *descriptive* or *prescriptive/normative*, establishing the model to follow. Prescriptivism is an approach that attempts to determine the rules of correct usage of a language, while descriptivism is an approach that analyses and describes how the speakers of a language actually use it.

Method. The methodological approach adopted can be of a *semasiological*, *onomasiological*, or *mixed* nature. In a semasiological approach, one starts from the lexical unit to identify the meaning(s). In an onomasiological approach, we begin from the concept to identify the lexical unit designating it. Finally, in a mixed approach, as adopted in this thesis, lexical units are treated according to lexicographic and terminological assumptions that consider the double dimension (conceptual and linguistic) of terminology.



Based on what we have just explained, the academy dictionaries under study are: semasiological; prescriptive in different degrees, as we will see in the following chapters; contemporary, since their content is the subject of constant updating; monolingual; based on structured data formats. In terms of medium, our research focuses on the printed dictionaries – DLPC (2001) and DLE (2014) – and the digital versions of DLE and DAF, as well as the DLP resulting from the retrodigitised version of the DLPC PDF (Figure 13).

ACADEMY DICTIONARIES	
Method	semasiological
Normativity	prescriptive in different degrees
Temporal perspective	contemporary
Number of languages	monolingual
Format	structured data formats
Medium	print + digital

Figure 13: Classification of the Academy Dictionaries under study

2.3 Dictionary Structure

The dictionary structure is the sum of all the parts of a dictionary. A semasiological-oriented dictionary is always organised into a relatively stable structure that interconnects its different parts (Bergenholtz & Tarp, 1995, p. 188).

The *megastructure* – the dictionary as a whole, referring to the general structure of the parts that compose it – comprises two different sections: the first is the *main body* of the dictionary and the second is its *outside matter*. The *outside matter* includes the *front*, *middle* and *back matter*. Although Müller-Spitzer (2013, p. 374) prefers the term *outer features* for digital dictionaries, since not every element in the external domain of online dictionaries belongs to the text category (Klosa & Gouws, 2015, p. 148), we will enlist the term *outside matter*, since our analysis will be based on the texts of the latest printed editions of the Portuguese and Spanish academy dictionaries, except for the French dictionary, whose latest edition is restricted to the online version. Hartmann and James (1998/2002) describe in more detail these components:

- (i) the *outside matter* or the section of metadata – the set of texts external to a dictionary's lemma list such as the front matter (e.g., preface, user's guide, collaborators list), located before the lemma list, is a mediator

between the dictionary and the users that enables them to take advantage of the available resources. In simpler terms, we could call this first component the introduction to a dictionary;

(ii) the *middle matter*, located between the macro- and microstructures, is the interruption between these components (e.g., illustrations, encyclopaedic information);

(iii) the *back matter*, located after the word list, brings information such as verbal conjugation, grammar sections, in the form of appendices.

The *main body* of a dictionary has three structures: *macrostructure*, *microstructure*, and *mediostructure*.

The terms *macrostructure* and *microstructure* are the most used within the lexicographic community. Baldinger (1960, p. 524) was the first to use these terms when he stated that microstructures must be organised within a macrostructure. In the following decade, Rey-Debove (1971) defined *macrostructure* as ‘the set of entries’ (p. 21). In the same vein, Hausmann and Wiegand (1989) referred to it as ‘the ordered set of all the lemmas in the dictionary’³⁶ (p. 328). Indeed, the term *macrostructure* has been commonly used in two senses: as a synonym for ‘*nomenclature*’ (Rey-Debove, 1971), ‘word list’ (Béjoint, 2000) or ‘lemma list’ (Svensén, 2009), and as a reference to how the body of the dictionary is organised (the entire structure of the main components of a dictionary) – for which the term *megastructure* is adopted here. All the aspects related to the number of lexical items, the type of registered lexical units, and their arrangement in the dictionary are related to the macrostructural scope. Thus, in this work, we understand *macrostructure* as the set of the lexical units included in the dictionary making up the lemma list and their respective organisation (e.g., alphabetical order, arrangement of homographs, sublemma organisation).

The *microstructure* includes all the ordered lexical information present in each dictionary entry (Rey-Debove, 1971, p. 21). In this study, we used the term *microstructure* to refer to the multiple lexicographic components that constitute a lexicographic article. The type of information given varies depending on the type,

³⁶ For this sense, in Portuguese, the term *nomenclatura* [nomenclature] is currently used; in Brazil, also *nominata*; in English, it is more common to use word-list; in Spanish, *nomenclatura* or *macroestructura*.

purpose and size of the dictionary. Typically, dictionaries include the following information: grammatical information, such as part of speech, gender, number; usage labels; meaning; examples; etymology; and elements of representation (e.g., icons or symbols). In summary, the *microstructure* provides information on the form, meaning or semantic information, syntagmatic information on fixed combinations, and paradigmatic information involving synonyms, hyponyms, etc. The format of a lexicographic article is defined by certain typographic conventions, explanatory texts and symbols.

The *mediostructure* (Wiegand, 1989) corresponds to what Hartmann and James (1998/2001) cite as the ‘cross-reference structure’, i.e., the cross-referencing of different components of a dictionary, particularly between lexicographic articles. This definition, however, conveys a misconception about this component, since it can also refer to related terms, hypernyms, hyponyms and hypertexts. While in print dictionaries, there are cross-references, in digital dictionaries, there are hyperlinks that point to a certain lexicographic article or a particular sense. The main difference is that, in print dictionaries, you were stuck to the object – there was no getting out of the book; and, in a digital environment, you can ‘get out of the box’ with the insertion of external links. Wiegand (1996/2011, pp. 1164–1168) distinguishes between different types of mediostructures, such as (i) dictionary-internal mediostructures (cross-referring within the same dictionary), (ii) dictionary-linking mediostructures (cross-references linking lexicographical data in one dictionary by means of references to data in another dictionary), (iii) source-related mediostructures (cross-referring to external sources), (iv) literature-related mediostructures (cross-referring to literature).

A scheme of a dictionary structure can be visualised in Figure 14.

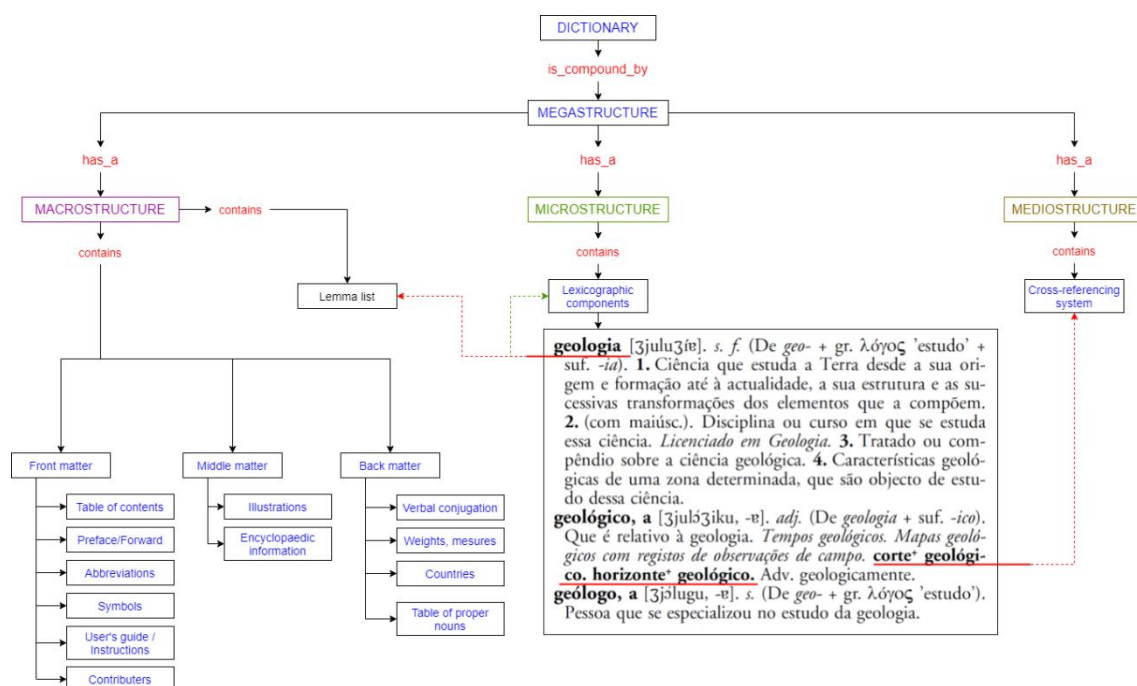


Figure 14: Model of a Dictionary Structure

The concepts of macrostructure and microstructure will be explored in more detail in Chapter 6 accompanied by the analysis of the front matter and lexicographic articles of the academy dictionaries. The *lemma* is both part of the macrostructure as well as of the microstructure and therefore plays a pivotal role. In most European languages dictionaries, the lemma is usually singular if there is a variation in number; the masculine form is used if there is a variation in gender, whereas the infinitive form is used for all verbs.

While the conversion of printed dictionaries signalled a paradigm shift, the dissemination of the web has forced us to rethink the concept of lexicographic work. This effective exchange of content between systems always depends on metadata that describe content so that the systems involved can effectively profile the material received and combine it with their internal structures.

2.4 Going Further: Modelling and Standardising Lexicographic Resources

Conceiving digital lexicographic resources increasingly requires the application of adapted standards and tools capable of guaranteeing the availability of structured

data and ensuring interoperability between systems. To transform a raw document into a structured one, we need to define the different data types that comprise it to model it according to a standardised data model, rendering interoperability feasible.

Actually, the digital revolution (Trap-Jensen, 2018; L’Homme & Cormier, 2014) increasingly requires the application of standards and adapted software to be capable of guaranteeing the structured publication of data for different systems, especially when the lexicographic production scenario is very heterogeneous due to its nature, form and content. There are several types of dictionaries, in several languages, with disparate structures and different functions, purposes and users. Many of them adopt a hierarchical data structure representation, mainly based on Extensible Markup Language (XML).

The application of standards undergoes a few processes, such as *modelling* and *encoding*. *Modelling* refers to how researchers conceptualize external representations (Godfrey-Smith, 2009) – the process of creating a data model that can account for all the lexical data and their components. *Encoding* refers to the process of expressing an abstract, conceptual model using a specific data format (e.g., TEI Lex-0). Essentially, *modelling* is a design task, and encoding is an implementation task. This is a crucial issue for lexicography to ensure interoperability between the software components of heterogeneous lexicographic resources (Romary & Wegstein, 2012).

Although a reasonable number of lexicographic works can currently be consulted online, these dictionary resources end up being static, failing to take real advantage of the digital environment. Now, more than ever, any lexicographer needs to know how to take advantage and explore the possibilities of the digital environment (Trap-Jensen, 2018; Bergenholtz, Nielsen & Tarp, 2009) to create dynamic, more robust lexicons, enriched with semantic, conceptual and statistical information, where data from different resources can be linked (i.e., linked data).

We propose to apply these new principles – computational methods and interoperable standards (Chapter 8) facilitating the organisation of large amounts of data and lexical metadata – according to the defined methodology in this thesis and essentially base it on linguistic knowledge, which is often ignored.

CHAPTER 3

European Lexicographic Tradition

*A story about dictionaries is a story about books, but it is also,
most importantly, a story about people.*

CONSIDINE (2014, p. 8)

Lexicography, boasting of old history, has undergone substantial evolution. Clay tablets, lists of difficult words, glosses and glossaries were replaced by what we now call dictionaries. There are two decisive moments in this process: the invention of typographic printing in Europe by Johannes Gutenberg during the 16th century and the development of computer technology accompanied by the digital revolution in the 20th century. Dictionaries of various types, compiled from the early age of civilisation, were indispensable to preserving and disseminating linguistic conventions and cultural factors in a language community.

As a complete survey of the world's lexicographic production is beyond our scope, we have limited ourselves to presenting a brief retrospective from the first lexicographic works to the emergence of national academies and, more precisely, the representative selected academies in this study. We propose to highlight the production of monolingual general language dictionaries and locate our object of study (academy dictionaries) within the tradition.

Academy lexicographic works represent a large-scale and long-term dictionary project initiated and compiled by official national bodies established to record, maintain and promote authoritative accounts of language use. A contextualisation of the beginning of the academy tradition is presented, with the publication of the *Vocabolario degli Accademici della Crusca* (1612) and the *Dictionnaire de l'Académie Française* (DAF, 1694), which spreads throughout Europe, encompassing several prestigious dictionaries compiled by academies or inspired by this academy principle during the 17th and 18th centuries. We see how the Enlightenment was the golden age of the academy dictionary, when these texts served as an authoritative resource for the study of European vernacular languages. Then, the three European academic institutions

selected in this work are presented, described and analysed, as well as the chosen dictionaries. We begin by referring to the emergence of the Académie Française, which will serve as a model for the others, that is, the Real Academia Española and the Academia das Ciências de Lisboa. A brief retrospective of the various editions of academy dictionaries is made, from the beginning to the present day.

3.1 The Origins of Lexicography

One of the oldest lexicographic works that we know of can be traced back to pre-classical antiquity, to a time when the invention of writing revolutionised human communication. The tabular prototypes, distant ancestors of what we would call a dictionary, are lists of words³⁷ in Sumerian, in cuneiform script, engraved on clay tablets found in the city of Uruk (situated on the eastern banks of a channel of the Euphrates River). These clay tablets were used to teach writing. The students were required to make copies of these lists, thus training their handwriting and learning how to write new words by thematic groups.

The items discovered in the ancient city of Ebla (in Tell Mardikh, modern Syria), which are notably bilingual, are noteworthy. There are 24 clay tablets in cuneiform script from the Sumerian civilisation, from ancient Mesopotamia, dating from around 3200 BCE (Lynch, 2016). They contain lists of words in Sumerian and Akkadian (they were called *HAR-ra* = *hubullu* or *Urra-hubullu*)³⁸ and resemble glossaries that covered all kinds of words to name occupations, animals or vegetable life.

In addition to the compilation of thematically ordered Egyptian lists of hieroglyphs, such as *Ramesseum Onomasticon* and *Onomasticon Tebtunis*, Greek lexicons occupied a prominent position in the early days of lexicography. Philitas de Cos and Simias de Rhodes compiled the first, extensive collections of glosses of erudite words from Ancient Greece around 300 BCE. The study of Homeric texts and the desire to understand ancient legal texts led to the elaboration of the first glossaries, where words that were difficult to understand were listed and defined to facilitate their

³⁷ For school lists from the Sumerian archaic period, see Englund & Nissen (1993).

³⁸ Which is, indeed, the first entry; a word that means 'debt with interest'.

reading. These are the modern lexicographer's predecessors, philologists concerned with understanding previous literary texts and correcting errors.

The compilation of the first surviving Chinese dictionary, 尔雅 [*Erya* or *The Ready Guide*], dating from the third century BCE, has no known author, and its title literally means '*próximo da língua padrão, visando aproximar a língua dos utilizadores da língua padrão*' [close to standard language, aiming to bring the language of users closer to standard language] (Wang, 2016, p. 277). The work is divided into 19 chapters: the first three define lexical units, and the remaining 16 explain the meaning of objects, animals, plants, etc., much like an encyclopaedic-type dictionary (Yong & Pen, 2008). Furthermore, it is considered the first prescriptive dictionary made on Chinese soil.

At the beginning of ancient Latin lexicography, in the 1st century BCE, we can find works such as *Liber glossematorum* by Lucius Ateius Philologus or *De verborum significatu* by Marcus Verrius Flaccus, the latter being the most significant lexicon of the language. Hellenistic and Roman culture established a model of studies based on the analysis of a few texts by certain classic authors who, due to their style and moral teaching, deserved to be part of a canon. Here lies the origin of the quotations of the current dictionaries.

In the Middle Ages, Latin, known as Vulgar Latin, already had many differences compared to Classical Latin, the language of instruction in universities, liturgy or law. Thus, the practice of glossing texts – explaining the meaning of difficult words through notes – came to life. The glosses were written between the lines or in the margins of the texts, hence leading to the introduction of the designation interlinear gloss (written between one line and another), which later changed to marginal gloss (written in the margins). Medieval bilingual glossary listings (Latin-Vernacular) were published primarily to assist the learning of Latin throughout the period.

With the advent of the Renaissance, more precisely at the beginning of the 16th century, '*a lexicografia começou a estruturar-se como disciplina linguística [...] em vários centros humanísticos europeus*' [lexicography started to be structured as a linguistic discipline [...] in several European humanistic centres] (Verdelho, 2007, p. 14). The translation of the two classical languages, Greek and Latin, into ordinary languages also progressively increased.

One of the most celebrated volumes of the Renaissance era is the Latin-Italian *Dictionary Latino* by the Italian monk Ambrogio Calepino (c. 1440–1510), published in 1502. In later editions, compiled by other dictionaries, this work included as many as 11 languages; 210 editions were printed, the last one in 1779. The book became so famous that the term *calepino* became synonymous with ‘dictionary’. The humanist lexicographic works that have emerged use the *calepino* and *Diccionario latino-español* (1492) by the Spanish philologist Elio Antonio de Nebrija (1441–1522) or *Thesaurus Linguae Latinae* (1531) by Robert Estienne (1503–1559) as reference sources.

However, the Renaissance individual increasingly required linguistic exchange instruments that enabled communication between the various European nations and, therefore, bilingual dictionaries multiplied throughout the 16th century. Despite the significance of these publications, it is known that many 17th-century dictionaries copied each other (Biderman, 1984), and they have many gaps. To clarify this last point, we must note that compiling a dictionary was a herculean task before the computer age. We have to remember that these dictionaries resulted from the work of individual authors who copied and collected the lexical information into paper slips or index cards without any computerised corpora, editing tools or even spellcheckers available to swiftly verify inconsistencies.

3.2 The First Monolingual Dictionaries

The Enlightenment brought renewal to several fields of knowledge, especially concerning the description of living languages, when Latin was still the language of instruction, redesigning the dictionary role as a metalinguistic instrument. Across Europe, there was an appreciation of Vernaculars directly related to the emergence of nation-states (Burke, 2010), which sought to build a national cultural and linguistic heritage. Consequently, the publication of dictionaries became a tool of this construction, for the purpose of normalisation and affirmation of national languages, promoted by several European academies. The gold standard languages were seen as an instrument of power, a power that academies seized to relegate minority languages or dialects to a secondary position.

Despite previous experiences, we can confidently say that modern, monolingual lexicography in a common language initially emerged in the 17th century in the region shared between Italy, France and Spain. The first work with these characteristics is the *Tesoro de la lengua castellana, o española* (1611) by Sebastián de Covarrubias (1539–1613) or its continuation by Juan Francisco Ayala Manrique with the *Tesoro de la lengua castellana, en que se añaden muchos vocablos, etimologías y advertencias sobre el que escrivió el doctíssimo Don Sebastian de Cobarruvias* (1693), which was never finished.

Before moving on to the academy work on our object of study, a reference to the production of the French dictionary is necessary due to the influence it had on subsequent works, with the 17th century being considered its *grand siècle* [great century]. The most prominent works in this context are Father César-Pierre Richelet's (1626–1698) work, *Dictionnaire françois, contenant les mots et les choses, plusieurs nouvelles remarques sur la langue françoise* (1680) with 25,000 entries, and Antoine Furetière's (1619–1688) *Dictionnaire universel* (Furetière, 1690).

In 1662, Furetière was elected to the Académie Française (AF), which had been trying to produce its dictionary for decades. He began his academic activity with great promise. However, given his colleagues' lack of interest and the restrictions imposed on the word list – they rejected a certain encyclopedism –, he eventually decided to elaborate his own dictionary, *Essais d'un dictionnaire universel*, which later scandalised the *immortels*.³⁹ Thus, Furetière was expelled from the *Académie* in 1685 and died in 1688. The dictionary, in three volumes, was posthumously published in 1690 in the Netherlands by Pierre Baile. Furetière had compiled a fine encyclopaedic dictionary, emphasising the arts and sciences, and his great *dictionnaire* was soon recognised as more comprehensive than the French Academy's.

Among the modern European monolingual dictionaries, we also find a Portuguese reference worth mentioning: *Vocabulario Portuguez e Latino*⁴⁰ by Rafael Bluteau (1712–1728), which served as a basis for future dictionary writers and many authors who reused the encyclopaedic and metalinguistic precepts supported by him.

³⁹ The members of the Académie Française are nicknamed the immortals because of the inscription 'À l'immortalité' [for immortality], which is on the official seal of the institution and was offered by Richelieu.

⁴⁰ For a detailed analysis of Bluteau's work, see Silvestre (2008).

Bluteau marks the transition between the Latin-Portuguese dictionaries and the first monolingual dictionary, i.e., the *Dicionário da Língua Portuguesa* (1789) by António de Morais Silva (1755–1824), commonly known by antonomasia as the *Morais*⁴¹ dictionary, which inaugurates the modern lexicography of the Portuguese language (Biderman, 1984, p. 5; Verdelho, 2002, p. 473).

3.3 The Rise of the Academy Tradition

The academy tradition⁴² of producing dictionaries of living languages spread throughout Europe as a dictionary model in the 17th century. But why call it academy tradition? As Considine (2014) said: ‘because the dictionaries which constituted it were often the work of learned bodies called academies’ (p. 2).

Throughout the 17th and 18th centuries, scientific academies began to appear throughout Europe, intending to boost research and disseminate and promote the application of new scientific knowledge. Academies allowed direct contact between scientists and encouraged the progress of science.

The beginning of this movement can be traced to the project of the members of the Florentine society, Accademia della Crusca, when they published the *Vocabolario degli Accademici della Crusca* in 1612, which was created in the previous century in Florence.

⁴¹ It is a condensed version of Bluteau’s work, to which Morais added new entries, ‘reformed and accredited’, so it was said to be the first edition, with Morais taking over the authorship only in 1813, for the second edition. After his death, it continued to be edited and updated. Furthermore, the author of the current work is involved with other colleagues in the digitisation of the first three editions of this historically significant Portuguese dictionary as part of the already mentioned Portuguese national project [MORDigital – PTDC/LLT-LIN/6841/2020].

⁴² Considine (2014, p. 2) points to this concept of academy tradition since dictionaries were the result of the work of these national societies. He recognises, however, that the term is rarely used by lexicographers, historians or researchers. Referring mainly to studies of Scandinavian lexicography, he talks about the academic principle (‘academy principle’), used in 1907 by Verner Dahlerup (Danish form: ‘*akademiprincip*’), in a paper (*Ordbog over det danske sprog*) that described the guiding principles of the Danish national dictionary: ‘The principle is that which takes its most typical expression in the French Academy dictionary, namely that the dictionary will contain only good words: it must, so to speak, be an honour for a word to find a place in the dictionary, just as it is an honour for a work of art to find a place in the national art collections’ (Considine, 2014, p. 3). The same author concludes with a reflection, matching this ‘academy principle’ to the more recent term ‘metalexicography’.

The first academies arose during the Renaissance in Italy. The origin of the term 'academy' from the ancient Greek *Ἀκαδημία* is attributed to Plato, who named his school in honour of Academus, owner of the gardens where he met with his disciples. During the Renaissance, academies began to designate gatherings where philosophy, science or literature were discussed. These groups are at the origin of the academies of sciences, understood as institutions dedicated to the research, discussion and dissemination of science that were eventually financed by the State and still are. They comprised select groups of academicians distinguished for their scientific work, who found a place to debate and publicise their work. An essential part of the scientific development of Europe in the 17th and 18th centuries lies in the activities carried out by these institutions (Peixoto, 1997, p. 71).

One of those first Renaissance corporations of sages, the Brigata dei Crusconi in Florence, gave rise to the Accademia della Crusca⁴³ founded in 1585. The name '*crusca*', i.e., the bran (the thickest part of the flour after being sieved) implies that academics, with their sieve, should be able to separate the superfluous and unsatisfactory customs of the language. The normative intention in fixing the language was thus present from the beginning, including the institution's symbology, to signify the work of 'cleaning up' the language⁴⁴.

Crusca produced the first academy dictionary, *Vocabulario degli Accademici della Crusca* (1612), to reduce the various Italian dialects, defend the common language of Tuscany and establish a linguistic standard based on Dante, Petrarca and Boccaccio. This first academy lexicographic work demonstrated how academies could successfully establish themselves as dictionary makers. From then on, it served as a model for future dictionaries in other countries. As Considine (2014) concludes: 'This dictionary, more than any other, was the foundation of the scholarly lexicography of the living languages of Europe' (p. 27).

The *Vocabulario degli Accademici della Crusca* was followed by the *Dictionnaire de l'Académie Française*, which was started in the 1630s and published in 1694 in Paris. This institutional model was very successful and was followed by the Royal Society of

⁴³ A summary of the history of the Accademia della Crusca can be found in Grazzini (1991).

⁴⁴ The bran is the part of the wheat that is discarded when the grain is cleaned up.

London (1662), the Paris Académie Royale des Sciences (1666) and the Berlin-Brandenburgische Akademie der Wissenschaften (1700), among others. An identical premise governs its foundation: to write a dictionary to preserve and improve the language as well as to regulate the use, vocabulary and grammar of languages.

The academy dictionaries as a *cultural object* have been used as tools for nation-building. Hence, they constitute a significant part of ‘cultural memory’ (cf. Ahumada, 2002, p. 20; Rey, 2008, p. 120). Correia (2009) states, ‘*Quando uma língua se torna oficial, procura-se imediatamente que ela passe a dispor de um dicionário geral monolíngue que descreva o seu vocabulário essencial e que fixe os seus modos de dizer, os seus padrões linguísticos*’ [When a language becomes official, measures are immediately taken to produce a general monolingual dictionary that describes its essential vocabulary and fixes its ways of saying, its linguistic patterns] (p. 16). Thus, academy dictionaries are a good indication of the setting of a standard insofar as the dictionary is a reference work whose object is to represent, as closely as possible, the norm of the linguistic community to which it is intended. In the words of Rey (1983), ‘*La fonction du dictionnaire est de fournir à ses usagers une référence sur la norme*’ [The function of the dictionary is to provide its users with a reference on the standard].

During the 17th and 18th centuries, with an eye on this lexicographic legacy, several national language academies embarked on projects to compile dictionaries (Spain and Portugal, for example) as a way of asserting that specific languages, or varieties of languages, were sufficiently unified and stable to be an object of study, thus seeking to promote the coherence and stability of the language. With the dawn of the 18th century, several English lexicography projects inspired by this academy principle began to emerge, aiming not only to take stock and define all the words in English but also fixing the language, even if not promoted by an academy. As stated by Klein (2015), ‘the bulk of lexicographic work, however, was always done by enterprising publishers and engaged individuals, such as Dr Samuel Johnson’. Samuel Johnson’s⁴⁵ work, *A Dictionary of the English Language* (Johnson, 1755), is a good example of how a

⁴⁵ In 1746, Samuel Johnson signed a contract with a consortium of booksellers to produce a new English dictionary.

dictionary may be written in a commercial enterprise's scope, regardless of any official support.

Let us now turn our gaze to academy dictionaries, which will be the target of our study and in which 'All of them depended on the belief that the languages or language varieties which they treated were sufficiently unified and stable to be coherent objects of study, and some of them sought to promote the continuing coherence and stability of a language' (Considine, 2014, p. 3).

3.3.1 Académie Française

The origin of the *Académie* goes back to the years 1620 and 1630 when a group of *gens de lettres*, an assembly of writers and scholars, held informal meetings at the house of the civil servant Valentin Conrart (1603–1675) in Paris, where they had discussed all sorts of things; they met to talk about literary topics and to read and mutually review their works. Contrary to the case of *Crusca*, the impetus for the constitution of a society did not come from the members themselves but an external authority. Cardinal Richelieu (1585–1642) protected the group, preparing it to establish a French-language academy.



Figure 15: Emblem of the Académie Française (AF)

The main function of the new institution, according to its Charter (Figure 16), was '*travailler avec tout le soin et toute la diligence possibles à donner des règles certaines à notre langue et à la rendre pure, éloquente et capable de traiter les arts et les sciences*' [to work with all the care and diligence possible to provide our language with specific rules and to and make it pure, eloquent and capable of treating the arts and sciences] (AF, 1635/1995).

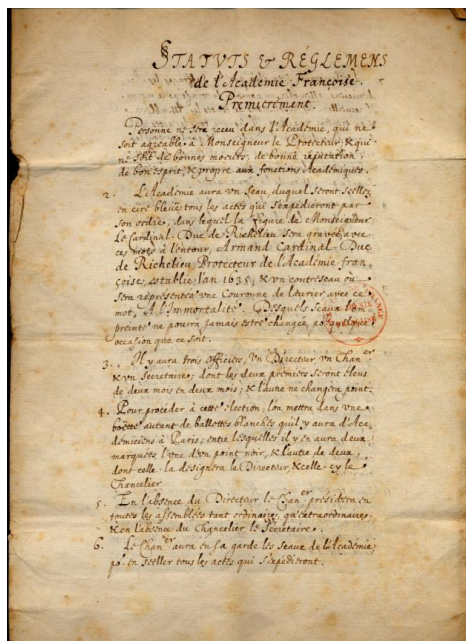


Figure 16: Charter of the Académie Française (1635)

XXIV

La principale fonction de l'Académie sera de travailler avec tout le soin et toute la diligence possibles à donner des règles certaines à notre langue et à la rendre pure, éloquente et capable de traiter les arts et les sciences.

XXV

Les meilleurs auteurs de la langue française seront distribués aux académiciens pour observer tant les dictionnaires que les phrases qui peuvent servir de règles générales et en faire rapport à la Compagnie, qui jugera de leur travail et s'en servira aux occasions.

XXVI

Il sera composé un dictionnaire, une grammaire, une rhétorique et une poétique sur les observations de l'Académie. (AF, 1635/1995)

To this day, the *Académie* maintains its status as the guardian of good practice and witness to the evolution of the French language. This mission is, therefore, enshrined in the very up-to-date statutes of the academy today: '*fixer la langue française, de lui donner des règles, de la rendre pure et compréhensible par tous*' [to fix the French language, to give it rules, to make it pure and understandable by all] (AF, 1635/1995).

All the words of the *bon usage* should appear in the *Académie* dictionary, helping French become a communication system suitable for the arts and sciences. Finally, the *Académie* is required to develop, in addition to a dictionary, grammar and rhetoric textbooks – '*Il sera composé un dictionnaire, une grammaire, une rhétorique et une poétique*' (AF, 1635/1995). The textbook was never produced. The grammar only appeared in the 20th century; as stated previously, the dictionary saw the light of day in the year 1694.

The institution has been operating up to the present day, except for an interruption during the French Revolution. The AF was born equipped with the mission of creating a dictionary of the French language, which would be a treasure of the

language and represent a linguistic authority in the style of the times of authoritarian monarchical rule.

3.3.1.1 Dictionnaire de l'Académie. The first edition of this dictionary, published in 1694, represents a milestone in the history of France and had a significant impact on Europe. Despite the delay, it served as a model for similar publications and academies for several years.⁴⁶

This lexicographic project was born in 1635, with the foundation of the AF by Cardinal Richelieu. Started in 1638 by invitation of Richelieu, the writing of the DAF was directed by Claude Favre de Vaugelas (1585–1650). The first edition did not appear until 1694.

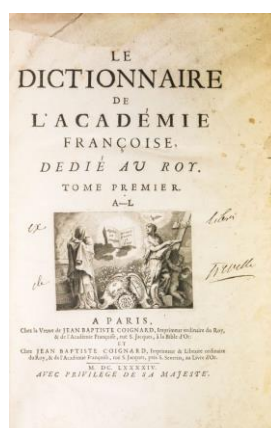


Figure 17: Title page of the *Dictionnaire de l'Académie Française*, engraved by Pierre-Jean Mariette in 1694

On 24 August 1694, a delegation from the AF presented to King Louis XIV in Versailles the first copy of the long-awaited French language dictionary, *Dictionnaire de l'Académie françoise*⁴⁷ – see Figure 17. ‘Messieurs, voicy un Ouvrage attendu depuis

⁴⁶ We must not forget that the French language was considered very prestigious at that time but also the influence that French dictionaries exercised in methodological terms. As far as language dictionaries are concerned, we can cite, in addition to DAF, Richelet’s dictionary (*Dictionnaire François*, 1680), and for encyclopaedic dictionaries, Furetière (*Dictionnaire universel*, 1690), Trévoux (*Dictionnaire universel françois et latin*, 1704–1771), and, of course, the *Encyclopédie*, by Diderot and D’Alembert (1751–1777).

⁴⁷ The orthographic form ‘françoise’ will only appear on the title page of the French dictionary since 1835.

longtemps' [Gentlemen, here it is, a long-awaited work], must have said the king, sardonically, considering the time it took to elaborate the DAF. As Rey (1989, p. 375) points out: 'a remark that could have been seen less as praise for the result than as an ironic reference to the snail's pace at which it had been achieved'. '*Enfim, Madame, toute la France va être contente*' [At last, Madam, all of France will be happy] is the famous phrase with which *Le Mercure Galant*⁴⁸ welcomed the publication of the dictionary.

The first edition comprises two volumes and includes approximately 15,000 words, classified by families with the same root. *Mots primitifs* (words which were not derived from other words) were printed in capitals and followed by derived and compound forms in small capitals.

Comme la Langue Françoisse a des mots Primitifs, & des mots Derivez & Composez, on a jugé qu'il seroit agreable & instructif de disposer le Dictionnaire par Racines, c'est à dire de ranger tous les mots Derivez & Composez après les mots Primitifs dont ils descendent, soit que ces Primitifs soient d'origine purement Françoisse, soit qu'ils viennent du Latin ou de quelqu'autre Langue. On s'est pourtant quelquefois dispensé de suivre cet ordre dans quelques mots, qui sortant d'une mesme souche Latine, ont fait des branches assez differentes en François pour estre mis chacun à part; & on s'en est aussi dispensé dans quelques autres mots dont le Primitif Latin n'a point formé de mot Primitif en François, ou a esté aboli par l'usage, & dont par consequent les Derivez & Composez sont en quelque façon independans les uns des autres; comme les mots construire & destruire qui viennent du mot Latin struere, qui n'a point passé en François. [As the French Language has Primitive words, & Derivative & Compound words, it was judged that it would be pleasant & instructive to arrange the Dictionary by Roots, that is, to put all the Derivative & Compounds words after the Primitive words from which they descend, either that these Primitives are of purely French origin, or that they come from Latin or some other language. However, we have sometimes dispensed with following this order in a few words, which, coming out of the same Latin lineage, have made quite different branches in French to be set apart; & it has also been dispensed with in a few other words of which the Primitive Latin did not form a Primitive word in French, or was abolished by use, & of which, therefore, the Derivates & Composes are in some way independent of the from each other; like the words build & destroy which come from the Latin word struere, which did not pass into French.] (DAF, 1694, s. p.)

⁴⁸ *Le Mercure Galant* (August 1694, tome 8, p. 296): <https://obvil.sorbonne-universite.fr/corpus/mercure-galant/MG-1694-08>.

For example, in Figure 18, the reader will have to look for the entry 'croistre' in order to look up the meaning of 'croissance' [growth] and 'croissant' [crescent-shaped]. Within this lexicographic article, the reader will then find the meaning of the derived words.

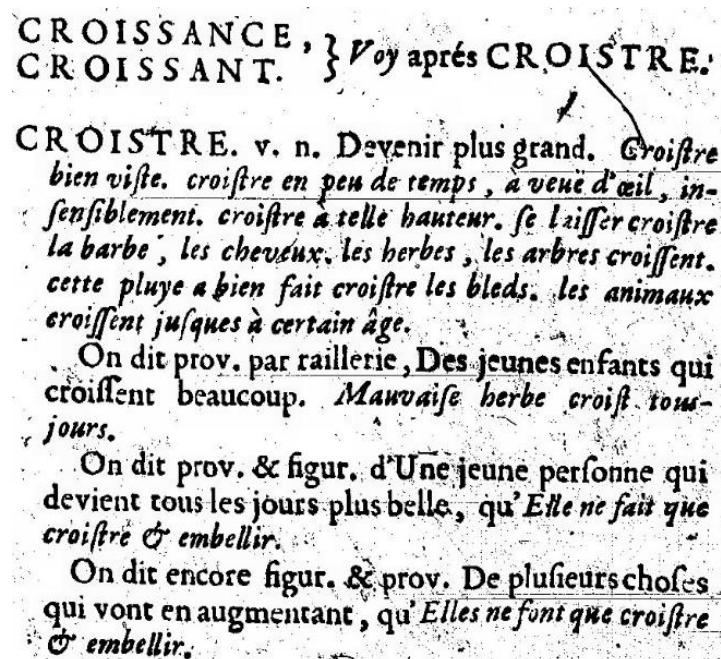


Figure 18: Le Dictionnaire de l'Académie Française, Dédié au Roy, 1st edition (DAF, 1694, p. 289)

Concerning the following editions, the second (1718) adopts the alphabetical order to facilitate the process of looking up a word – Figure 19.

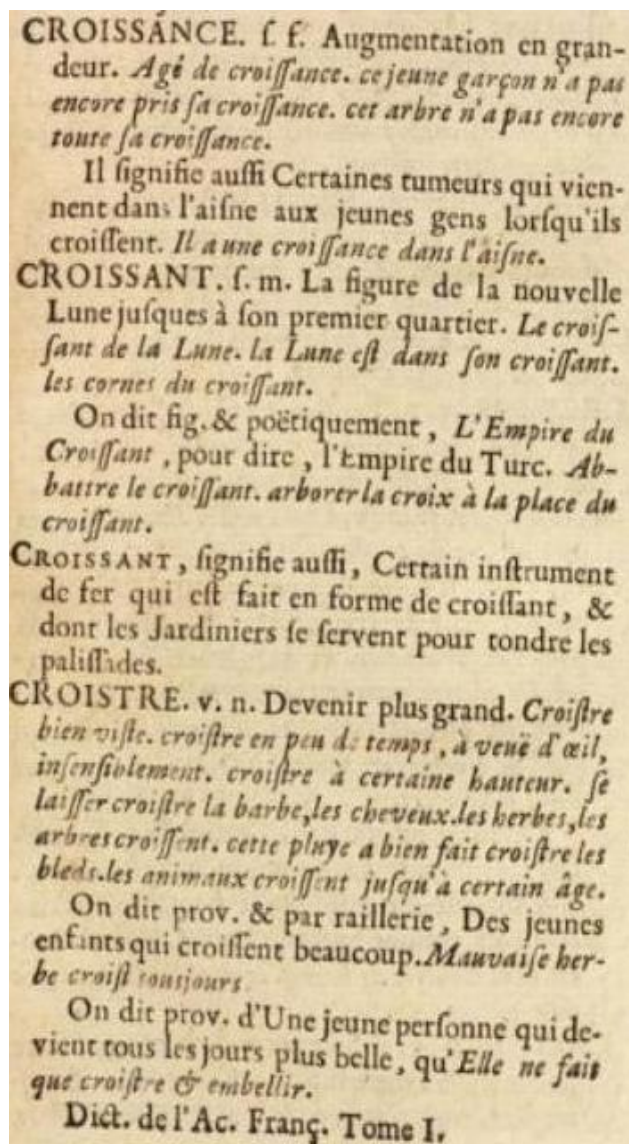


Figure 19: Nouveau Dictionnaire de l'Académie Française Dedié au Roy, 2nd edition⁴⁹

The third and fourth (1740 and 1762) editions were very progressive; the most remarkable innovation was an extensively revised orthography and the integration of the words that la '*Révolution et la République ont ajoutés à la langue*' [the Revolution and the Republic added to the language] (DAF, 1798), which emerged through a supplement to the fifth edition in 1798. In 1835, the sixth edition defined nearly 30,000 words. The seventh edition was published in 1878 and the eighth in the 1932–1935 period. At the end of the 20th century, the ninth edition was issued first in the form of fascicles, starting in 1982, then with the first volume printed in 1992 (A–Enz) followed

⁴⁹ <https://gallica.bnf.fr/ark:/12148/bpt6k12803909/f417.item.zoom>

by the second volume (Eoc–Map) in 2000. The first volume contains 14,024 words, including 5,500 new words, and the second approximately 11,500 words, including 4,000 new words (Souffi, 2009).

The AF wanted its dictionary to be made available to the public free of charge via the internet, which was achieved through the Institut National de la Langue Française (INALF, CNRS) and Analyse et Traitement Informatique de la Langue Française (ATILF) in 2001, in collaboration with the Service du Dictionnaire de l'Académie Française. It was in 1996–1997 that the Nancy laboratory digitised the eighth edition of the DAF. The first two volumes of the ninth edition were digitised in 2000–2001, and the fascicles published in the *Journal officiel de la République française* (which will constitute the material for volume 3 to be published) were posted.

The prefaces of all editions were compiled and studied by knowledgeable scholars in a reference work published in 2000 and edited by Quemada (1997).

3.3.1.2 Le Dictionnaire de l'Académie française est en ligne. In February 2019, DAF was made available to the public through a free and open-access web portal. This platform currently provides access to the dictionary's ninth (nearing completion) and eighth editions. For the first time, the public, via the internet, is privy to the whole lexicographic enterprise carried out by the *Académie* since 1694. For the launch of its new web portal, the AF first proposed the text of the ninth edition, which is almost completed and currently available for searches up to the letter S (any research concerning the end of the alphabet will be automatically redirected to the eighth edition, fully accessible). All the other editions of the dictionary will also be digitised to be made publicly available. It will then be possible to circulate from one edition to another based on the definition of a word. Additionally, the AF plans to update its web portal regularly as its work progresses. The portal has a new user-friendly interface, with responsive design, and a full hypertext navigation by a simple click on any lexical unit – Figure 20.



Figure 20: Front page of Dictionnaire de l'Académie Française (2021), AF

Finally, we have to mention the linking to several lexical data: lexical notes, i.e., notes regularly published by the AF, concerning difficulties or curiosities of the French language; spelling notes, about the French spelling reform; the official terminology database, *FranceTerme*⁵⁰, *Base de données lexicographiques panfrancophone* (BDLP)⁵¹, containing diatopic variations of the French language.

3.3.2 Real Academia Española

The primary goal of the Real Academia Española (RAE) is to watch over the changes that the Spanish language experiences, guaranteeing the essential unity of the entire Hispanic scope.

The RAE, founded in Madrid in 1713, has the official tutelage of the Spanish language, among other functions. Under the reign of Philip V, the initiative was the goal of Juan Manuel Fernández Pacheco (1650–1725), Marquis of Villena and Duke of Escalona, who created it with the purpose of '*cultivar, y fijar la pureza, y elegancia de lengua Castellána*' [to cultivate, and fix the purity, and elegance of Castilian language]

⁵⁰ <http://www.culture.fr/franceterme>

⁵¹ <https://www.bdlp.org/>

(RAE, 1715, p. 11). In other words, to establish the criteria for its correct and proper use in order to contribute to the splendour of the language. Its constitution was approved on 3 October 1714 by King Philip V, who welcomed it under his own personal, as well as official royal, protection. The RAE was modelled after the French Academy, and it has been tasked with safeguarding the correct use of the Spanish language since its inception.

Its emblem, a crucible on fire, is accompanied by the motto '*Limpia, fija y da esplendor*' [To cleanse, fix and enhance], reflecting its prescriptive nature. The symbolism might have been influenced by the Paduan academic's emblem featuring Hercules on fire (Figure 21)⁵².



Figure 21: Paduan academic's emblem and the emblem of the RAE

In its first Charter, the creation of a Spanish language dictionary was immediately established, '*el más copioso que pudiera hacerse*' [the most copious that can be made] (RAE, 1715, p. 12) – see Figure 22.

⁵² For a detailed description of the RAE emblem, see Blecua, 2006, pp. 22–25.

CAPÍTULO PRIMERO.
DEL INTENTO, Y MOTIVO DE LA
Fundación de la Academia.
ESTATUTO ÚNICO.



SIENDO el fin principal de la Fundación de esta Academia, cultivar, y fijar la pureza, y elegancia de la lengua Castellana, deserrando todos los errores, que en sus vocablos, en sus modos de hablar, ò en su construcción há introducido la ignorancia, la vana afectación, el descuido, y la demasiada libertad de innovar; será su empleo distinguir los vocablos, frases, ò construcciones estrangeras de las propias, las antiquadas de las usadas, las baxas, y rústicas de las Cortesanas, y leuantadas, las burlescas de las féricas, y finalmente las propias de las figuradas. En cuya

B 2 con-

12

consecuencia tiene por conveniente dar principio desde luego por la formación de un Diconario de la lengua, el mas copioso que pudiere hazerse: en el qual se anotarán aquellas voces, y frases, que están recibidas debidamente por el uso cortesano, y las que están antiquadas, como también las que fueren baxas, ò bárbaras: observando en todo las reglas, y preceptos, que están puestos en la planta acordada por la Academia, impresa en el año de 1713.

Figure 22: Charter of the Real Academia Española (RAE, 1715), 1st edition

Since 1993, the RAE has maintained the Instituto de Lexicografía (Ilex)⁵³ to organise the academy's lexicographic works, first by the hand of Dámaso Alonso (1898–1990). The Ilex's main job is to prepare the institution's lexicographic works, especially the DLE and, for example, the *Diccionario del Estudiante* and the *Diccionario Esencial*.⁵⁴ The DLE was never completely revised, indicating that the revision was never carried out entirely from A to Z according to specific criteria, including the specialised areas and grammar categories, among others.

Currently, the RAE and 23 other academies, one for each country where Spanish is spoken, form the Asociación de Academias de la Lengua Española (ASALE), which plays a very active and intervening role through the promulgation of standards aimed at fostering international unity in the language. The RAE has taken on the task of ensuring

⁵³ We took advantage of the ELEXIS Transnational Research Visit Grant that we received to visit the RAE and work in Ilex from 11 to 30 November 2018.

⁵⁴ Lexicographic works developed at Ilex: *Diccionario de la lengua española*; *Diccionario del estudiante*; *Diccionario práctico del estudiante*; *Diccionario esencial de la lengua española*; *Diccionario de americanismos*.

that changes in the spoken language do not break its unity maintained throughout the Spanish-speaking world. Apart from the publication of several studies on the knowledge about and research on Spanish language and literature, there are three essential publications concerning the RAE's lexicographic work: the *Gramática*, *Ortografía* and *Diccionario*.

3.3.2.1 *Diccionario de la Lengua Española*. The *Diccionario de la Lengua Española*, known as the dictionary of the Real Academia, is the broadest normative dictionary of the Spanish language.

The first Spanish academy lexicographic work, *Diccionario de la Lengua Castellana*, which came to be known as the *Diccionario de Autoridades* (illustrated by the best literary authorities), appeared between 1726 and 1739 in six folio volumes. The organisation followed the alphabetical order, and 'each of them would be followed by its derivatives and compounds as by phraseological information, as had been done with the *mots primitifs* and their derivatives in the first edition of the *Dictionnaire de l'Académie Française*' (Considine, 2014, p. 114). The expensive format of the so-called *Diccionario de Autoridades* limited its circulation; the second edition took much longer to make. Based on this work, a new version of the dictionary, created in a single volume compendium (no longer including quotes from authors), was produced in 1780. It will be the first edition of what we know today as *Diccionario de la Lengua Española* or *Diccionario de la Real Academia Española* (Figure 23).

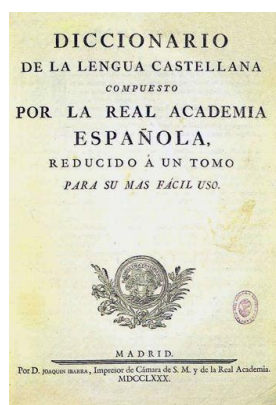


Figure 23: Title page of the *Diccionario de la Lengua Castellana*, RAE (1780)

The dictionary title was altered several times: *Diccionario de la lengua castellana reducido a un tomo para su más fácil uso* [Dictionary of the Castilian language reduced to a tome for easier use] between the first (1780) and fourth editions (1803); *Diccionario de la lengua castellana por la Real Academia Española* [Dictionary of the Castilian Language by the Real Academia Española] between the fifth (1817) and 14th editions (1914); *Diccionario de la lengua española* [Dictionary of the Spanish language] since the 15th edition (1925) till the 22nd edition (2001); and from the 23rd edition (2014) onward – which coincided with the celebration of the third centenary of the foundation of the RAE –, the acronym DLE has been used.

As seen in the preamble to the 23rd edition (DLE, 2014), the DLE then had 93,111 entries, with a total of 195,439 senses. This dictionary resulted from the collaborative work of the ASALE that brought together the lexicons used in Spain and all the other Spanish-speaking countries.

The dictionary includes common words used extensively, at least in a representative range of places where Spanish is spoken as the primary language, along with numerous archaisms and words now in disuse. The main reason for this choice was to facilitate the understanding of early Spanish literature.

3.3.2.2 *Diccionario de la Lengua Española en línea.* Until the 21st edition (1992), the medium used was paper. In 1992, the dictionary was circulated through CD-ROM and in two-pocket editions in addition to the traditional book format. Since 2001, it has also been available for an online search. The digital version of the 23rd edition was made available to the public free of charge on 21 October 2015 – Figure 24 – and the last update was in 2020 (electronic version 23.4).⁵⁵

⁵⁵ https://dle.rae.es/docs/Novedades_DLE_23.4-Seleccion.pdf



Figure 24: Front page of the *Diccionario de Lengua Española en línea* (2021), RAE

The Enclave RAE⁵⁶ is a new RAE language resource and service platform that anyone can access through a monthly or annual subscription. Although the DLE is free, the user can subscribe to this service to attain access to more linguistic tools, such as the *Diccionario avanzado*, where filters including a search by domains can be used (this type of search is not possible in the DLE) and *Diccionarios*, a module that houses all the current dictionaries of the Academy – *Diccionario de la lengua española* (DLE), *Diccionario del español jurídico*, *Diccionario del español juridical*, *Diccionario panhispánico de dudas*, *Diccionario de americanismos* and *Diccionario del estudiante* – among other modules.

3.3.3 Academia das Ciências de Lisboa

The Academia das Ciências de Lisboa (ACL), originally Academia Real das Sciencias de Lisboa due to its royal protection, was founded in 1779, during the reign of Dona Maria I. The main proponents of this academic project, D. João Carlos de Bragança e Ligne de Sousa Tavares Mascarenhas da Silva (1719–1806), second Duke of Lafões, and José Francisco Correia da Serra (1750–1823), better known as *Abade* Correia da Serra,

⁵⁶ <https://enclave.rae.es/que-es>

were influenced by Enlightenment trends and institutions that were already emerging across Europe.

The institution's emblem represents Minerva, the goddess of Wisdom and War, with the mercury rod and the shield with the Portuguese royal arms, under the inspiring sign of a verse by Phaedrus: '*Nisi utile est quod facimus stulta est gloria*' [If what we do is not useful, glory is in vain], symbolising the alliance between knowledge and royal power (Figure 25).



Figure 25: Emblem of the Academia das Ciências de Lisboa (ACL)

Since its foundation, the ACL has established that among its '*utilísimos intentos, que a composição de hum Diccionario da mesma lingua fizesse parte dos seus primeiros trabalhos*' [useful intentions, that the composition of a Dictionary of that language was part of its first works] (ACL, 1793, s. p.). A '*Planta para se formar o Dicionário*' [Plan to form the Dictionary] was presented at an academic session on 4 July 1780.

Nowadays, the plan of the first Charter, dating back to 1780, highlights the utilitarian perspective of the creation of the institution '*consagrada à glória, e felicidade pública para adiantamento da indústria nacional, perfeição das ciências, e aumento da indústria popular*' [consecrated to the public glory and happiness for the advancement of national industry, the perfection of the sciences and the increase of popular industry] ACL (1780, p. 3).

Currently, among its missions, the ACL is responsible for encouraging scientific research, stimulating the study of the Portuguese language and literature and promoting the study of Portuguese history and its relations with other countries.

Pursuant to its current Charter, the ACL remains an '*órgão consultivo do Governo português em matéria linguística*' [advisory body to the Portuguese Government on linguistic matters] (Decreto-Lei n. 157/2015, art. 5).

The lexicographic activities of the ACL are part of the responsibilities of the Instituto de Lexicologia e Lexicografia da Língua Portuguesa (ILLLP), an organisation tasked with

promover a criação e apoiar a atividade de núcleos de estudos necessários para a defesa e enriquecimento do léxico da língua portuguesa e promover a realização de colóquios e seminários, dentro das áreas da lexicologia e da lexicografia do português [promoting the creation and supporting the activity of study centres necessary for the defence and enrichment of the lexicon of the Portuguese language and fostering the realisation of colloquia and seminars within the areas of lexicology and lexicography of the Portuguese language]. (Decreto-Lei n. 157/2015, art. 20)

The proposal for the creation of the ILLLP had been approved on a plenary session of the ACL (s. d.), and the first time that the ILLLP appears enshrined in legislation is in the Decreto-Lei n. 390/87. Additionally, in 1989, an ILLLP leaflet was published by the ACL (ACL, 1987), whose section '*Actividades em curso*' [Activities in progress] refers to the 'drafting of a New Dictionary of the Portuguese Language'.

3.3.3.1 The First Attempts at Making a Dictionary. The ACL's first lexicographic works are incomplete. Its successive attempts at undertaking lexicographic projects that ended up being suspended are (in)famous – twice the Portuguese academy's dictionaries stopped at the letter A.

The first volume of ACL's dictionary, from '*a*' to '*azurrar*'⁵⁷ [to bray], is dated 1793, entitled *Diccionario da Lingoa Portuguesa* (DLP) and is an unfinished work (Figure 26).

⁵⁷ The fact that it ends with the '*azurrar*' entry led to some scathing comments. Criticisms, such as that from Alexandre Herculano in *Dama Pé-de-Cabra*, did not spare the organisers: '*O onagro fitou as orelhas e... começou a azurrar, começou por onde, às vezes, as academias acabam.*' [The onager looked at his ears and... started to bray, started where, sometimes, academies stop.]

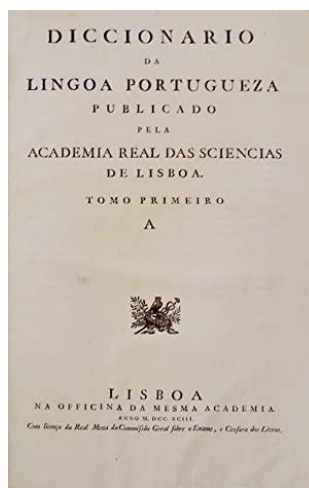


Figure 26: *Diccionario da Lingoa Portuguesa* (1793), ACL

The main advocates of this work, planned in 1780, were Pedro José da Fonseca (1737–1816), a royal professor of rhetoric and poetry at Colégio dos Nobres who had produced a Portuguese-Latin dictionary in 1771, Agostinho José da Costa (1745–1822), a royal professor of rational and moral philosophy and Bartolomeu Inácio Jorge (?–?), a professor of philosophy at Colégio das Necessidades.

As Considine (2014) observed, ‘The immediate sense which the printed dictionary gives is one of grandeur’ (p. 158). It has an introduction, a plan for the dictionary, a comprehensive list of authors from whose works the excerpts were taken and a bio-bibliographical list of authorities, perhaps the most elaborate one that has ever been prefixed in a dictionary spanning more than two hundred pages. Although it stopped at the letter A, its value is indisputable, bearing ‘*testemunho de um saber lexicográfico moderno, apoiado em reflexão teórica*’ [witness to modern lexicographic knowledge, supported by theoretical reflection] (Verdelho, 2007, p. 27), or, as Casteleiro (2008) recognizes ‘*constitui um monumento lexicográfico, pela sua riqueza, pelo seu rigor, pela sua amplitude, assim como pela metodologia inovadora que consagra*’ [it constitutes a lexicographical monument, due to its richness, rigor, breadth, as well as the innovative methodology that it enshrines] (p. 351). An important point of reference was evidently the *Diccionario de la lengua castellana*, mentioned in the preliminaries before the *Vocabolario della Crusca* and the *Dictionnaire de l’Académie*. Finally, and well summed up by Casteleiro (1981), the introduction reveals ‘*um sólido conhecimento dos*

problemas que se põem à elaboração de um dicionário [a solid knowledge of the problems that arise in the creation of a dictionary] (p. 59).

The content of the DLP, which was purist, had a normative purpose. The following excerpt (ACL, 1793, s. p.) illustrates its normative nature well:

Não intenta a Academia dar á luz debaixo deste titulo hum simples Vocabulario de palavras Portuguezas; mas fixar em geral no idioma patrio (quanto se permite nos existentes) pela autoridade dos nossos melhores Escritores, a differença dos significados em seus vocabulos, a variedade de seus usos, as suas syntaxes, frases, anomalias, elegancias. [...] [o dicionario quer] até ajudar de hum certo modo a composição, ministrandolhe cópia no socorro dos epithetos, na multiplicidade das locuções, e na frequencia dos excellentes modelos da nossa lingoagem, que a tudo, quanto fica referido, servem de confirmação. [The Academia does not intend to give birth under this title to a simple vocabulary of Portuguese words; but to broadly fix in the nation's idiom (as far as existing ones are allowed) by the authority of our best writers, the difference of meanings in their words, the variety of their usages, their syntaxes, sentences, anomalies, elegencies. [...] [the dictionary wants to] help even in the composition somehow, giving it a copy in the aid of the epithets, the multiplicity of phrases and the frequency of the excellent models of our language, which serve as confirmation, as mentioned, of everything.] (ACL, 1793, s. p.)

Quotations illustrating the different meanings were chosen according to this normative tone, and justifiably so because it was a dictionary made by an Academy.

Although this dictionary was basically synchronous, it had a retrospective flavour, as did the *Vocabolario della Crusca*, since the authorities cited about 150 authors and 500 works in all, confined to the period from the mid-14th century to the end of the 17th century. Therefore, the archaic words these authors used must be recorded.

However, despite being incomplete, the dictionary presents crucial information: grammatical classification, such as gender, number, verb irregularities and usage; indications about usage or variety; definitions; etymology; and spelling variants, to name a few. The ACL thus produced ambitious work, with each word worked out in meticulous detail.

Perhaps ambition and quality condemned the project, however. Unable to maintain the established level, the ACL could not sustain the enterprise and, hence, the

dictionary did not venture beyond the letter A nor did it become the instrument that it was promised to be. In turn, another one emerged, occupying the symbolic place of a great Portuguese dictionary utilised by those in the following decades – the *Morais* dictionary, mentioned previously.

Despite successive academic attempts, the publication of a Portuguese academy dictionary only arose once more in the 20th century. In 1976, the ACL published a new work, DLP, in a 678-page volume coordinated by Jacinto Prado Coelho (1920–1984) – at the time of the creation of the ILLLP. His plan (Coelho, 1974) foresaw the elaboration of a selective dictionary in three double volumes comprising a set of six volumes. Similarly to the first edition, this work did not venture past the letter A: from ‘a’ to ‘azuverte’ [the designation of a Timor-Leste bird].⁵⁸

No research justifies the ‘incompleteness’ of these dictionaries, but below we will try to put together why this happened⁵⁹:

(1) It is true that since its foundation, the ACL promoted the creation of a dictionary. The first project was highly ambitious, and intended to provide information about the various uses of words (Casteleiro, 1981, p. 50). From the sample of the letter A from the 1793 dictionary, we are aware of the editors’ task, the high commitment of the authors to carry out such exhaustive work and, consequently, how time-consuming such work would be.

(2) At a certain point in the 20th century the institution began to conceive the publication of an orthographic vocabulary as a priority (vocabularies were printed in 1940, 1947, 1970 and finally in 2012). Working on vocabularies, which are faster to edit, the academicians specialised in

⁵⁸ Interestingly, there was a clear academic concern to introduce another entry so as not to end with the laughable *azurrar* [to bray]. The *azuverte* entry [a bird], thus, became the last entry of the new volume.

⁵⁹ These arguments are informed by the literature (e.g., Dias, 2018; Amaral, 2012; *III Jubileu*, 1931; Ayres, 1927) and the direct contact of the author of this thesis with the ACL and its partners, especially Professors Telmo Verdelho and M. J. Lemos de Sousa. Despite having consulted the various ACL Statutes and, when appropriate, the respective Regulations, none of the documents examined, i.e., all the texts of the Statutes and Regulations from 1822 to those currently in force, refer that the ACL would be in charge of preparing a language dictionary. However, it should be emphasized that there is notice that in the first public session of the ACL, Pedro José da Fonseca presented a paper on the ‘*Composição do Dicionário da Língua*’ [Composition of the Language Dictionary], whose content is, unfortunately, unknown. The development of this topic is beyond the main objective of our study, but we wanted to leave at least this short note.

lexicographic issues (who are small in number) have always ended up not being available to dedicate themselves to a dictionary project.

(3) Funding is a relevant aspect of any scientific project. Since funds from the Portuguese State are scarce, the only viable option is to apply for funding for lexicographic projects. Thus, in 2001, the ACL managed to publish its first complete dictionary, primarily thanks to João Malaca Casteleiro's effort and commitment, since he was the one who secured funding for this publication for about 12 years.

The reasons that have determined the unsuccessful and practically unfeasible undertaking of the Portuguese academy dictionary are institutional, emphasising two very unfavourable factors: the first one stems from the traditionally austere, insufficient and unmotivating financial framework for any demanding work schedule; the second is related to the number of philologists and linguists with a place in the framework of the ACL. The ACL is open to the broadest range of knowledge and has a proportionally minimal representation of lexicographic scholars.

The many ups and downs of the project experienced and suffered since the beginning are only by-products of the actual difficulties resulting from the composition and functioning of this institution, which is not exactly an '*Academia da Língua*' [Language Academy]. After the wearying, incipient and unappreciated first volume (1793) and after the ambitious and inglorious catalogue of books to be read for the continuation of the Portuguese language dictionary published by the Academia Real das Sciencias de Lisboa (ACL, 1799), all attempts withered for more than two centuries. The attempt of 1976 was also unsuccessful. Finally, in the 21st century, more precisely in 2001, under the coordination of Malaca Casteleiro (1936–2020), the ACL finally published a complete dictionary, the DLPC.

3.3.3.2 *Dicionário da Língua Portuguesa Contemporânea.* With financial support from the Fundação Calouste Gulbenkian (FCG), in addition to other funding institutions, and more than 200 years after the publication of the first attempt, the ACL launched a complete Portuguese dictionary, the DLPC, published by Editorial Verbo in 2001. The publication was coordinated by the then Chairman of the ILLLP, João Malaca

Casteleiro, enlisting the support of the Ministry of Education, the Instituto Camões of the Government of Portugal and the FCG, gathering a vast team of linguists whose work had begun in 1988.

The dictionary was published in two volumes. The first with the letters A–F, from pages 1 to 1846; the second with the letters G–Z, from pages 1847 to 3809. Together, the two volumes have a total of 3880 pages (Figure 27). The word list of the DLPC has a total of 69,426 entries with 167,556 senses.

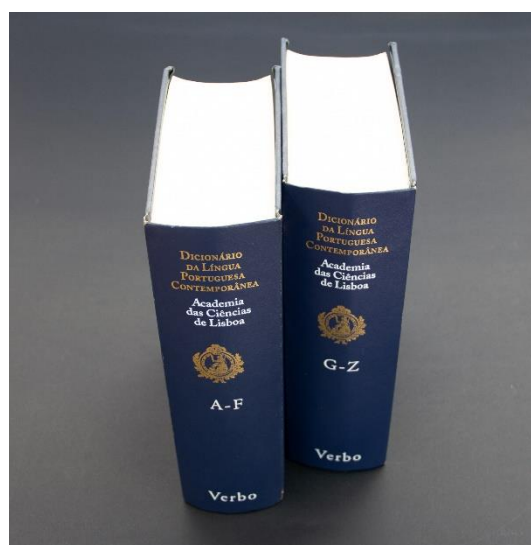


Figure 27: *Dicionário da Língua Portuguesa Contemporânea* (2001), ACL

In addition to the definition and explanation of words, the dictionary includes their etymology and phonetic transcription, presents examples of the lemma in various contexts reflecting its multiple uses (literary, scientific texts, etc.) and indicates pure or approximate synonyms.

One of the DLPC's main features, which differentiates it from other contemporary Portuguese dictionaries (e.g., GDLP; HOUAISS), is the treatment of the part-of-speech (POS) homonyms. Homonyms of the same etymological family belonging to different POS are described in each entry and distinguished by numerical superscripts on the right of the lemma (e.g., 'paleozóico¹, adj.', 'Paleozóico², s. m.')

 as an adjective and a noun. According to the editors in the Introduction, splitting entries '*justifica-se por*

razões de natureza semântica, morfológica e sintáctica' [is justified for reasons of a semantic, morphological and syntactic nature] (DLPC, p. xvii).

After this publication, Casteleiro (2008, pp. 321–322) states that the elaboration of a second edition of the DLPC would be in progress to correct errors and gaps in the first one and increase the list of lemmas from 90,000 to 95,000. However, this edition was never published.

The publication of the dictionary generated a great wave of controversy in the national public opinion, with several personalities from the Portuguese cultural scene pointing out gaps and inconsistencies.⁶⁰ The promised revised second edition was abandoned due to a disagreement between João Malaca Casteleiro and the ACL, which evolved into legal disputes and public exchanges of accusations between the mentioned parties and José Pina Martins (1920–2010), the then-Chairman of the ACL.⁶¹

3.3.3.3 Dicionário da Língua Portuguesa. The DLP, a scholarly dictionary of the Portuguese language now being developed by the ACL, is a retro-digitised dictionary created by converting the DLPC, last published in 2001. Currently, it is being prepared under the ILLLP's supervision in collaboration with researchers and invited collaborators. Between 2015 and 2016, some preparatory work for the Portuguese academy digital dictionary was performed through the ILLLP and a database was developed by a team working in natural language processing (NLP) at the University of Minho (Simões, Almeida & Salgado, 2016), which now includes the Instituto Politécnico do Cávado e do Ave (IPCA) and the Centro de Linguística da Universidade NOVA de Lisboa (CLUNL) (Salgado et al., 2019). This project is supported by a small annual Community Support Fund Portuguese National Fund (Fundo de Apoio à Comunidade – FAC) through the Fundação para a Ciência e a Tecnologia (FCT). It will be the first academy Portuguese digital dictionary.

⁶⁰ Cf. <https://ciberduvidas.iscte-iul.pt/artigos/rubricas/controversias/reflexoes-acerca-do-dicionario-da-lingua-portuguesa-contemporanea-da-academia-das-ciencias-de-lisboa/886>

⁶¹ Cf. <https://www.dn.pt/arquivo/2006/presidente-da-academia-das-ciencias-ataca-trabalho-de-malaca-casteleiro-639622.html>

3.4 Final Considerations

The lexicographic corpus employed for this research consists of the latest editions published by the last three academies mentioned above. Prescriptivism characterises academy dictionaries; this normative vein is visible in the very foundations of these institutions. All of them registered an inventory of the vocabulary normatively and authoritatively. In the first editions, one of the main goals was to record good usage; the use of words was illustrated with quotations from canonical literary authors, plainly assuming that these writers treat the vernacular language with greatest propriety and elegance. Remember: the Charter of the *Académie* stated that '*Les meilleurs auteurs de la langue françoise seront distribués aux académiciens pour observer tant les diction que les phrases qui peuvent servir de règles générales et en faire rapport à la Compagnie, qui jugera de leur travail et s'en servira aux occasions*' [The best authors of the French language will be distributed to academics to observe both the dictions and the sentences that can serve as general rules and report back to the Company, which will judge their work and use it on occasion] (Livet, 1858, p. 493). Meanwhile, the first Spanish dictionary is called *Diccionario de Autoridades*, keeping in mind that a language requires a standard based on the use of the best writers (those who, as noted in the prologue, '*han tratado la Lengua Española con la mayor propiedad y elegancia: conociéndose por ellos su buen juicio, claridad y proporción, con cuyas autoridades están afianzadas las voces*' [have treated the Spanish Language with the greatest propriety and elegance: getting to know through them their good judgment, clarity and proportion, with whose authorities the entries are consolidated] (DA, 1770), and the Portuguese bio-bibliographical list of authorities has more than one hundred folio pages (ACL, 1793, pp. LIII–CC). The concept of authority goes back to the Ciceronian *auctoritas* on whose tradition the moderns are based. As stated by Gonçalves (2002), '*A auctoritas correspondia ao mérito ou valor lingüístico-literário dos autores, sendo dilucidada ou determinada em função de um conjunto de critérios.*' [The *auctoritas* corresponded to the literary-linguistic merit or value of the authors, being diluted or determined according to a set of criteria.] (s. p.). Of course, all of this is related to the missions of the institutions described earlier. The French and Spanish dictionaries retain, perhaps more clearly, their normative role compared to the Portuguese due to political reasons beyond the scope of this research.

It is fascinating to observe each of the emblems and mottos of these academic institutions: the AF presents an image of the building, seemingly mirroring the solidity of this institution that speaks for itself; the ACL, with Phaedrus' verse emphasising the importance of the scientific contributions of each of the members of the Letters and Science classes; finally, the RAE makes its mission regarding language very clear: '*Limpia, fija y da esplendor*' [Clean, fix and gives splendor].

In Portugal, as in Europe generally at the time, the dictionary was developed due to the necessity of enhancing the linguistic and literary heritage. With only works available whose purpose was to predominantly cater to the description of Latin, there was a real need for a dictionary that expanded the vernacular nomenclature.

The digital age has opened up new paths for the production, elaboration and sharing of these resources. The three dictionaries already possess digital versions, although the ACL dictionary will only be made publicly available later next year. In fact, the availability of dictionaries on the web definitely carves out a path for further innovation, even though many of the available resources do not yet truly explore the possibilities of the digital environment, merely copying and somehow echoing the structure adopted on paper, as we will explore in Chapter 6. In order to observe the structure of each academy dictionary, their respective user guides were made available (see Annexes 1, 2, and 3).

CHAPTER 4

Usage Labels in General Language Dictionaries

*There's quite a lot of work involved in putting together
a consistent policy on labels in a dictionary.*

ATKINS & RUNDELL (2008, p. 231)

This chapter discusses the treatment of usage labels in general language dictionaries. We begin to explain the notions of deviations and restrictions to discuss the so-called marked or diasystematic marking or usage labelling in dictionaries. Recognising that labelling is a recurrent and ancient lexicographic practice, we then clarify the concept of label, the form and position in which it usually appears in dictionaries, and detailed its function. Different classifications are referred to and we emphasised the lack of agreement on the designations to classify them. Finally, we enumerate the different types of diasystematic marking with examples taken from the dictionaries under study: diachronic marking; diatopic marking; diaintegrative marking; diastratic, diaphasic and diatextual marking; diafrequentative marking; diaevaluative marking; dianormative marking; diasemantic marking; and, finally, diatechnical marking. Given the importance of the domain label to this thesis, an entire section has been dedicated to this specific topic. After describing the domain label, we identify the different types of domain labels, the difficulties that any lexicographer found when dealing with specialised data and finally, we introduce the need to build a structured organisation arguing for the benefits of establishing the concepts of superordinate domain, domain and subdomain.

4.1 Labelling Practices

Dictionary makers have long known that a definition (in the case of monolingual dictionaries) or its equivalent (in the case of bilingual or multilingual dictionaries) is not sufficient to describe a lexical item *per se*. Applying a usage label to a lexical unit implies that it moves away 'in a certain respect, from the main bulk of items described in a dictionary, and that its use is subject to some kind of restriction' (Svensén, 2009, p. 313). The need to label certain deviations (e.g., when the language register is familiar) and

restrictions (if a particular unit belongs to a domain field) originated what is currently, in general, called marking or diasystematic⁶² marking (Hausmann, 1989, p. 651). Along the same vein, Fajardo (1996/1997) mentions that labels are '*informaciones concretas sobre los muy diversos tipos de particularidades que restringen o condicionan el uso de las unidades léxicas*' [concrete information about many different types of peculiarities that restrict the use of lexical units] (p. 32).

Our interest in lexicographic markers stems from two different perspectives: (1) labels are important lexicographic mechanisms that are highly useful for lexicographers as an identification marker for specialised senses and consequently as a terminology control tool for scholars and users facilitating research, for instance, in tasks concerning the disambiguation of meaning, terminology extraction or automatic translation. Nevertheless, they are also devices that, being compact and short, often hide the complexity of the dynamic sociolinguistic, cultural and ideological processes that they intend to convey; (2) labels present a specific conceptual and infrastructural challenge for the creation of interoperable lexical resources, and their inclusion usually is not hierarchical, corresponding to simple listings of domains in alphabetical order.

Dictionaries rarely communicate the reductive nature of labels to their users or the details of the decision-making process that led them to apply certain labels. Analysing, integrating and combining high-quality lexicographic data from different sources and across different languages requires, among other things, a clear understanding of the mutual (in)compatibility of the labels used in different dictionaries around the world.

The term *usage labelling* is commonly used to designate the system concerning the restrictions and indications of constraints on the use of lexical items.

Labelling is a recurring and ancient lexicographic practice. The practice of marking lexical units and meanings with labels in English dictionaries, for example, dates back to the 18th century, a tradition established by Nathan Bailey (1691–1742), the author of several dictionaries, such as *An Universal Etymological English Dictionary*, and Samuel Johnson (1709–1784). In Richelet's (1680) dictionary (*Dictionnaire François*), we

⁶² Svensén (2009) explains this term: it 'means that we are concerned with varieties within a (language) system' (p. 315).

already find some classifiers – typographic symbols and textual markers – that complement the language description, albeit irregularly used.

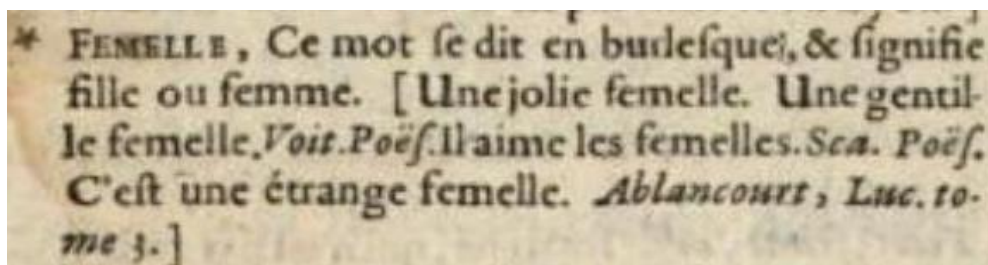


Figure 28: Entry 'femelle' [female], *Dictionnaire François* (1680), AF

The lexicographic articles marked with an asterisk, such as 'femelle' [female] in Figure 28, are the lexical units used figuratively. Those marked with a cross would be used humorously, in a burlesque or satirical fashion. Classifiers such as 'Terme de...' were used as textual markers, referring to the domain in which a lexical unit is used, as we can see in Figure 29:

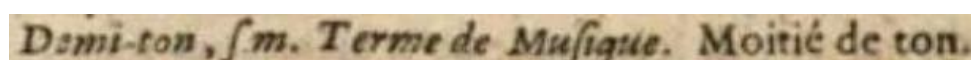


Figure 29: Entry 'demi-ton' [semitone], *Dictionnaire François* (1680), AF

Actually, current labels descend from old dictionary systems modified to standardise the options and usage of various markers. Over time, labelling mechanisms have developed to convey analytical knowledge, taxonomic will, and value judgments of a social nature roughly linked to standard and usage notions (cf. Rey, 1990). What Rey calls '*jugements de valeur*' (Rey, 1990, p. 19) reminds us of the choices the lexicographer must make, which are not always based on objective criteria⁶³ but are directly related to the use of lexical items in a specific context.

⁶³ Somehow, lexicographic discourse is never impartial or neutral.

4.2 Labels: Definition and Practices

Most dictionaries provide restrictive labels⁶⁴, but to proceed with our research, we have to clarify what a label actually is, what it indicates, what form it takes and the position it occupies within the lexicographic article, along with its respective implications, purposes and roles.

Yet another aspect we must elucidate is the concept of deviation. Languages are not monolithic entities. Any language varies according to geographic origin, level of education, formality or many other factors. 'A label is understood to be indicating a marked periphery vis-a-vis an unmarked center' (Tasovac, 2020, p. 165). The labelling system is arranged into many scales, or a 'number of part-systems' (Svensén, 2009, p. 315), with different items located at different distances from the central zone, i.e., an unmarked/neutral zone. The unmarked/neutral core of all these scales is the general language; all the others must be marked. The standard language is an unmarked centre; a regionalism is considered substandard speech, language usage that deviates from the accepted norm, so it is a marked periphery. A label always represents a zone that has a given extension between the central zone and the periphery.

4.2.1 What Is a Label, Really?

A label is a metalinguistic marker defined as an element that indicates the restricted use of a lexical item. Dictionary labels are usually indicated in paper versions through certain conventions (see 4.2.3, *Form and Position of Usage Labels*, p. 99). However, some researchers use this concept more comprehensively. In Spanish metalexicography, for example, the lexicographer Porto Dapena (2002, p. 250) considers part of speech categories to be '*marcas lexicográficas*' [lexicographic markers], attending to the idea of deviation and restrictive features: '*nosotros preferimos partir de un concepto más amplio que incluya no solo rasgos restrictivos, sino de cualquier otro*

⁶⁴ Exploring all the usage labels is beyond this doctoral project. For each of the different labels we present only a few examples of entries extracted from the DLPC: diachronic or time labels ('beque' [the back of a dress], *ant.*, 'antiguado' [old-fashioned]), diatopic or geographic labels ('parabenizar' [congratulate], *Bras.*, 'Brasil' [Brazil]), diatechnical or domain labels ('linfoma' [lymphoma], *Med.*, 'Medicina' [Medicine]), level or register labels ('paleio' [chat], *fam.*, 'familiar' [familiar]), connotative labels ('maralha' [riffraff], *Dep.*, 'depreciativo', [depreciative]) and frequency labels ('saturno' [lead], *des.*, 'desusado', [in disuse]).

tipo, como por ejemplo la pertenencia a una determinada categoría y subcategoría gramatical o semántica [we prefer to start from a broader concept that includes not only restrictive features but also any other kind of features, e.g., belonging to a certain category and grammatical or semantic subcategory]. Porto Dapena (2002, pp. 250–265) thus establishes three types of markers: grammatical (part of speech), semantic transition (e.g., figurative) and diasystematic (diachronic, diatopic, diastratic and diaphasic markers). Fajardo (1996/1997, p. 388), on the other hand, does not consider the indications of the part of speech after each lemma as a label as it is *‘fuera del concepto de marcación todo lo que es regular y constante en cada uno de los artículos del diccionario’* [excluded from the concept of marking everything that is regular and constant in each article of the dictionary]. This is a position we can agree on since we consider the restricted use of a lexical item as a preponderant identifying element of a label.

4.2.2 What Does a Label Label?

Atkins and Rundell (2008, p. 227) already asked themselves the question: ‘What does a label label?’ The answer is: multiple things. A label can refer to different pieces of information (e.g., diatechnical and diatopical markings, among others). However, lexicographers also use labels to signal the inclusion in a specific domain, immediately reducing the possibilities of interpretation and making it possible for the user to locate a specialised sense.

Moreover, in the digital age, ‘domain labels have an important role to play in lexical databases [...] where the domain label is useful in word sense disambiguation’ (Atkins & Rundell, 2008, p. 227). Considering labels aid users in searching for a specific lexical unit, they can also enable the generation of word lists containing specialised units, which in turn can be used to support automatic word sense disambiguation in lexical databases.

4.2.3 Form and Position of Usage Labels

Labels have adopted various forms. Printed editions usually implied the need to save space by condensing text, and therefore labels were generally spelt as abbreviations. Abbreviations in dictionaries are considered a by-product of the print format, which required condensed typographic solutions – literally, for economy of space.

The tradition of using abbreviations in lexicography is mentioned in the DAF webpage presentation, stating that they are often '*opaque et rebutante*' [opaque and off-putting], contrasting with their unabbreviated form in the digital version:

L'usage des abréviations constitue une tradition très ancrée dans l'histoire des dictionnaires, et renforce le côté très 'codé' de ceux-ci. Cependant, cette codification, parfois opaque et rebutante, semble peu adaptée au lecteur 'numérique' et aux usages d'aujourd'hui, ainsi qu'à l'élargissement considérable du lectorat (éducation, francophonie) que permet le support numérique. Dans cette perspective, la nouvelle mise en pages du Dictionnaire intègre la mise au long d'un certain nombre d'abréviations utilisées habituellement dans les éditions imprimées: sur les noms de domaines: BEAUX-ARTS, PHYSIQUE, ASTRONOMIE, etc.; sur les catégories grammaticales figurant à la suite de l'entrée principale; sur certaines marques de métalangue, comme 'Par extension', 'Par analogie', 'Spécialement', etc. [The use of abbreviations is a tradition deeply rooted in the history of dictionaries and reinforces the very 'coded' side of them. However, this codification, sometimes opaque and off-putting, seems ill-suited to the 'digital' reader and to today's uses, as well as to the considerable expansion of the readership (education, Francophonie) that digital media allows. From this perspective, the dictionary's new layout incorporates the expansion of a number of abbreviations usually used in print editions: on domain names: BEAUX-ARTS, PHYSIQUE, ASTRONOMIE, etc.; on the grammatical categories appearing after the main entry; on certain labels of metalanguage, such as 'Par extension', 'Par analogie', 'Spécialement', etc.] (AF, 2021)

When a dictionary is displayed on a computer screen (as opposed to the printed page), lexicographers do not have to abide by the same constraints, and some researchers have argued that abbreviations are therefore unnecessary in e-lexicography.

We will move on to exhibit a few examples where this does not always happen.

eluviação [iluvjêw]. *s. m.* (Do lat. *eluvio*, *-ōnis* 'inundação').
Geol. Depósito sedimentar, resultante da desagregação das rochas, acumulado no local de origem, ou seja, sem sofrer transporte.

Figure 30: Entry 'eluviação' [eluvium] in the DLPC (ACL)

In the DLPC, the dictionary entry "eluviação" [eluvium] (Figure 30) presents the abbreviated label *Geol.*; the lemma is a term belonging to the GEOLOGY domain. However, they can also appear in non-abbreviated forms (e.g., ARTE [art]), as shown in Figure 31 retrieved from the DLE.

musivario, ria

Del lat. tardío *musivarius* 'artífice de mosaicos', y este de *musivum* 'mosaico²' y *-arius* '-ero'.

1. **adj.** *Arte.* Pertenciente o relativo a los mosaicos. *Técnica musivaria.*
2. **adj.** *Arte.* Dicho de una persona: Que hace mosaicos. *U. t. c. s.*
3. **f.** *Arte.* Arte del mosaico.

Figure 31: Entry 'musivario' [mosaic, mosaicist, mosaicking] in the DLE (RAE)

Labels typically occupy the position before their corresponding meanings. The position of a label in a lexicographic article indicates its scope in every article and the particular meanings of lexical units or sense(s)⁶⁵:

- (1) At the lemma level, it indicates that the label applies to the lexicographic article as a whole, preceding any information related to the particular senses it conveys. In the example of Figure 32, the lexicographic article 'abcesso' [abscess] in the DLPC with its respective Brazilian spelling variant, 'abscesso', presents the abbreviated label *Bras.* for 'Brasileirismo' [Brazilianism]

⁶⁵ Sense here refers to a meaning conveyed by the lexical unit; one of the several meanings it can convey.

that is, associated with the Brazilian spelling variant, directly addressing the lemma.

abcesso, abscesso (Bras.) [ɐbsɛ́su], [ɐbʃsɛ́su]. *s. m.* (Do lat. *abscessus* 'afastamento'). *Med.* Acumulação de pus numa cavidade que se forma devido a infecção; inchaço causado por acumulação de pus. ≈ APOSTEMA. *Tomou anti-biótico porque tinha um abcesso num dente.*

Figure 32: Entry 'abcesso' [abscess] in the DLPC (ACL)

The following figure (Figure 33), featuring the entry 'escanteio' [corner] in the DLPC, illustrates the case of a label encompassing the entire lexicographic article, i.e., all the senses of the entry:

escanteio [ɛ/kẽtēju]. *s. m.* (De *canto*). *Bras. Fut. 1.* Infracção que consiste na saída da bola pela linha de cabeceira do campo de futebol, tocada por um jogador que defende; tiro de canto. ≈ CANTO. **2.** Chuto na bola com esta colocada num ângulo formado pela linha lateral e pela linha de cabeceira do campo de futebol, para cobrar essa infracção; cobrança dessa falta. ≈ CANTO. **chutar para escanteio**, pôr alguém de lado; largar.

Figure 33: Entry 'escanteio' [corner] in the DLPC (ACL)

At the sense level, by restricting the use of a certain sense, it appears as the first element following the given sense number and/or preceding the definition or descriptions in most monolingual dictionaries.

In Figure 34, the entry "cratera" [crater] has several senses, where senses 2, 3, 5 and 6 have usage labels or, more specifically, domain labels. Sense 2, *Geol.*, indicates that this sense belongs to the domain of GEOLOGY and sense 3, to industry (*Ind.*). Sense 5 belongs to the MILITARY domain (*Mil.*), while sense 6 is related to the field of ASTRONOMY (*Astr.*).

cratera [kretére]. *s. f.* (Do lat. *cratēra* < gr. κρατήρ).
 1. Recipiente ou vaso antigo, geralmente de grandes dimensões, com duas asas, onde os gregos misturavam o vinho com água, para ser servido durante as refeições.
 2. *Geol.* Abertura larga, geralmente arredondada e em forma de funil, por onde o vulcão deixa sair a lava, as matérias incandescentes ou os gases. *A cratera do vulcão.*
 3. *Ind.* Abertura na parte superior de um forno de uma fábrica de vidros. 4. Buraco grande. *Caminho cheio de crateras.* 5. *Mil.* Buraco no solo provocado pela explosão de uma granada, projectil, bomba ou outra carga explosiva.
 6. *Astr.* Depressão de origem vulcânica ou meteorítica, de dimensões variáveis, na superfície de um astro. *Cratera lunar: cratera meteorítica*, a que é produzida pelo impacto de um meteorito com a superfície de um planeta. 7. O que pode dar origem a infelicidade, calamidades, grandes acidentes... 8. Zona ou local onde alguma coisa ferve, fervilha, ou mostra grande actividade. $\approx$ AZÁFAMA.

Figure 34: Entry ‘cratera’ [crater] in the DLPC (ACL)

Additionally, the labels can be used in polylexical units (collocations or fixed expressions) or even in synonyms, as illustrated in the case of ‘pança’ [paunch, belly], sense 2, (*Fam.* or familiar) in Figure 35.

pança [pẽse]. *s. f.* (Do lat. *pantex*, *-tīcis*, pelo cast. *panza*).
 1. *Zool.* Primeiro e o maior dos compartimentos do estômago dos ruminantes. $\approx$ BANDULHO, RUMINADOURO.
 2. *Fam.* Barriga. $\approx$ BANDULHO (*Fam.*). *Encheu a pança até mais não poder e agora diz que não se está a sentir bem.*
 3. *Fam. Deprec.* Barriga grande.

Figure 35: Entry ‘pança’ [paunch, belly] in the DLPC (ACL)

4.2.4 Purpose and Role of Usage Labels

According to Svensén (2009, p. 317), a label can have two different functions: description and differentiation. The former points to the description of a particular lexical unit, providing information about it and restricting its scope of usage – this is the primary function of usage labels, i.e., marking any kind of variations from the so-called unmarked core. The other function is to differentiate between an item and other similar units.

From the user's perspective, labels can be used as signposts to locate specialised senses. However, speaking in more abstract terms, labelling can be seen as a lexicographic device for knowledge organisation in a given lexical resource (see 4.4.3 *Organisation of Domain Labels*, p. 113).

On the other hand, apart from playing a semantic role, labels also play a pragmatic role, referring to the use of a lexical item in a communicative situation that is directly dependent on the context, situation, person, etc.

In prescriptive dictionaries, the marking system imposes the appropriate or considered correct use – the idea of lexicographers as ‘censors’ (Iamartino, 2014). For Beaujot (1989), this imposition ‘*contraindre les usagers à respecter une norme socio-culturel, linguistiquement debatable*’ [compels users to respect a socio-cultural norm, linguistically debatable] (p. 91), which is controversial because the lexicographer is never an authority but the institution for whom they work can be. However, we have to recognise that ‘Dictionaries only succeed because of an act of faith on the part of their users, and that act of faith is dependent on those users believing their dictionaries both authoritative and beyond subjectivity’ (Moon, 1989, p. 59).

4.3 Classifying Usage Labels: An Overview

Researchers are acutely aware that we are still far from labelling practices that encourage consistent classification and transparent criteria for consistent labelling policies (e.g., Atkins & Rundell, 2008; Sakwa, 2011; Fedorova, 2004). Even though ‘diasystematic’ is the most recurrent term in the lexicographic literature describing the kind of information provided by dictionary labels, there is no universal agreement. As referred to above, both Svensén (2009) and Hausmann (1989) prefer the designation ‘diasystematic marking’ as a synonym for ‘diasystematic information’; Atkins and Rundell (2008) make use of the term ‘linguistic labels’, emphasising the linguistic nature of the information provided; whereas Yong and Peng (2007) opt for ‘stylistic glosses’, Landau (2001) favours ‘usage information’, while Monson (1973) speaks of ‘restrictive labels’.

A review of the existing literature (Salgado, Costa & Tasovac, 2019) has allowed us to compare different classifications of diasystematic labels. The most comprehensive classification was proposed by Hausmann (1989, p. 651), who identified 11 types of labels that were later adopted by other authors, such as Bergenholtz and Tarp (1995, pp. 131–134) and Svensén (2009, pp. 326–332). Atkins and Rundell (2008, pp. 182–186), in turn, distinguish nine types – called ‘linguistic labels’ – whereas Landau (2001, pp. 217–272) presents eight distinct types that he considers usage information, and Jackson (2002, pp. 109–115) describes seven types of usage labels.

Milroy and Milroy (1990) suggest distinguishing ‘group labels’ from ‘register labels’. The former indicates that a lexical item is restricted in its use, and the latter assists the speakers of a language in choosing the right words in the right contexts. Hausmann (1989) is the only one who integrated the label ‘diaintegrative information’ in his classification, whereas Milroy and Milroy (1990) are the only ones who adopted the term ‘diafrequentative information’. All the other researchers omit these labels from their classifications.

A survey of the different classification proposals with the different types of marking can be found in Table 1.

Hausmann (1989) adopted by others	Milroy and Milroy (1990)	Landau (1989)	Jackson (2002)	Atkins and Rundell (2008)	Criterion	Examples
diachronic	temporal	currency/temporality	history	time	Time	<i>arcaico, neologismo</i>
diatopic	geographical	regional/geographic variation	dialect ^a	region dialect	Place	<i>América, regionalismo</i>
diaintegrative	–	–	–	–	Nationality	<i>Latim, Inglês</i>
diamedial	–	style, functional variety/ register	–	–	Medium	<i>spoken</i>
diastatic	–	restricted or taboo sexual scatological usage and slang	status	slang and jargon offensive terms	socio-cultural	<i>calão, gíria</i>
diaphasic	register	style, functional variety/ register	formality	register	Formality	<i>familiar,</i>
diatextual	–	style, functional variety/ register	–	style	Text type	<i>literário, poético</i>
diatechnical	field	technical or specialized terminology	topic or field	domains	Speciality	<i>anatomia, desporto, medicina</i>
diafrequentative	frequency	–	–	–	Frequency	<i>pouco usado, raro</i>
diaevaluative	–	insult style, functional variety/register ^b	effect	attitude	Attitude	<i>eufemismo, irónico</i>
dianormative	–	status or cultural level	disputed usage	–	Normativity	<i>non-standard</i>

^aAlthough using the term “dialect”, Jackson says that “dialect labels refer to geographical restrictions, and we can take this to include both national varieties and regional dialects within a national variety” (Jackson 2002, p. 110)

^bLandau (1989) divides the “diaevaluative information” in two groups: those that denote a positive connotation (*humorous*) belonging to “style, functional variety, or register”, and those of negative connotation (*offensive*) belonging to this classification and designated as an insult

Table 1: Classifications of diasystematic information proposed by different researchers (retrieved from Salgado, Costa & Tasovac, 2019)

Despite all these classification efforts, none of these authors presents rules or explanations on how to represent diasystematic information in dictionaries, which would be of great use to a lexicographer. The existing literature on lexicographic usage labels and the mapping represented in Table 1 above exemplify, above all, a lack of agreement on the designations used to classify them. These various designations are relevant as they imply different conceptualisations of the processes or categories they signify. For instance, do temporal labels describe a lexical unit's 'currency' (Landau, 2001), 'history' (Jackson, 2002) or 'time' (Atkins & Rundell, 2008)? What does it mean when an author states that diaevaluative labels describe the 'effect' of lexical units (Jackson, 2002) instead of the speaker's 'attitude' (Atkins & Rundell, 2008)? It would be difficult to answer these questions based on the current literature because metalexicographers, as a rule, do not provide explicit definitions of their classification types, just as lexicographers fail to provide explicit definitions of the usage labels themselves.

We will now explore the different types of marking that create restrictions on the use of certain lexical units in the contexts in which they occur in more detail. We present the definitions (Salgado, Costa & Tasovac, 2019) for each usage label type to better understand their application and as the first step towards harmonising and standardising usage labels in dictionaries.

4.3.1 Diachronic Marking

Diachronic marking refers to the time dimension and associates a lexical item with a specific period in a language's history. In general, these markers are temporal labels that represent a chronological scale in which the archaisms and neologisms are at the extremes. These labels identify the use of a given lexical unit on a scale from old (archaisms) to new (neologisms). An example of an archaism could be '*haut-de-chausses*' [breeches] (Figure 36) in DAF, marked with the label '*Anciennement*'.

'HAUT-DE-CHAUSSES (*h* initial est aspiré) **nom**
masculin (pluriel *Hauts-de-chausses*).

xvi^e siècle.

Anciennement. Partie du vêtement masculin qui allait de la ceinture aux genoux. *Porter le pourpoint et le haut-de-chausses.*

Figure 36: Entry 'haut-de-chausses' [breeches] in the DAF (AF)

4.3.2 Diatopic Marking

Diatopic marking refers to the geographic dimension and associates a lexical item with a language community of speakers. In the centre, the standard language remains unmarked in dictionaries; in the periphery regionalisms, dialect units are marked. These labels identify the place or region where a lexical unit is predominantly used. However, some dictionaries, instead of identifying a specific place, identify whether the lexical unit is generally used in every geographic area or not (e.g., *regionalismo*). In the following figure (Figure 37), the lemma "banana" is an Americanism indicated by the abbreviated geographic labels Arg. [Argentina], Col. [Colombia], Ec. [Ecuador] Par. [Paraguay], Urug. [Uruguay], corresponding to the Castilian "plátano", plant and fruit (senses 1, 2, 4).

banana

Voz del Congo.

1. f. *Arg., Col., Ec., Par. y Ur.* **plátano** (|| planta musácea).
2. f. *Arg., Bol., Ec., Par. y Ur.* **plátano** (|| fruto).
3. f. *Col.* Nombre de una variedad de confites.
4. f. *Hond. y Ur.* Variedad del **plátano** (|| fruto).
5. f. *despect. coloq. Ur.* Persona tonta, boba. **U. t. c. adj.**

banana split

1. **m.** Postre hecho principalmente de helado y plátano.

Figure 37: Entry 'banana' [banana] in the DAF (AF)

4.3.3 Diintegrative Marking

Diintegrative marking refers to the degree of integration of a lexical unit in the native lexicon of a language. Although native lexical units are not, as a general rule, marked, some dictionaries mark loanwords (we have to disagree with Svensén [2009, p. 327], who states that foreign words are marked, and loanwords are unmarked). In the DLPC, for instance, the “icebergue” entry, as shown in Figure 38, is a loanword and has the *Angl.* label (to identify this lexical unit as an anglicism). Sometimes this information and the information given in the field of etymology overlap.

iceberg [ájsibɛrg]. *s. m.* (Ingl.). V. *icebergue*.
icebergue [isibɛrgi ~ ajsibɛrgi]. *s. m.* (Do ing. *iceberg*, do nor. *isberg* ‘montanha de gelo’). *Angl.* Massa de gelo flu-tuante de grandes dimensões, separada de um glaci- ar po- lar, a maior parte submersa, impelida pelas correntes ma- rítimas. *A parte do icebergue que emerge da água chega a atingir 200 m. de altura. Depois de ter colidido com o ice- bergue, o barco afundou-se. ponta⁺ do icebergue.*

Figure 38: Entries ‘iceberg’ and ‘icebergue’ [iceberg] in the DLPC (ACL)

4.3.4 Diastratic/Diaphasic/Diatextual Marking

Diastratic marking usually includes all information related to style level in a broader sense. Therefore, we refer to several dimensions of usage corresponding to different labels, a label that identifies the typical use of a lexical unit in a particular discourse, such as literary or poetic language, formal as opposed to informal language or the socio-cultural label, which identifies the use of a given lexical unit by particular social groups and/or in certain types of communicative situations depending on their level of formality, such as the opposition formal versus informal.

4.3.5 Diafrequential Marking

Diafrequential marking is related to the frequency of the occurrence of a given lexical unit. As a rule of thumb, dictionaries tend to mark words that are either very frequent or rare, based on an often-subjective assessment, which can be founded on a

quantitative analysis of a corpus or a lexicographer's intuition. Found in numerous dictionaries, these labels, termed 'frequency labels', determine a lexical unit's relative rate of occurrence in a given textual context.

4.3.6 Diaevaluative Marking

Diaevaluative marking refers to the attitude dimension of the speaker. We call it an attitude label as it identifies the speaker's subjective point of view, be it positive or negative, regarding the object referred to by a given lexical unit. The values can be humorous, ironic, depreciative, etc. For example, in DLE, 'friolero' [*chilly* as an adjective, *trifle* as a feminine noun and ironically something that is clearly not a trifle, but the opposite of it, like a boatload of money] in its ironic sense (sense 3) is the opposite of the denotative sense recorded in 2 (Figure 39).

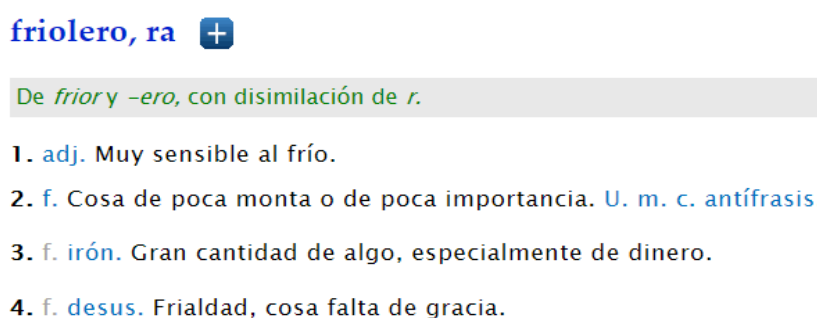


Figure 39: Entry 'friolero' [sensitive to the cold] in the DLE (RAE)

4.3.7 Dianormative Marking

Dianormative marking refers to the notion of correct and incorrect. The normativity label identifies the use of a given lexical unit, where acceptability is assessed regarding its correctness. For example, 'círculo' [circle]⁶⁶ (INFOPÉDIA), in sense 2 is marked as '*uso indevido mas generalizado*' [misused but widespread], since circle should not be taken as synonymous with circumference. Some authors, viz. Svensén (2009, p. 331), include labels such as 'Anglicism' in this group. However, the use of such labels

⁶⁶ <https://www.infopedia.pt/dicionarios/lingua-portuguesa/c%C3%ADrculo>

could merely serve to signal the language of origin of the word as we saw in the case of the ‘icebergue’ entry in the DLPC in Figure 38.

4.3.8 Diasemantic Marking

Following Hausmann’s (1989, p. 651) classification, we added a new type of marking, the diasemantic marking, to encompass any semantic extension of a particular lexical unit’s sense. However, figurative or metaphorical meanings are not strictly related to the labelling system; for practical reasons the information has the form of labels, the same function and same position.

PRINTEMPS nom masculin

xii^e siècle, *printans* ; xiii^e siècle, *printemps*. Composé de l’ancien français *prin*, « premier », et de *temps*.

- Saison qui s’étend du 20 ou 21 mars au 21 ou 22 juin et se caractérise, dans l’hémisphère Nord, par un adoucissement des températures et par la renaissance de la végétation. *L’équinoxe de printemps. La célébration de la pâque se situe au printemps. Une belle journée de printemps. « Le Printemps », tableau de Botticelli.*
- Expr. proverbiale empruntée à Aristote. *Une hirondelle ne fait pas le printemps*, un premier signe favorable ne garantit pas une fin heureuse.
- **CUISINE** *Rouleau de printemps*, mets vietnamien composé d’une galette de riz farcie de légumes crus et de viande ou de crevettes, qui se consomme frais.
- Par métonymie. Parfois plaisant. *Année. Elle compte seize printemps. Il va fêter ses quatre-vingt-dix printemps.*
- Fig. Temps de la jeunesse. *Il est encore dans le printemps de sa vie, au printemps de son âge.*
- **HISTOIRE** *Printemps des peuples*, nom donné aux révolutions européennes de 1848, qui marquèrent l’émergence du droit des peuples à disposer d’eux-mêmes. *Le printemps de Prague*, mouvement de libéralisation du régime communiste de Tchécoslovaquie, au printemps de 1968, qui fut écrasé au mois d’août de la même année par les armées du pacte de Varsovie.

📖 Titre célèbre : *Le Sacre du printemps*, ballet d’Igor Stravinski (1913).

Figure 40: Entry ‘printemps’ [spring] in the DAF (AF)

In Figure 40, we are interested in highlighting the meaning that refers to ‘Année’ [years of age] or ‘Temps de la jeunesse’ [youth]. In the DAF, there are two different labels, ‘Par métonymie’ [By metonymy] and ‘Fig.’ [figurative], which correspond to diasemantic marketing.

4.3.9 Diatechnical Marking

Diatechnical information/marking indicates that a given unit belongs to a particular domain. Bearing in mind that knowledge is complex, Sager (1990) states, ‘In practice, no individual or group of individuals possesses the whole structure of a community’s knowledge; conventionally, we divide knowledge up into subject areas, or disciplines, which is equivalent to defining subspaces of the knowledge space.’ (p. 16). In sum, a domain is a ‘field of special knowledge’ (ISO 1087, 2019, p. 1). This definition has the advantage of being transparent and sufficiently comprehensive.

In the universe of the labelling system commonly used in lexicography, the labels assigned to these specialised senses are called ‘domain labels’, which are defined as a ‘marker which identifies the specialised field of knowledge in which a lexical unit is mainly used’ (Salgado, Costa & Tasovac, 2019). Given its significance in the present work, this label will be analysed in more detail in the next section.

4.4 The Domain Label

The designation domain label is not consensual. Atkins and Rundell (2008), referring to ‘linguistic labels’, classified specialised vocabulary as ‘domains’ (p. 182); they are termed ‘field labels’ according to Verkuyl, Janssen and Jansen (2003, p. 7), ‘*marcas técnicas*’ by Fajardo (1994; 1996/1997), ‘*marca de materia*’ (Martínez de Sousa, 1995), ‘*marca terminológica*’ in Lara (1997), ‘*marcas temáticas*’ in Estopà (1998), ‘field label’ (Hartmann & James, 1998/2002), ‘*marca de especialidad*’ (Nomdedeu Rull, 2008), or ‘*diatechnical information/marking*’ (Hausmann, 1989; Svensén, 2009). In our research framework, we prefer the term ‘domain label’ because it seems to be a transparent and recognisable designation for lexicographers, as well as a beacon for terminologists. Therefore, we use ‘label’ to indicate abbreviations (e.g., *Geol.*) collected in our lexicographic corpus and ‘domain’ to mention the designations of each of the abbreviations written in full GEOLOGIA [GEOLOGY].

As a general rule, a domain label informs the user that a lexical item does not belong to the general language, restricting a certain meaning to the field of activity or knowledge. These labels are used ‘*para señalar el léxico temáticamente especializado*,

en contraposición al léxico común’ [to signal the thematically specialised lexicon in contrast to the common lexicon] (Estopà, 1998, p. 1) and are generally expressed in the form of abbreviations (remember the economy of space rationale in the paper format).

Regarding a diachronic study of domain labels in the RAE dictionaries, Paz Battaner (1996) considered that *‘la presencia de marca temática parece aleatoria en la tradición académica, y en todas las que la siguen*’ [the presence of a thematic label seems random in academic tradition, and in every other tradition that follows it] (p. 104). Nevertheless, strictly speaking, we have to ask what the domain label is for and what it intends to mark.

Domain labels serve multiple functions:

- aiding lexicographers by providing specific information that identifies specialised lexica in general language dictionaries, which can serve as terminology-control mechanisms;
- facilitating user searches used as signposts by grouping lexical items according to a field, enabling the user to determine beforehand whether the complete lexicographic article is relevant for them;
- assisting end user word sense disambiguation tasks;
- advancing terminology extraction in diverse languages;
- enhancing machine translation and NLP projects.

In our understanding, the use of domain labels is intended not so much to point out a specialised sense in a general language dictionary but to further clearly distinguish the different meanings, which is very useful for polysemic entries. Their function is essentially representational and distinctive of meanings (which is very useful in bilingual or multilingual dictionaries in multiple equivalence cases, so that the user can quickly locate a term used in a given field). Despite this utility as a distinctive descriptor of meanings, dictionaries also mark monosemic entries. Therefore, we agree with Lépinette (1990) when emphasising the specificity of this label functioning only as *‘la spécification d’un domain de reference*’ [the specification of a reference domain] (p. 502).

Candel (1979, p. 100) identified two main functions in the attribution of a domain label: (i) the semantic criterion that *'peut signifier que la définition du terme implique une appartenance thématique'* [can mean that the definition of the term implies a thematic similarity] linked to the notion (concept) and class of objects to which the word corresponds; (ii) the pragmatic criterion, when it refers to a situation that may concern signifieds or referents, indicating that the term's usage is linked to a milieu. The semantic function assumes information related to the concept and establishes relationships with a particular activity or field of knowledge. Conversely, its pragmatic function points to a situation where the lexical item's concept can be used and related to the term of a given domain.

4.4.1 Types of Domain Labels

A domain can be the designation of a field where a specific knowledge area is developed (GEOLOGY) or the specific object of the knowledge area (SHOEMAKING). Lexicographers often engage in subjective assignments in accordance with a certain tradition they subscribe to (Ptaszyński, 2010, p. 413). For instance, the dictionaries we analysed contained labels for domains such as 'CHAPELARIA/CHAPPELERIE' [millinery] and 'VENATÓRIO/VÈNERIE' [hunting] (DLPC, DAF) but not for MANAGEMENT or TOURISM.

According to Rey (1979, pp. 85–86), who identified two fields, theoretical and technical, the theoretical domains (philosophy, science, etc.) allow the apprehension of reality to derive knowledge from it. In contrast, technical domains act on reality that the author views as pragmatic domains. This classification can be found in many language dictionaries, where a domain label has the function of delimiting the use of a lexical unit and whose purpose is to restrict its meaning. The quantity and diversity of fields is a fact in any dictionary, combining theoretical and technical fields, activities, sectors and others. Svensén (2009, p. 50) argued that some fields are more represented in general language dictionaries since their terminologies are more common.

Rey (1985, p. 5) believes that a language dictionary must mark the linguistic nature of the term, which can be assigned *'à un registre d'usage marqué (comme technique, scientifique, didactique, et éventuellement par une marque plus précise –*

nom d'une technique ou d'une science') [to a marked usage register (such as technical, scientific, didactic and possibly by a more precise marking – the name of a technique or a science)].

Other scholars have distinguished between (1) domain of knowledge and (2) domain of activities or (3) sector of activities. There are those who consider a domain of knowledge as '*un savoir constitué, structuré, systématisé selon une thématique*' [a knowledge constituted, structured, systematised according to a topic] (De Bessé, 2000, p. 184). In this structured and systematised knowledge, we find '*les sciences pures, les sciences dures, les sciences molles, les techniques, les systèmes conceptuels dépendant d'un discours*' [pure sciences, hard sciences, soft sciences, techniques, concept systems depending on a discourse] (De Bessé, 2000, p. 184) (e.g., ZOOLOGY, LAW, PHILOSOPHY, GEOLOGY). By contrast, a domain of activities '*permet d'identifier un champ d'action, un ensemble d'actes coordonnés, une activité réglée, une pratique*' [allows one to identify a field of action, a set of coordinated acts, a regulated activity, a practice] (De Bessé, 2000, p. 184) and consists of '*un ensemble de procédés bien définis destinés à produire certains résultats*' [a set of well-defined processes intended to produce certain results] (De Bessé, 2000, p. 184).

Another distinction is made between '*domain propre*' [proper domain] (Pavel & Nolet, 2001, p. 5) or '*domaine d'origine*' [domain of origin] (Depecker, 2003, pp. 146–147) and '*domaine d'application*' [domain of application]. The proper domain or domain of origin, is '*le domaine dans lequel est créé le concept auquel renvoie le terme*' [the domain in which the concept to which the term refers was created] (Depecker, *ibidem*), and the domain of application is le '*domaine dans lequel le concept correspond[ant] [au] terme est utilisé*' [the field in which the concept that corresponds [to] [the] term is used] (*ibidem*).

Therefore, with these authors, we must recognise that the concept of domain is neither entirely satisfactory nor consistently operative insofar as it is only a pure artefact.

4.4.2 The Domain Label as a Challenging Lexicographic Issue

The real problem is that reference works have different criteria. For instance, the DLE do not label certain lexical units that can be assigned to certain specialised fields, and sometimes lexicographers do not apply any label when the subject field is evident from the definition.

Meanwhile, assigning domain labels has always been a challenging issue for any lexicographer. They are faced with difficult decisions such as: What domain label should I assign to this specialised meaning? Should I assign a domain label to a meaning that has lost its status as a term? This last decision results from the fact that the term may have gone through a process of determinologisation (see Chapter 6, p. 124), thus losing its status as a term. These are decisions that the lexicographer makes in a very solitary way.

In addition to the domain label, it goes without saying that linguistic formulae used in the definitions, contexts and other indicators generally point to specialised meanings.

4.4.3 Organisation of Domain Labels

Atkins and Rundell (2008) argued that instead of conceiving 'a totally flat non-hierarchical list of domains, it is more practicable to try to build a domain list with a certain hierarchical structure' (p. 184). Applying previously organised hierarchical structures is gainful when composing and editing a lexicographic resource because it helps the lexicographer control the terminological data.

Assuming that the unmarked lexicon belongs to the general lexicon, as we shall see, is a controversial matter. The criteria differ from dictionary to dictionary. In fact, not every lexical unit that can be classified as a term is actually marked; it is unclear if this is due to forgetfulness or the adoption of different criteria. In most cases, we can only limit ourselves to making assumptions, given the lack of introductory and explanatory texts on the methodology and criteria followed. On the other hand, some domains seem to be segmented, allowing the identification of some overlapping areas, which mainly result from the use of lexicographic material.

A domain is always an organised set of concepts (Depecker, 2003, p. 145; Cabré, 1999, p. 99). This structure, which is classically represented under the tree shape of the domain, is generally divided into substructures, which in turn are divided into other substructures of finer levels, etc., so that each substructure refers to a particular subdomain (Cabré, 1998, p. 174). Thus, we believe that it would be convenient to establish hierarchical concepts as a way to organise the domains registered in lexicographic resources. In this sense, we argue for the benefits of establishing three possible levels (superdomain, domain, subdomain, see Chapter 7). Therefore, 'If a domain is subdivided, the result is again a domain' (ISO 1087 (2019, p. 1). For instance, we can consider FOOTBALL, which can be integrated into a generic domain: SPORTS. The same procedure can be considered for other sports integrated into dictionaries. Entries related to HANDBALL, BASKETBALL, VOLLEYBALL, etc., can still be classified under the SPORTS domain. In terms of interoperability, the elaboration of a taxonomic classification for domain labels is advantageous: it allows labels to be similar in different dictionaries and enables their reusability.

Concerning domain labelling, in Chapter 6, we will analyse the flat (non-hierarchical) lists of domain labels that appear in the dictionaries under study. Then, in Chapter 7, we conceptually structure and organise the selected domains (GEOLOGY and FOOTBALL). We consider three possible levels (superdomain, domain, subdomain) to better structure and organise terminological data in general language dictionaries and improve search engines. Lastly, in Chapter 9, we highlight and discuss the importance of having hierarchical domain labels in TEI.

CHAPTER 5

Terms in General Language Dictionaries

Personne ne met en doute la nécessité de la présence des technolectes dans les dictionnaires à l'usage de tous.

BOULANGER & L'HOMME (1991, p. 26)

In the present context, it would be inconceivable to imagine a general language dictionary that did not include terms; however, it was not always like this. There was some hesitation, discussion, disturbance and even resistance, especially in academic circles, which the passage of time and the evolution of society can justify. This chapter begins with an overview of this discussion about including terms in monolingual general dictionaries, focusing on the academy dictionaries under study. Then, we highlight the source of lexical renewal represented by terms in current lexicographic works, justifying the interest and concerns of our research. We progressively move forward to clarify some of the key concepts of this doctoral research project, namely the *term*, which necessarily brings the *concept* along. Because we will deal with specialised lexical units in a particular field of knowledge, the concept of *domain* will be explored again. The delimitation of the domain and its organisation is an essential task in terminological work, which supports the close link between *term* and *definition*. We highlight the recommendations of ISO standards 1087 and 704 concerning the formulation of definitions, emphasising the guidelines regarding the *intensional definition*⁶⁷ that should be used whenever possible.

5.1 Terms in General Dictionaries: To Include or Not To Include?

Macrostructurally speaking, the inclusion of terms in general dictionaries is a long-standing tradition (Walczak, 1991, p. 126). However, centuries ago, when the debate surrounded the inclusion of terms in language dictionary projects, opinions were

⁶⁷ An *intensional definition* is defined as 'definition (3.3.1) that conveys the *intension* of a *concept* by stating the immediate *generic concept* and the *delimiting characteristic(s)*' (ISO 1087, 2019, p. 7).

divided. As this research focuses on dictionaries published by academies, we will dedicate some words to the inclusion of terms in those dictionaries.

We begin by referring to the first of the academy dictionaries – the *Académie* dictionary. Rey (1984/2001), in the preface to the *Grand Robert de la langue française*, summarises the doctrine followed by French academicians in the elaboration of their dictionary: ‘*définir, par des choix dictés par le bon goût, un usage du français excluant les variétés régionales – surtout méridionales –, les archaïsmes, les vulgarismes, ainsi que les termes ‘d’art’, c’est-à-dire scientifiques et techniques*’ [to define, by choices dictated by good taste, a use of French excluding regional varieties – especially southern ones – archaisms, vulgarisms, as well as the terms of ‘art’, i.e., scientific and technical] (p. xviii). We thus observe that, according to the methodology applied, the DAF 1st edition would exclude terms from its lemma list; that is, it rejected a general trend at the end of the 17th century towards encyclopedism. It is, above all, a reflection of the dominant ideology in a monarchic society: ‘*il y avait d’une part le langage de la cour et des écrivains bien en cour, d’autre part le langage des métiers et des sciences qui ne relevait pas de la culture de l’honnête homme*’⁶⁸ [on the one hand, there was the language of the court and of the writers, which was very much alive; on the other hand, the language of the trades and the sciences, which did not belong to the culture of the honnête homme] (Guilbert, 1973, p. 5). Furthermore, this will be the point that dictates the distance between Antoine Furetière (1619–1688) and his academic confreres.

Furetière, also a follower of the *bon usage*, was equally interested in accurately describing the meanings designated by words specifically having to do with scientific notions and rational knowledge. Pierre Bayle (Bray, 1990) explains in the preface to Furetière’s *Dictionnaire universel* that ‘*le langage commun n’est icy qu’en qualité d’accessoire*’ [common language is here only as an accessory] (p. 1800). The description of terms is its purpose: ‘*c’est dans les termes affectez aux Arts, aux Sciences, & aux professions, que consiste le principal*’ [the most significant importance is in the terms assigned to Arts, Sciences and occupations] (Furetière, 1685, p. 4). Furthermore, this

⁶⁸ In the French 17th and 18th centuries, the figure of a *honnête homme* [honest man] represents a man with a broad general culture and the social qualities necessary to make him pleasant by demonstrating a social ease in accordance with the ideal of the moment.

concern is evident in the complete title of Furetière's work: '*contenant généralement tous les MOTS FRANÇOIS tant vieux que modernes, & les termes de toutes les SCIENCES ET DES ARTS*' [generally containing all FRENCH WORDS, both old and modern, and the terms of all SCIENCES AND ARTS]. Furetière, as early as 1685, had criticised the usefulness of the academy dictionary. He returns to it many times in his *Factums*:

Les termes des Arts & des Sciences sont tellement engagez avec les mots communs de la Langue, qu'il n'est pas plus aisé de les separer que les eaux de deux rivières à quelque distance de leur confluent. [The terms of the arts and sciences are so interwoven with the common words of the language that it is no easier to separate them than the waters of two rivers at some distance from their confluence.] (Furetière, 1685, p. 19)

In Furetière's view, the academy dictionary would have little use without including terms; he thus defends a nomenclature as comprehensive as possible. Therefore, this is the major difference between Furetière's dictionary and the guidelines of the *Académie* dictionary. When the DAF was published in 1694⁶⁹, the Prologue stated:

L'Académie en bannissant de son Dictionnaire les termes des Arts & des Sciences, n'a pas creu devoir estendre cette exclusion jusques sur ceux qui sont devenus fort communs, ou qui ayant passé dans le discours ordinaire, ont formé des façons de parler figurées; comme celles-cy, Je luy ay porté une botte franche. Ce jeune homme a pris l'Essor, qui sont façons de parler tirées, l'une de l'Art de l'Escrime, l'autre de la Fauconnerie. On a usé de mesme à l'esgard des autres Arts & de quelques expressions tant du style Dogmatique, que de la Pratique du Palais ou des Finances, parce qu'elles entrent quelquefois dans la conversation. [The *Académie*, by banning the terms of the arts and sciences from its dictionary, did not think it necessary to extend this exclusion even to those that have become very common, or have gone into ordinary discourse, have formed figurative ways of speaking; like these, *Je luy ay porté une botte franche. Ce jeune homme a pris l'Essor*, which are specific ways of speaking, one of the art of fencing, the other of falconry. We have used the same with regard to the other arts and a few expressions both of the dogmatic style and of the practice of the palace or of finances, because they sometimes enter the conversation.] (DAF, 1694, s. p.)

⁶⁹ Thomas Corneille (1625–1709), a French academician, publishes *Dictionnaire des Arts & des Sciences* in the same year.

In this way, the *Académie* justifies the exclusion of terms that are only used in specialised contexts and includes those that have become widespread in everyday discourse. Johnson (1747) also references this point in his Preface:

The academicians of France, indeed, rejected terms of science in their first essay, but found afterwards a necessity of relaxing the rigour of their determination; and, though they would not naturalise them at once by a single act, permitted them by degrees to settle themselves among the natives, with little opposition; and it would surely be no proof of judgment to imitate them in an error which they have now retracted, and deprive the book of its chief use, by scrupulous distinctions. (Johnson, 1747)

The first edition of the Spanish academy dictionary, the *Diccionario de Autoridades* (DA, 1770), makes some references to terms. In the Prologue of the first edition of this dictionary, it is explained that the work is composed of '*todas las voces de la Lengua, estén, è no en uso, con algunas pertenecientes à las Artes y Ciéncias*' [all the entries of the language, which are or are not in use, with some belonging to the Arts and Sciences' (DA, 1770, p. II, parag. 4). The RAE justifies its moderated inclusion with the intention to publish a terminological dictionary – which would not be published: '*de las voces propias pertenecientes à las Artes liberales e mecánicas há discorrido la Académia hacer un Diccionario separado, quando este se haya concluido: por cuya razón se ponen solo las que hana parecido mas comunes y precisas al uso, y que se podían echar de menos*' [of the entries belonging to the liberal and mechanical arts, the Academy discussed the possibility of making a separate dictionary after this has been concluded: for that reason, only those that seemed more common and necessary, and that could be missed were included] (DA, 1770, p. v, parag. 8) – i.e., an analysis had to be conducted to determine whether a term should be included in a general language dictionary or if it should only be included in specialised dictionaries, which also denotes a certain concern with the selection criteria.

As already noticed by Paz Battaner (1996, p. 6), Spanish academy dictionaries use the expression '*voz de...*' [entry of...] to point to terms. See, for example, in Figure 41, *Agr. – Voz de la Agricultura*, or *Mit. – Voz de la Mitología*.

EXPLICACION

De varias abreviaturas que se ponen en este Diccionario para denotar la calidad y censura de las voces.

adj.	Adjetivo.	Mancj.	Voz del Manejo de los caballos.
adv.	Adverbio.	Maq.	Voz de la Maquinaria.
adv. l.	Adverbio de lugar.	Mar.	Voz de la Marinería.
adv. m.	Adverbio de modo.	Mat.	Voz de la Matemática.
adv. t.	Adverbio de tiempo.	Med.	Voz de la Medicina.
Agri.	Voz de la Agricultura.	met.	Metáfora, ó metafóricamente.
Alb.	Voz de la Albeytería.	Milic.	Voz de la Milicia.
Albal.	Voz de la Albaladería.	Mina.	Voz de las Minas.
Alg.	Voz de la Aígebra.	Mit.	Voz de la Mitología.
amb.	Ambiguo.	mod. adv.	Modo adverbial.
Anat.	Voz de la Anatomía.	Mon.	Voz de la Monería.
ant.	Voz, ó frase antiquada.	Mus.	Voz de la Música.
Arif.	Voz de la Aritmética.	n.	Nombre.
Arg.	Voz de la Arquitectura.	n. p.	Nombre propio.
Art.	Voz de la Artillería.	Naut.	Voz de la Náutica.
Astrof.	Voz de la Astrología.	num.	Número.
Astron.	Voz de la Astronomía.	Opt.	Voz de la Óptica.
augm.	Aumentativo.	Orat.	Voz de la Oratoria.
Ballest.	Voz de la Ballestería.	Orth.	Voz de la Ortografía.
bax.	Voz, ó frase baxa.	otr. par.	Otras partes.
Blaz.	Voz del Blason.	p.	Plural.
Bot.	Voz de la Botánica.	p. a.	Participio activo.
Cabestr.	Voz de la Cabestrería.	p. And.	Provincial de Andalucía.
Can.	Voz de la Cantería.	p. Ar.	Provincial de Aragón.
Corp.	Voz de la Carpintería.	p. Ast.	Provincial de Asturias.
Catoptr.	Voz de la Catóptrica.	p. Ast. de Santill.	Provincial de Asturias de Santillana.
Cer.	Voz de la Cerería.	p. Extr.	Provincial de Extremadura.
Cerrag.	Voz de la Cerrajería.	p. Gal.	Provincial de Galicia.
Cetr.	Voz de la Cetrería.	p. Gran.	Provincial de Granada.
Chim.	Voz de la Química.	p. Manch.	Provincial de la Mancha.
Cir.	Voz de la Cirugía.	p. Mont.	Provincial de las Montañas.
cocin.	Voz del arte de Cocina.	p. Montañ. de Burg.	Provincial de las Montañas de Burgos.
Com.	Voz del Comercio.	p. Mur.	Provincial de Murcia.

Figure 41: List of abbreviations of the Diccionario de Autoridades (1770), RAE

This methodology and concerns about the selection and treatment of terms were followed and referred to in the prologues of several editions. To cite one more example, in the Prologue of the DA (1770), one can read: *‘De las voces de ciencias, artes y oficios se ponen aquellas que están recibidas en el uso comun de la lengua’* [From the entries of sciences, arts and trades are included those that are received in the everyday use of the language] (DA, 1770, p. 1). In the last paper edition, the criterion is to mark only the senses that are not considered to be of general use:

El Diccionario da cabida a aquellas voces y acepciones procedentes de los distintos campos del saber y de las actividades profesionales cuyo empleo actual – se excluyen también los arcaísmos técnicos – ha desbordado su ámbito de origen y se ha extendido al uso, frecuente u ocasional, de la lengua común y culta. Siempre que tal uso no se haya hecho general, las acepciones tienen una marca que las individualiza. [The Dictionary includes those entries and senses coming from the different fields of knowledge and of the professional activities whose current employment – excluding also the technical archaisms – has overflowed its scope of origin and has been extended to frequent or occasional use of the common and cultured language. Whenever such use is not general, they have a label that individualises them.] (DLE, 2014)

Concerning the first attempt of the ACL dictionary, in 1793, academicians comment on terms in the Introduction: *‘Admitirseñão também as vozes peculiares às Sciencias, às Artes liberais e mecânicas, se estas vozes se achavam impressas nos Autores*

*aprovados*⁷⁰ e *Diccionarios Portuguezes*’ [The entries peculiar to the sciences, the liberal and mechanical arts will also be admitted, if these entries were found in the approved Authors and Portuguese Dictionaries] (ACL, 1793, p. xiv).

After almost a hundred years, in the *‘Relatório da Comissão encarregada de propor à Academia Real das Sciencias de Lisboa o modo de levar a efeito a publicação do Diccionario da Lingua Portuguesa’* [Report of the Commission in charge of proposing to the Academia Real das Sciencias de Lisboa how to carry out the publication of the *Diccionario da Lingua Portuguesa*] (ACL, 1870, p. 5), we can read that *‘desde logo se levanta a questão de se havemos de incluir no Diccionario apenas os termos da lingua vulgar e da litteraria, ou além d’estes os tecnologicos e os obsoletos’* [from the onset the question arises as to whether we should include in the dictionary only the terms of the common and literary language, or in addition to these the technological and obsolete terms]. In other words, the inclusion of terms was still a matter of debate and concern among the Portuguese academicians. The commission, recognising that *‘No estado da civilização actual, em que a sciencia deixando de ser o apanágio exclusivo dos sábios, invade todos os espíritos e por assim se democratizar’* [In the current state of civilisation, in which science is no longer the exclusive attribute of sages, it invades all minds and thus becomes democratised] (ACL, 1870, p. 5), concludes that *‘não parece racional excluir do Diccionario todos os vocabulos scientificos’* [it does not seem rational to exclude all scientific words from the dictionary] (*ibidem*), excluding only those that are of *‘uso tão peculiar ás profissões especiaes’* [very particular use to special occupations] and privileging those that are *‘indispensaveis’* [indispensable] (*ibidem*). From 1793 to 1870, much had changed in society at large, which certainly justified this approach, supporting the aforementioned democratisation of science. And this same question is debated again as early as the 20th century. In 1936, Júlio Dantas (1876–1962), while he was Chairman of the ACL, reinforced the need to include terms in an academic session dedicated to *‘Nomenclaturas científicas no Dicionário da Academia’* [Scientific Nomenclatures in the Academy Dictionary]. Dantas (1936) specified that *‘Não, porém, todas as terminologias de cada ciência ou de cada técnica; mas a parte delas que possa considerar-se definitivamente incorporada na língua portuguesa’* [Not, however, all the

⁷⁰ *‘Autores aprovados’* [approved Authors], that is, the concept of *‘auctoritas’*. See Chapter 3, p. 92.

terminologies of each science or each technique; but that part that can be considered definitively incorporated into the Portuguese language] (p. 301) should be registered in the dictionary. He then talks about the methodology to be used, considering that it is not a dictionary or a special vocabulary of any particular science, but a language dictionary, excluding the terms, scientific neologisms still not reviewed and words rejected by international committees.

Another point that deserves some attention is the reference to the need to '*vernaculização da linguagem tecnológica*' [popularise technological language] (Dantas, 1936, p. 302) because the use of too many foreign words was already resented. This topic reveals the normative concern of the Portuguese academic institution. Years later, in 1974, Jacinto Prado Coelho, in the presentation of the plan for a new academic dictionary, notes that some terms will appear: '*os tecnicismos mais generalizados na linguagem usual; os tecnicismos que, embora não generalizados correspondem a noções ou classificações e a aparelhos fundamentais em cada ciência ou técnica*' [the most generalised technicalities in the usual language; the technicalities that, although not generalised, correspond to notions or classifications and fundamental devices in each science or technique] (Coelho, 1974, pp. 250–251). This is a sentence that will be used by the editors of the 2001 edition, as we will discuss in Chapter 6.

Finally, and although our research does not focus on English dictionaries, we intend to leave here a brief note about the inclusion of terms in English general dictionaries. For some scholars (Landau, 2001, p. 46–52; Jessen, 1996, p. 68), it seems to date back to John Bullokar's *An English Expositor* (1616), included in the 'hard words' tradition (Landau, 2001, pp. 46–52). Bullokar – who was a physician – included terms from medicine, logic, philosophy, law, astronomy and heraldry.

Concerning Samuel Johnson's dictionary, one of his guiding principles was that 'the value of a work must be estimated by its use' (Johnson, 1747). 'It is not enough', he continues, 'that a dictionary delights the critick, unless, at the same time, it instructs the learner'. As the English lexicographer continues: 'and the words that most want explanation are generally terms of art'. Johnson thus legitimises the inclusion of terms in general dictionaries.

Of such words, however, all are not equally to be considered as parts of our language; for some of them are naturalised and incorporated; but others still continue aliens, and are rather auxiliaries than subjects. This naturalisation is produced either by an admission into common speech, in some metaphorical signification, which is the acquisition of a kind of property among us; as we say, the zenith of advancement, the meridian of life, the cynosure of neighbouring eyes; or it is the consequence of long intermixture and frequent use, by which the ear is accustomed to the sound of words, till their original is forgotten, as in equator, satellites; or of the change of a foreign to an English termination, and a conformity to the laws of the speech into which they are adopted; as in category, cachexy, peripneumony.

Of those which still continue in the state of aliens, and have made no approaches towards assimilation, some seem necessary to be retained, because the purchasers of the dictionary will expect to find them. Such are many words in the common law, as *capias*, *habeas corpus*, *præmunire*, *nisi prius*: such are some terms of controversial divinity, as hypostasis; and of physick, as the names of diseases; and, in general, all terms which can be found in books not written professedly upon particular arts, or can be supposed necessary to those who do not regularly study them. (Johnson, 1747)

Johnson (1747) remains clear that the use of terms in non-specialised contexts justifies their inclusion in a general dictionary. He discusses the criteria for their inclusion and the difficulty of defining them. On this basis, as stated by Landau (2001), 'it is unwise to exclude terms of science and art' (p. 59), even the terms with 'alien' status, as the end user may need them and look up their meaning in the dictionary. Boulanger (2001), in turn, considers that the lexicographer makes a double choice: '*d'abord il établit le catalogue des mots; ensuite il sélectionne les vocabulaires thématiques appropriés, puis, à l'intérieur de ceux-ci, il procède à un nouveau tri afin de recruter un certain nombre d'unités pertinentes*' [first he establishes the inventory of words; then he selects the appropriate thematic vocabularies, then, within these, he proceeds to a new sorting in order to recruit a certain number of relevant units] (p. 247).

Even today, if the inclusion of information that is too highly specialised in language dictionaries is discussed – because it may be unclear to the target audience to whom they are addressed (Correia, 2009) –, the inclusion of terms in a general language dictionary is mandatory. The advances in science in general and technology in particular, accompanied by the spread of scientific concepts among native speakers, dictated a mandatory presence of terms in general dictionaries. More: the interest in terms is also justified by the fact that they are one of the privileged sources of lexical renewal and

enrichment of the linguistic systems, and, by their identification, structuring and storage, fundamental for the organisation of data. There is a strong likelihood that an ordinary user will look for terms in a general dictionary rather than specialised dictionaries.

5.2 Research on the Inclusion of Terms in General Dictionaries

Many researchers have conducted studies on the presence of terms in general dictionaries based on monolingual dictionaries (Rey, 1985; Béjoint, 1988; Tournier, 1992; Cabré, 1994; Paz Battaner, 1996; Estopà, 1998; Boulanger, 2001; Roberts, 2004; Guerra Salas & Gómez Sánchez, 2005; Nomdedeu Rull, 2008). For example, Estopà (1998) analyses marking mechanisms; Boulanger (2001) studied the development of technolectal usage labels in general French bilingual dictionaries; Guerra Salas and Gómez Sánchez (2005) also studied technolectal usage labels dictionaries for learners; and Nomdedeu Rull (2008) studied the sport domain label in DLE.

Landau (1974), Boulanger and L'Homme (1991), Wiegand (1984) and Ahumada (2002), among others, claim that terms in an unabridged dictionary make up between 40 and 50 percent of the content. Casteleiro (2008), noting that the DLPC registers around 70,000 entries, points out that around 32,000 of these units are terms or meanings from different domains (cf. p. 317). Rey and Delesalle (1979, p. 23) had already recognised that the proportion was high. Rondeau (1984, pp. 1–4) lists several reasons that justify the general increasing presence of terms in general dictionaries – the advancement of science, the technological boom, the growth of communication media that contribute to scientific popularisation, and so on.

We saw in the previous section that the French academicians began by making a distinction between *mots communs* [common words] and *termes des arts et des sciences* [arts and science terms], or, to abbreviate, *words* and *terms*. Although the use of *term* is consolidated, the concept itself is quite intricate, and there is some terminological variation around it. In the lexicographic scenario, it is common to find the terms '*technolectes*' (e.g., Boulanger & L'Homme, 1991; Verdelho, 1994) and '*tecnicismos*' (e.g., Haensch, 1997, p. 148; DLPC, p. xiv) referring to what we are here considering

terms. The unit ‘terminologies’ or ‘*terminologias*’ can also be found in practically all research in the field. According to L’Homme (2004, p. 31), one can speak interchangeably of ‘term’, ‘terminological unit’, ‘specialised lexical unit’, ‘terminologism’, or ‘technical term’. Another way the literature refers to terms in dictionaries is ‘scientific and technical words’ (e.g., Béjoint, 1988, pp. 354–368). There are even scholars who make a distinction between ‘scientific term’ and ‘technical term’. This is the case for Landau (1974, p. 241): a term is ‘scientific’ when its meaning is restricted and only applied in a particular field; on the contrary, if a term does not refer to a particular scientific field but specialised technical contexts, it is a ‘technical term’. This distinction always raises many obstacles to lexicographic work – while it is tough to separate what belongs to the general lexicon from what belongs to a specialised field, distinguishing between a technical and a scientific term increases that difficulty, even more so when both are specialised. For the purpose of this thesis, we do not make this distinction.

Guilbert (1973, p. 35), recognising that the ideal source with which to observe the inclusion of terms in the general language is the dictionary, states that this inclusion does not necessarily prove that they are integrated ‘*dans l’usage et font partie du lexique commun*’ [in everyday usage and are part of the common lexicon]. General language dictionaries illustrate the ‘*va-et-vient entre les termes et la circulation sociale de leur expression linguistique*’ [back-and-forth between the terms and the social circulation of their linguistic expression].

This *va-et-vient* between the terms leads us to the process by which terms move from specialised language to everyday language, i.e., the use of terms in a non-specialised context. This linguistic phenomenon has different understandings and may be considered as ‘*banalisation lexicale*’ [lexical banalisation]⁷¹ (Galisson, 1978), ‘*vulgarisation scientifique*’ (Guilbert, 1975) or ‘determinologisation’ (Meyer & Mackintosh, 2000) – a term that we adopt here because we consider it very evocative.

⁷¹ Galisson is considered the creator of this term. In its original sense, this term does not have the same sense that we adopt here for determinologisation. For Galisson (1978, pp. 71–128), ‘*banalisation lexicale*’ points to ‘*la manifestation socialisée du processus d’accommodation*’ [the socialised manifestation of the accommodation process] while ‘*vulgarisation scientifique*’ is ‘*la manifestation individualisée*’ [the individualized manifestation] (*ibidem*). However, in the literature, their use is often found to be synonymous (e.g., Josselin-Relay & Roberts, 2014).

In our research, we privilege determinologisation processes that we describe as ‘the process by which a term is transformed into a general language word or expression’ (Costa et al., 2021b). In these cases, it no longer refers to a concept and, consequently, it is no longer part of a concept system within a given domain. Determinologisation does not mean that specialists no longer use the term. The term loses the link to a certain concept and is therefore no longer part of a concept system within a given domain, acquiring new properties. Nová (2018) goes further and considers that determinologisation corresponds to the process by which ‘a scientific term, during its way from a field specialist to a layperson, loses its accuracy, gets new connotations, and the word can be even moved to refer to a completely different thing’ (p. 387).

The terms that have undergone a process of determinologisation are indeed recorded in the dictionaries. Interestingly, their registration is usually no longer accompanied by a domain label in these cases. For some authors (e.g., Reboul (1994) cited by Delavigne (2002), as soon as a term leaves specialised discourse, it can no longer be considered a term. *‘Lorsque le terme est vulgarisé [...], la valeur se diffuse; la notion n’est plus celle du spécialiste; il n’y a d’ailleurs plus de notion. Il ne semble plus possible de parler de terme’* (p. 228) [When the term is popularised [...], the value is diffused; the notion is no longer that of the specialist; besides, there is no longer any notion. It no longer seems possible to speak of a term]. On the other hand, Delavigne (2002, p. 225, 227, 230) states that terms found in popular science discourses can be truly considered terms: *‘Les termes dans les discours de vulgarisation sont amenés à certains bouleversements sémantiques et référentiels. Nous n’y voyons cependant pas une raison suffisante pour ne pas les considérer encore comme des unités terminologiques’*. [The terms in popularised discourses have led to certain semantic and referential upheavals. However, we do not see this as sufficient reason not to consider them as terminological units yet.]

For Carras (2002), *‘les discours de vulgarisation qui accompagnent la diffusion de certains thèmes scientifiques périodiquement médiatisés [...] font migrer vers la langue commune des termes que le public va s’approprier’* [the popular science discourses that accompany the dissemination of certain periodically mediated scientific topics [...] migrate into the common language terms that the public will appropriate]. Thus, terms

exist in popular science discourse as well as in specialised discourse, and it is now well recognised that there are constant back-and-forth movements or interference between the general language – or common language – and specialised language. Thus, on the one hand, we are witnessing a ‘terminologisation of words in the general language’ (Cabr , 1994, p. 593, also cf. Sager, 2000, p. 43), and, on the other, a phenomenon of ‘de-terminologisation’ (Meyer & Mackintosh, 2000).

Finally, we summarise some relevant cases of determinologisation recorded in general language dictionaries. Accordingly, we identified three types:

- 1) Determinologisation *sensu stricto*: Speakers begin to use a given term in a context different from the original domain or specialised context. Thus, the term originates a new meaning. In the DLP, this type of phenomenon corresponds to separate meanings – the original terminological meaning and the determinologised meaning generally based on metaphor. When the determinologised meaning is lexicalised, lexicographers usually record it in dictionaries using the label ‘figurative.’ The new unit loses its specialised features since the core meaning is used figuratively. This phenomenon is verified in specific sports terms, namely in football terms, as is the case of the “cart o vermelho” [red card]. In addition to being a term widely used in the context of certain sports, it is figuratively used as any punishment. In this case, in the DLP, we will add the ‘*figurado*’ [figurative] label, which was not used in the previous edition.
- 2) Determinologisation *sensu lato*: The term’s connotation changes when used in contexts other than the domain of origin. The term “granito” [granite] or “m rmore” [marble] is an example. In a geological context, granite is an igneous rock whose essential minerals are quartz and alkali feldspar. However, the use of the term granite is current in industrial sectors with an understanding different from a geologist’s. In industry, all polished igneous rocks are often called granite. The same phenomenon happens with marble; the fundamental sense of meaning is retained, but the concept undergoes some changes. This is a phenomenon not always easy to illustrate in general language dictionaries. Collaborating with an expert enables detecting these

details. The lexicographer can open a new meaning with an extension of the meaning label or introduce a note.

- 3) Blurring of the meaning: The concept of the term changes in popular usage.

We recognise, however, as per Nová (2018), that ‘there is probably no universal way to treat determinologised words, but many of them need a special approach’ (p. 397).

5.3 Dealing with Terms in General Dictionaries

So far, we have been discussing the inclusion of terms in general dictionaries and their implications for lexicographic work. Nevertheless, to proceed with our research, we now aim to clarify some of the key terminology concepts of this doctoral research project, namely, the *term*, which necessarily brings the *concept* along. Furthermore, microstructurally speaking, we must point to the *domain* and the *definition*.

5.3.1 Term and Concept

These two core keywords have been defined quite differently by the various theoretical approaches in terminology (e.g., Wüster, 1979/1998; Felber, 1987; Cabré, 1999; Temmerman, 2000; Gaudin, 2007; Faber, 2009).

Terms are objects of interest for terminology as a linguistic representation of a concept that belongs to a given domain of knowledge or as a denomination of a concept, verbally formulating the people’s perception.

Many of the earlier definitions of *term* did not clearly distinguish between *term* and *word*, which did not benefit the definition of something already complex. Rondeau (1984, p. 19) defines the term as ‘*un signe linguistique [...], c’est-à dire une unité linguistique comportant un signifiant et un signifié*’ [a linguistic sign [...], i.e., a linguistic unit comprising a signifier and a signified]. In line with Saussure, Rondeau (1984, pp. 21–23) considered the term to be a linguistic sign in itself, consisting, on the one hand, of a signifier called a ‘denomination’, and, on the other, a signified called a ‘notion’. This idea of the term as a linguistic sign by itself is now shared by authors such as Depecker (2003,

p. 20), who, however, talks about ‘designation’ and ‘concept’ in line with ISO 1087 (2019). We agree with Sager (1990, p. 57) when stating that ‘terms are the linguistic representation of concepts’.

We bear in mind that a term is a ‘*designation* that represents a *general concept* by linguistic means’ (ISO 1087, 2019, p. 7). According to terminological ISO standards, the concept ‘should be viewed not only as a unit of thought but also as a unit of knowledge’ (ISO 704, 2009, p. 3). However, we adopt the ISO 1087 (2019) definition, according to which the concept is a ‘unit of knowledge created by a unique combination of *characteristics*’ (ISO 1087, 2019, p. 3).

Another concept we aim to clarify is the one of *characteristic* – an ‘abstraction of a property’ (ISO 1087, 2019, p. 3). We pay attention only to the so-called *essential characteristics* – ‘characteristic of a concept that is indispensable to understand[ing] that concept’ (*ibidem*). As we will see in Chapter 7, the distinctive characteristics of a concept are fundamental to the creation of concept systems and for drafting definitions.

As we see in Figure 42, the concept – a non-linguistic element – is designated by the term, and the term – a linguistic element – in turn lexically designates the concept.

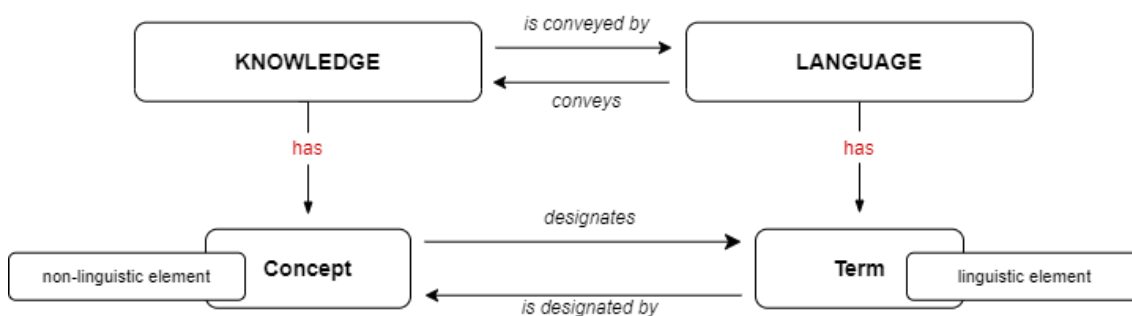


Figure 42: The Relationship of Concept and Term mirroring the double dimension of terminology
(adapted from Costa, 2021)

Observing Figure 42, it is impossible not to see the relationship between concepts and terms. The texts (language discourse) do not in themselves contain concepts, as they are extra-linguistic elements, containing only the linguistic uses of the terms they designate. However, this does not prevent us from finding linguistic

manifestations pointing to a particular conceptual organisation. Our concern is, in fact, a better description of the language, but to achieve this we argue that we have first to understand the knowledge about a field but also the ways in which that knowledge is conveyed by language (cf. Costa, 2013, p. 40).

Although term and concept are independent elements, in practice it is not always easy to isolate them when working in lexicography. Even though it is hard to establish a boundary between the conceptual and the linguistic dimensions, the two should not be seen as antagonistic but as quite the opposite: *'la perspective linguistique, plutôt sémasiologique et la perspective conceptuelle, plutôt onomasiologique, [...] ne s'excluent pas mutuellement, mais se complètent'* [the linguistic perspective, rather semasiological, and the conceptual perspective, rather onomasiological, [...] are not mutually exclusive; more so, they complement each other] (Costa, 2006b, p. 85). This way a mixed approach supports the theoretical assumptions. As Costa (2013) explains, we 'can shift from the concept to the term and from the term to the concept' (p. 40). So, throughout this work, we follow two complementary methodological approaches:

- 1) An onomasiological approach, rooted in Wüsterian doctrine, advancing from the concept to the term, modelling (always with the help of the expert) concept systems⁷²;
- 2) A semasiological approach, advancing from the term to the concept and its relations in a textual environment by analysing the terminological data extracted from the dictionaries under study.

In lexicography, we adopt a semasiological analysis of the lexicographic articles related to terms. As we argue for a mixed approach, the onomasiological approach, i.e., the delimitation and organisation of the domains under analysis and the analysis of the concepts and the linking to other concepts within a specific concept system, which is 'the process of discovering and representing the conceptual structures underlying the terms of a domain' (Meyer & Mackintosh, 1996, p. 261), will be iteratively introduced in our methodology (Chapter 7). The relations between concepts and the location of the concept in a particular system are not always easy to establish. As lexicographers, we

⁷² A concept system is understood as a 'set of concepts structured in one or more related domains according to the concept relations among its concepts' (ISO 1087, 2019, p. 6).

could not aim to work with all identified concepts, but we consider it important to analyse the relations among relevant concepts and to organise them into concept systems, which will benefit the drafting of definitions.

All that remains to be mentioned is that a concept can designate a simple term or a complex term⁷³ or, in our preferred words, terms may be monolexical or polylexical units.

5.3.2 Term as a Polylexical Unit

In specialised literature, different authors with different theoretical backgrounds (e.g., Gantar et al., 2018; Fellbaum, 2016; Baldwin & Kim, 2010; Calzolari, Zampolli & Lenci, 2002; Moon, 1998; Cowie, 1994; Mel'čuk et al., 1984/1999) have referred to polylexical units as multiword expressions, collocations, phrasemes, phraseologies, idiomatic expressions, lexical combinations and so forth. Each of these designations is often defined within a particular theoretical linguistic framework. These morphosyntactic sequences are generally described as complex units.

We recognise that the term *multiword expression* (MWE) is already widely used, including in the LMF standard (ISO/FDIS 24613-1, 2019), but the terminology used in this research aims to be supra-theoretical and, consequently, as neutral as possible, hence our preference for *polylexical unit*. For our purpose, a polylexical unit can be defined as a stable and recurrent sequence of units (a lexical unit composed of two or more lexical items) perceived as an independent lexical unit by the speakers of a language. Terminologically, a polylexical unit is always recognised when the concept to which it refers is identified within a subject field.

We will not explore the morphosyntactic properties of polylexical terms but rather identify the polylexical terms that can be found in lexicographic practice and their encoding. Scholars (e.g., Svensén, 2009; Atkins & Rundell, 2008; Fontenelle, 1997; Mel'čuk et al., 1984/1999; Zgusta, 1971) have long recognised that polylexical units are essential components of lexical resources. When including a polylexical item in a

⁷³ ISO 1087 (2019) defines a complex term as 'term that consists of more than one word or lexical unit' (p. 8).

dictionary, lexicographers must decide on the degree of its lexical independence based on several criteria from different fields of knowledge, including statistics, semantics, morphosyntax, pragmatics and/or, broadly speaking, culture. This kind of lexicographic judgement, enacted through a particular editorial policy and influenced by the conventions of a given lexicographic tradition, necessarily leads to multiple ways of capturing, classifying and presenting lexicographic knowledge about polylexical units. There are some problems with placing polylexical units as sublemmas. First, lexicographers need to designate the unit component under which the entire unit should be registered, as well as other issues concerning variable components. The lack of a more general agreement within the lexicographic community makes the process of encoding dictionaries particularly challenging. This is due to a conundrum: how can we identify, describe and consistently represent this type of linguistic phenomena in lexical resources if we disagree on what they are and/or what to call them?

Structurally speaking, Salgado et al. (2019) identified four different types of headwords in the DLPC – monolexical units, polylexical units, affixes and abbreviations (Figure 43).

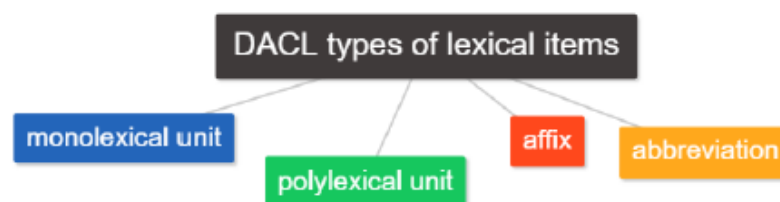


Figure 43: Formal representation of lexical entries in the DPLC (Salgado et al., 2019)

Monolexical and polylexical units can be divided into two types – lexical units (nouns, adjectives, verbs) and grammatical units (conjunctions, determiners, prepositions, pronouns). When polylexical units are headwords, they can be of two different types: (i) *palavras compostas* [compounds]⁷⁴ which are graphically realised as *palavras hifenizadas* [‘hyphenated words’] (DLPC, p. xiv) (e.g., *decreto-lei* [decree-law],

⁷⁴ By compounds, we mean every lexical unit formed by two or more elements with autonomy within the language that together form a new lexical unit with a new meaning.

franco-canadiano [French-Canadian], *pré-cristão* [pre-Christian]); and (ii) *locuções latinas* [Latin phrases] (e.g., *fiat lux* [let there be light]). Under this classification, we have included compounds and all kinds of lexical combinations, such as collocations or phrasemes.

Whereas in terminological or specialised dictionaries, a polylexical unit constitutes a headword (lemma), in general language dictionaries polylexical terms can be macrostructural and microstructural components of the lexicographic article. When they belong to the microstructure it is difficult to locate them. Two main challenges affect the modelling of polylexical units in general language dictionaries, both related to the typographical constraints of print-based dictionaries. These are as follows:

- (1) In most general language dictionaries, polylexical units do *not* appear as lemmas, i.e., independent lexical units in the dictionary macrostructure, but rather as sublemmas within entries that have a monolexical headword; and
- (2) Polylexical units are not always *explicitly* labelled as such in dictionaries: they may be typographically singled out, using a particular typeface, but they are not always accompanied by the label that identifies the given unit as a ‘collocation’, an ‘idiom’, or a ‘proverb’.

The position of polylexical units in the dictionary and the benefits of lemmatisation have been discussed before (see Jónsson (2009) and Lorentzen (1996), for instance). For our purposes, however, it is essential to note that when we suggest particular encodings of the new edition of the DLP, we will follow that very dictionary’s structure and conventions. This does not suggest an attempt to flatten the hierarchy or encode all polylexical units using the same set of tags. Instead, they will be encoded as they appear within the structure imposed by the dictionary itself – in this case, no change concerning the representation already adopted in DLPC will be made.

As for the lack of explicit labels for particular types of polylexical units, we will explain, in Chapter 9, the extent to which the types can be deduced from the entry structure.

5.3.3 Term and Domain

The notion of domain is one of the criteria we traditionally use to distinguish a term from a lexical unit. Boulanger (2001, p. 247) characterises terms as '*unités représentatives d'une sphère d'activité*' [representative units of a sphere of activity]. A term is always defined with consideration to the domain to which it belongs. But the same term can point to different concepts depending on the domain in question. Likewise, the concept is always defined in relation to other concepts within that domain (Cabr , 1994, p. 591). For instance, the Portuguese lexical item "*m o*" [hand], originally from the ANATOMY domain, is also used in the SPORTS domain – in FOOTBALL, it indicates a foul committed by a football player who deliberately touches the ball with that part of the body.

The interdependence among concept, term, domain and definition constitutes the meaning triangle that is useful for terminological work. Ogden and Richards (1923) developed the *semantic triangle* or the *meaning triangle* (Figure 44).

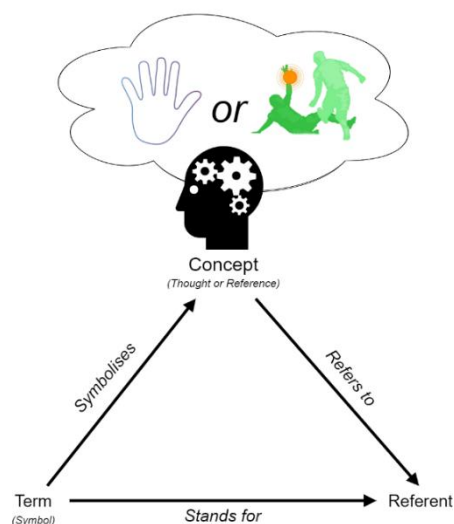


Figure 44: The Meaning Triangle (adapted from Ogden and Richards, 1923)

This diagram from Ogden and Richards (1923) has three vertices: *symbol*, *thought or reference* and the *referent*. We adapted this model in Figure 44, and we have the Term (*Symbol*), the Concept (*Thought or Reference*) and the Referent, and it shows correspondence among terms and concepts or referents. However, the relation

between a term and a referent is indirect, which means that concepts mediate the relationships between terms and referents.

Looking at Figure 44, we recognise that terms are lexical units that designate concepts and convey meanings, and that the same term can have several specialised meanings pertaining to different fields.

5.3.4 Term and Definition

The definition has been a hotly debated topic for centuries, not only in linguistics or terminology or lexicography but even more in philosophy or logic. The most challenging aspect is, without a doubt, the difference this word has in logic, philosophy, and even terminology, when compared with its meaning in lexicography.

Two concepts that appear in practically all definition theories need to be clarified:

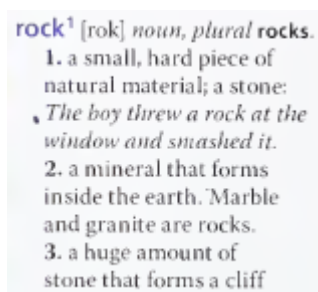
- *definiendum* (what is to be defined);
- *definiens* (how something is to be defined).

The origins of the debate go back to Ancient Greece, with Aristotle occupying a prominent place. The Aristotelian concepts of *genus* and *differentia* (specific difference) are still used today in the formulation of definitions and they impact terminological and lexicographic practices. According to Aristotle (Granger, 1983), the *genus* complemented by the *differentia* reveals knowledge of the essence of a thing. In Aristotelianism, the definition represents a philosophical concept that points out the essential nature of something, thus determining its similarities and differences in relation to other realities.

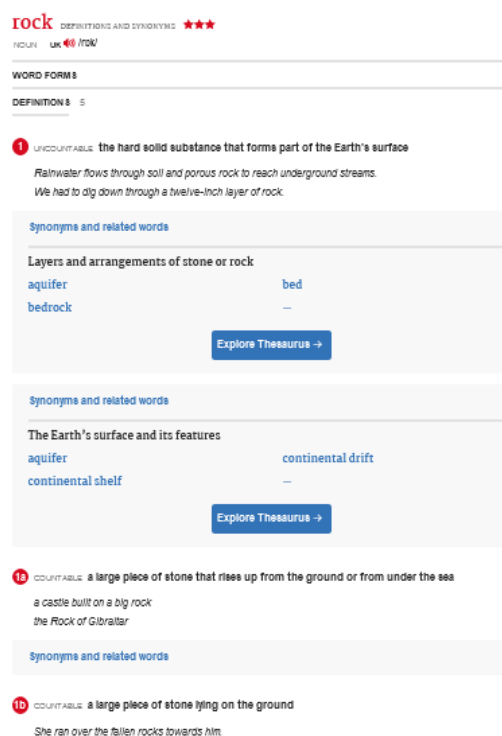
Much has been debated about the problematic issue of applying the term “definition” to explaining meanings in dictionary entries. We have found it very interesting to observe the use of the lexical unit “explanation” in Johnson’s Preface: ‘That part of my work on which I expect malignity most frequently to fasten, is the *explanation*; in which I cannot hope to satisfy those, who are perhaps not inclined to be pleased, since I have not always been able to satisfy myself.’ (Johnson, 1755, s. p.). Johnson seems to prefer “explanation” rather than “definition”. Wiegand (1984) also

employs the term “lexicographic explanation of meaning”. In fact, this is a better description of what lexicographers actually do. But for practical reasons, we decided to adopt the term “lexicographic definition” and the short form and more familiar term, i.e., “definition”.

In a general language dictionary, we foresee the need for a lexicographic definition. On the other hand, different dictionaries often define the same concept designated by a term in different ways. It is important to note that the dictionaries themselves can be addressed to different target audiences.



MACMILLAN Dictionary for CHILDREN



MACMILLAN Dictionary (2021, online)

rock [geol] An aggregate of one or more minerals, e.g., granite, shale, marble; or a body of undifferentiated mineral matter, e.g., obsidian; or a lithified organic material, e.g., coal.
rock [topog] Any prominent peak, cliff, or promontory, usually bare, when considered as a mass, e.g., the Rock of Gibraltar. Also, a rocky mass lying at or near the surface of a body of water, or along a jagged coastline, esp. where dangerous to shipping.
rockallite (rock'-al-lite) A coarse-grained, mafic, alkalic granite composed of quartz, aegirine, albite, and microcline. The name, given by Judd in 1897, is for Rockall Bank in the North Atlantic Ocean 350 km west of Scotland. Obsolete.
rock asphalt *asphalt rock*.
rock association A group of igneous rocks within a petrographic province that are related chemically and petrographically, generally in a systematic manner such that chemical data for the rocks plot as smooth curves on variation diagrams. Syn: *rock kindred*; *kindred*; *association* [petrology].

Glossary of Geology (Neuendorf, Mehl Jr. & Jackson, 2011)

Figure 45: The entry 'rock' in different English dictionaries

The different definitions we observe in Figure 45 have arisen because these dictionaries are designed for different target audiences – the first is a dictionary addressed to children (MACMILLAN, 2007), the second is the unabridged version of Macmillan dictionary (MACMILLAN, 2021), and the third is a glossary, a specialised resource (Neuendorf, Mehl Jr. & Jackson, 2011). As Landau (2001) stated, ‘lexicographers are concerned with explaining something their readers will understand’ (p. 154), while terminologists are focused on the internal coherence of their system.

The Latin etymon ‘*definitio*’ means ‘action of setting a limit’. The idea of limit is fundamental to understanding the relationship between term, definition and concept. As Costa (2013) explains:

Definitions are the main concern of terminological and lexicographical work alike since they allow us to establish the boundaries of a concept designated by a term. The definition allows for the establishment of a relationship between the concept and the term that is used to evoke it. (Costa, 2013, p. 40)

Our interest is in the definition in natural language. In our methodological proposal, we understand along the same lines as Silva (2014) that: ‘*Definir é fixar os limites do conceito recorrendo à língua, é distinguir os conceitos uns dos outros no seio de um sistema*’ [Defining is setting the limits of the concept using language, to distinguish the concepts from each other within a system] (p. 21). To fix the boundaries of the concept implies finding the distinctive characteristics that differentiate them within a concept system.

The definition simultaneously designates 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; as well as 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*. In Rey’s (1995) words, ‘it designates the operation and its result’ (p. 41). As stated by Rey (1979, p. 40), ‘*Le seul moyen pour exprimer ce système de distinctions réciproques est l’opération dite définition*’ [The only way of expressing this system of reciprocal differences is the operation definition].

We distinguish the terminological definition (cf. De Bessé, 1990; Rey, 1995; Sager, 2000; Temmerman, 2000) from the lexicographic definition (Mel'čuk & Polguère, 2018), which is generally suitable for general language dictionaries. Although terminology and lexicography favour definition by *intension*, their purposes are different. 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. In contrast, the lexicographic definition seeks to describe the (signified) meaning(s) of a lexical unit. As De Bessé (1990) notes, lexicography aims to define words (rather than concepts), following a primarily semasiological approach. However, the focus of terminological dictionaries is placed on domain knowledge. In terminology, the definition – what we will call terminological definition – establishes the relationship between the lexical unit (term) and the specialised concept from a domain of knowledge.

The terminological definition is related to the definition of the thing, as opposed to the lexicographic definition that relates to the usage of the word and is made by identifying the semantic features that characterise the meaning. The unit of meaning aimed at in the terminological definition is the concept (in terminology we define concepts, not terms, but the term is always inseparable from the concept it designates), which differs substantially from the meaning. The difference between the terminological definition and the lexicographic definition, therefore, leads to different approaches, although they do not exclude one another.

We also anticipate that many of the definitions in the dictionaries under analysis may be out of date. Knowledge evolves, which implies that the conceptual representation is constantly changing and, consequently, the discourse of a given scientific community that conveys the knowledge will also have to be reformulated. We will analyse the definitions of random concepts by the mean of terms to show, in a sustained manner, that the conceptual aspect and the relation between the concepts are relevant in the terminological definition, even if the audience is not made up of specialists.

To help lexicographers with the task of writing terminological definitions, we once again resorted to the ISO standards. Aiming to differentiate a given concept from another in a specific concept system belonging to a certain domain, the type of definition that interests us is the 'representation of a concept by a descriptive statement which serves to

differentiate it from related concepts' (ISO 1087, 2019, p. 6). The ISO 1087 (2019) standard itself highlights this setting of the concept's limits.

The ISO standards (ISO 704, 2009; ISO 1087, 2019) refer to the intensional definition and the extensional definition. The dichotomy between intensional (those specifying the close gender and specific difference) and extensional (those that enumerate the members of a given class or the subordinate concepts) is an Aristotelian legacy. The former consists of stating the immediate generic concept and the delimiting characteristics of the defined concept; the latter consists of enumerating all of its subordinate or partitive concepts. In our work, the formulation of definitions is based on the intensional definition model. As Eco (2001) explains, demonstrating *what a thing is* (extension) is not the same as proving that *a thing is a thing* (intension):

Não se define um homem dizendo que corre ou que está doente, mas dizendo que é animal racional de tal modo que o definiens seja co-extensivo ao definiendum e reciprocamente, isto é, que não haja nenhum animal que não seja animal racional. [A man is not defined by saying that he is running or sick, but by saying that he is a rational animal in such a way that the definiens is co-extensive with the definiendum and reciprocally, that is, that there is no animal that is not a rational animal]. (Eco, 2001, pp. 104–105)

The two referenced ISO standards and many scholars (e. g., Temmerman, 2000, p. 79; Cabré, 1999, p. 98; Sager, 1990, p. 24; Felber, 1987, p. 98) give preference to the intensional⁷⁵ type of definition, whenever possible, since this type of description makes the essential characteristics explicit and allows positioning of the concept in a concept system. In the context of our work, we are in line with Löckinger, Kockaert and Budin (2015) when we state that the intensional definitions become the 'standard way of illustrating concepts' (p. 66). Moreover, in Chapter 7, we will show how the modelling of concept systems can help the writing of well-formed definitions in natural language. The definitions must refer to the superordinate concept (*genus*) and the distinctive characteristics (*differentia*), which are domain dependent. Last, existing guidelines

⁷⁵ The term intensional also presents terminological variation, which can be said to be equivalent to 'definition by analysis' (Sager, 1990), '*définition par inclusion*' (Rey-Debove, 1971) or '*définition spécifique*' (Felber, 1987).

(definitional templates or models; category definitional frame-based approach) created to help write definitions can be found in the literature (e.g., Cabré, 1999; Atkins & Rundell, 2008, Swanepoel, 2010; ISO 704, 2009; Faber, 2012, 2015; Löckinger, Kockaert & Budin, 2015). Further, in the lexicographic literature, we find described principles for drafting a definition (Rey-Debove, 1966; Porto Dapena, 2002; Löckinger, Kockaert & Budin, 2015; Mel'čuk & Polguère, 2018), such as avoiding circularity, inaccuracies or irrelevant characteristics, defining every word used in a definition, complying with the replaceability principle and avoiding ambiguity and definitions in the negative, among others.

PART II

DATA ANALYSIS AND PROCESSING

CHAPTER 6

Coverage and Treatment of Terms in Academy Dictionaries

*C'est qu'un dictionnaire, c'est l'univers par ordre alphabétique.
A bien prendre les choses, le dictionnaire est le livre par excellence.
Tous les autres livres sont là-dedans; il ne s'agit plus que de les en tirer.*
FRANCE (1921)

This chapter is entirely devoted to the coverage and treatment of terms in academy dictionaries. We examined the front matter of the print editions of the DLPC and the DLE (2014), as well as the introductory texts available on the DAF webpage, to ascertain whether explicit references were made to the adopted labelling system and/or to any criterion or justification for the presence of diatechnical information. We explored labelling practices in those three dictionaries, focusing our attention on domain labels. Accordingly, we extracted the domain labels listed in those dictionaries to an Excel sheet. We started with the Portuguese dictionary and then analysed the same aspects in the Spanish and French dictionaries. After reviewing the listed domains, we evaluated whether there was any kind of organisation. We addressed the existing literature and showed how *metalabels* can be used to optimise the alignment of specialised senses in lexicographic works. Although the mapping was manual, this study's multilingual domain map can support future standardisation efforts concerning domain labelling processes and associated encoding tasks across various dictionaries and languages. Finally, we conducted a microstructural analysis comparing the definitions of selected terms from the domains in focus from the different lexicographic resources.

We should emphasise that it was not our main intention to check on the accuracy of the information they contain but only to comment on how they are presented by analysing and comparing them.

6.1 Lexicographic Data Analysis

We adopted a threefold methodology to analyse the chosen domains:

- (i) compilation and lexicographic data analysis: we began by analysing coverage, i.e., the domains included in each dictionary, and moved on to their microstructure, examining how these dictionaries treat terms;
- (ii) comparison between results: to systematise labels and detect overlapping, the compiled domain label lists were compared;
- (iii) domain mapping: we created new metadata to facilitate our analysis, namely a metalabel (the equivalent English term was assigned as a metalabel of the corresponding domain). Using this metalabel, we built a multilingual domain map. The domain labels were then manually mapped using semantic properties, such as exact and related and none.

In short, we aimed to (a) highlight the similarities and differences in the editorial practices of dictionaries and their approaches to knowledge organisation, (b) report on a manual mapping exercise for two particular domains (GEOLOGY and FOOTBALL), which can serve as test cases to establish procedural rules for the alignment of domain labels in general language dictionaries, and (c) highlight the problems and inconsistencies detected, which we will try to resolve in the following chapter with the methodology proposed.

6.1.1 Analysis of the Dictionaries' Front Matter

In methodological terms, the first step was to read the introductory pages, or front matter, of the print editions of the DLPC and the DLE (2014), as well as the introductory texts available on the DAF webpage, to ascertain whether there were explicit references to the treatment of terms, namely to the adopted labelling system and/or to any criterion or justification for the use of those labels. We began with the DLPC and subsequently analysed the same aspect in the DLE (2014) and the DAF.

As far as diatechnical information or domain labels are concerned, the DLPC's '*Introdução*' [Introduction] (pp. XIII–XXIII) describes, in very broad terms, the three types of specialised units registered, which the editors call '*tecnicismos*' [technicisms]:

No Dicionário registam-se ainda: tecnicismos generalizados na linguagem usual; tecnicismos que, embora de uso não generalizado, correspondem a noções ou

classificações e a aparelhos fundamentais em cada ciência ou técnica; tecnicismos que ocorrem em manuais escolares de natureza científica e técnica. (DLPC, p. XIV). [The Dictionary also registers generalised tecnicisms in the usual language; tecnicisms that, although not in general use, correspond to notions or classifications and fundamental devices in each science or technique; tecnicisms that occur in scientific and technical textbooks.]

In the case of the DLE (2014), in only the section ‘*Advertencias*’ [Warnings] (DLE, pp. LI–LIII) is there a brief mention of the labels, informing the user about the decisions made by the lexicographers when ordering meanings within lexicographic articles, whereby they arranged labels according to a specific order: register labels, domain labels, geographic and temporal labels:

De marcación: las acepciones no marcadas tienden a anteponerse a las marcadas. Dentro de estas, van primero las acepciones que tienen marcas correspondientes a los niveles de lengua o registros de habla, después las que llevan marcas técnicas, después las que tienen marcas geográficas (y dentro de ellas, primero las de España y luego las de América y Filipinas) y finalmente las que llevan una marca de vigencia. (DLE, p. LII) [About labelling: unmarked meanings tend to precede marked ones. Among these, the meanings that have labels corresponding to the levels of language or speech registers go first, followed by those that carry technical labels, those having geographical markings (and within them, first those of Spain and then those of America and the Philippines) and finally those with a temporal label.]

Subsequently, we turned our attention to the newly released online AF dictionary, namely to the page ‘*La nouvelle édition numérique du Dictionnaire de l’Académie française, dans ses différentes éditions*’ [The new digital edition of the *Dictionnaire de l’Académie française*, in its various editions], subsection ‘*La 9e édition*’ [The 9th edition] (AF, 2021). Here, we learnt that there has been an ‘[...] *introduction de la métalangue, qui compose un ensemble d’indicateurs linguistiques sur les usages et les domaines d’emploi d’un mot*’ [introduction of metalanguage, which makes up a set of linguistic indicators on the uses and fields of a word’s usage], although no example was provided. Further on, in the subsection entitled ‘*Présentation générale et mise en pages*’ [General presentation and layout], once again, there are some brief references to labels, although their employment is not justified and only their typographic distinction is mentioned:

la différenciation de la ‘métalange’, c’est-à-dire des indicateurs de domaines (maths, beaux-arts, etc.), et des marques d’usage (Fam., Par extension, etc.); ces éléments sont distingués par des attributs typographiques spécifiques.’ [the differentiation of ‘metalanguage’, i.e., indicators of domains (maths, fine arts, etc.), and usage labels (Fam., By extension, etc.), distinguished by specific typographic attributes.]

In this perspective, the new dictionary layout incorporates the list of several abbreviations usually employed on domain names – e.g., ‘BEAUX-ARTS’ [fine arts], ‘PHYSIQUE’ [physics], ‘ASTRONOMIE’ [astronomy] – distinguishing them from other abbreviations by the use of small caps. The editors also seem to distinguish domain labels, such as ‘*metalange*’ [metalanguage] and the remaining usage labels called ‘*marques d’usage*’ [usage labels].

6.1.2 List of Abbreviations

All three dictionaries include lists of abbreviations, but not all abbreviations are labels providing diasystematic information. Salgado, Costa and Tasovac (2019) made an exhaustive manual survey of the abbreviations employed in the three dictionaries, excluding grammatical markers (*adj.*, *n.*, and *v.*) and etymological markers (*esp.*, *lat.*, and *top.*).

After analysing the remaining labels, we compared them and reflected on them. There are two distinct columns in all these dictionaries: one with the abbreviation and the other with the unabbreviated denomination of the label in each language. The complete lists of abbreviations can be found in Annexes 4, 5 and 6.

In the DLPC, abbreviations are listed alphabetically in a section entitled ‘*Abreviaturas*’ [abbreviations] (DLPC, pp. xxxi–xxxiii). We noticed that the labels in this first list identify grammatical categories, etymological markers, different classes of diasystematic information, etc. For example, the labels ‘*antigo*’ [old] and ‘*Neologismo*’ [neologism] indicate diachronic or temporal information, ‘*Regionalismo*’ [regionalism] denotes diatopic or geographical information, and the labels ‘*calão*’ [slang] or ‘*gíria*’ [jargon] refer to diastratic information. Domain names are subsequently included in a separate list entitled ‘*Classificação do vocabulário quanto à repartição por ciências*,

técnicas e formas de actividade’ [Classification of the vocabulary broken down by sciences, crafts and forms of activity] (DPLC, pp. xxxv–xxxvi).

CLASSIFICAÇÃO DO VOCABULÁRIO QUANTO À REPARTIÇÃO POR CIÊNCIAS,
TÉCNICAS E FORMAS DE ACTIVIDADE

a			
<i>Acúst.</i>	= Acústica.	<i>Des.</i>	= Desenho.
<i>Aeron.</i>	= Aeronáutica.	<i>Desp.</i>	= Desporto.
<i>Agr.</i>	= Agricultura.	<i>Diplom.</i>	= Diplomática.
<i>Alg.</i>	= Álgebra.	<i>Dir.</i>	= Direito.
<i>Alveit.</i>	= Alveitaria.	<i>Dir. Can.</i>	= Direito Canónico.
<i>Alven.</i>	= Alvenaria.	<i>Dir. Civil</i>	= Direito Civil.
<i>Anat.</i>	= Anatomia.	<i>Dir. Comerc.</i>	= Direito Comercial.
<i>Antr.</i>	= Antropologia.	<i>Dir. Fiscal</i>	= Direito Fiscal.
<i>Apic.</i>	= Apicultura.	<i>Dir. Intern.</i>	= Direito Internacional.
<i>Arit.</i>	= Aritmética.	<i>Dir. Marít.</i>	= Direito Marítimo.
<i>Arm.</i>	= Armaria.		
<i>Arqueol.</i>	= Arqueologia.	e	
<i>Arquit.</i>	= Arquitetura.	<i>Ecl.</i>	= Eclesiástico.
<i>Artilh.</i>	= Artilharia.	<i>Econ.</i>	= Economia.
<i>Astr.</i>	= Astronomia.	<i>Econ. Pol.</i>	= Economia Política.
<i>Astronáut.</i>	= Astronáutica.	<i>Electr.</i>	= Electricidade.
<i>Astrol.</i>	= Astrologia.	<i>Electrotéc.</i>	= Electrotécnica.
<i>Autom.</i>	= Automobilismo.	<i>Embr.</i>	= Embriologia.
b		<i>Encad.</i>	= Encadernação.
<i>Bact.</i>	= Bacteriologia.	<i>Eng.</i>	= Engenharia.
<i>Ballist.</i>	= Balística.	<i>Equit.</i>	= Equitação.
<i>B.-Art.</i>	= Belas-Artes.	<i>Esc.</i>	= Escolar.
<i>Biol.</i>	= Biologia (Citologia, Histologia).	<i>Escol.</i>	= Escolástica.
<i>Bot.</i>	= Botânica.	<i>Escult.</i>	= Escultura.
<i>Bromat.</i>	= Bromatologia.	<i>Egr.</i>	= Egrima.
c		<i>Espir.</i>	= Espiritualismo.
		<i>Estát.</i>	= Estática.
		<i>Ética.</i>	
		<i>Etnog.</i>	= Etnografia.

Figure 46: Fragment of the DLPC list

The title of the section dedicated to domain labels (Figure 46) led us to think about what distinction the editors of the DLPC list made between ‘*ciências*’ [sciences], ‘*técnicas*’ [techniques] and ‘*formas de actividade*’ [forms of activity]. In finding the domains ALVEITARIA [animal healing], ALVENARIA [masonry], or CUTELARIA [cutlery], we believe this may be the reason for the use of forms of activity.

The DLE (2014) print edition lists all the labels used in a single general list of ‘*Abreviaturas y signos empleados*’ [Abbreviations and symbols used], from which we can also infer microsystems such as diatechnical and diatopic information (Figure 47).

Abreviaturas

a.	alto	affect.	afectivo
abl.	ablativo	afér.	aféresis
abrev.	abreviación	Agr.	agricultura
acep.	acepción	aim.	aimara
acort.	acortamiento	al.	alemán
acrón.	acrónimo	Al.	Álava
act.	activo	Alb.	Albacete
acus.	acusativo	Alm.	Almería
Acús.	acústica	Alq.	alquimia
adapt.	adaptación; adaptado	alterac.	alteración
adj.	adjetivo	alus.	alusión
adv.	adverbio; adverbial	Am.	América
advers.	adversativo	Am. Cen.	América Central
Aer.	aeronáutica	amer.	americano

Figure 47: Fragment of the DLE list

The new edition of the DAF presents a list entitled ‘*Tableau des abréviations utilisées dans le Dictionnaire*’ [Table of abbreviations used in the Dictionary] (Figure 48) in one of the modules of the digital dictionary page.

Tableau des abréviations	
Tableau des abréviations utilisées dans le Dictionnaire (9 ^e édition)	
Dans cette édition, le signe* indique les entrées nouvelles ; le signe ◇ renvoie aux recommandations de 1990. Note : de nombreuses abréviations ont été développées (mises au long) dans l'édition numérique.	
Absolt.	Absolument.
Abusivt.	Abusivement.
ACOUST.	Acoustique.
Adj.	Adjectif ou adjective.
Adj.t.	Adjectivement.
ADM.	Administration.
Adv.	Adverbe.
Advt.	Adverbialement
AÉRON.	Aéronautique.
AGRIC.	Agriculture.
AGRON.	Agronomie.
ALCH.	Alchimie.

Figure 48: Fragment of the DAF list

All three academy dictionaries lack explicit explanatory information regarding their labelling practices. The front matter of each of the DLPC, DLE (2014) and DAF includes only brief references to usage labelling. None of the dictionaries that we

analysed has published explicit criteria for the set of usage labels adopted⁷⁶. While we cannot pass judgment on the individual lexicographic workflows and the lexicographers' internal guidelines to produce these dictionaries, the lack of explicit criteria and an explicit typology of usage labels can affect the user's interaction with and interpretation of the dictionary content.

6.1.3 Exploring Labelling Practices

The task of exploring labelling practices started with the previous comparative study (Salgado & Costa, 2019), in which we only compared the domain labels from Iberian academy dictionaries. Our review of the existing literature (Salgado, Costa & Tasovac, 2019) allowed us to compare the different classifications of diasystematic labels proposed by different researchers and focus on the usage labelling in these scholarly lexicographic works.⁷⁷ We analysed all labels referring to diasystematic information. After collecting all the abbreviations included in the dictionaries, we found that the total number of labels was 438 in the DLPC, 336 in the DLE and 232 in the DAF.

In all these dictionaries, the apparent lack of reasoning for the options provided by the lexicographers prevented us from extending our analysis beyond deduction. However, through these lists, we can infer microsystems composed of diatechnical, diastratic, diaphasic information, etc. Despite little or no information on the selection criteria, by using domain labels, all three lists of abbreviations demonstrate that these general language dictionaries do indeed cover terms.

Given the importance of domain labels for our research, we conducted a thorough survey of all domain labels found in the lists provided by the academy dictionaries under study. In the case of the DLPC, we only had to extract the list shown in Figure 46 regarding the classification of 'specialised vocabulary' (DLPC, pp. xxxv–xxxvi).

⁷⁶ Interestingly, other dictionaries, e.g., *Le Petit Robert de la langue française* (2017) or the *Oxford Advanced Learner's Dictionary* (2014), provide explanations on label usage.

⁷⁷ This work stressed the importance of conducting a detailed analysis of any given dictionary before any lexical data modelling and semantic markup.

6.1.4 Domain Lists

The survey of all domain labels allowed us to determine the number of domains represented in the three dictionaries, both exclusive and shared, and in the case of the Portuguese and Spanish dictionaries, we also determined the frequency of their occurrence. We did not have access to the number of entries per domain in the French dictionary. We also assessed whether the use of domain labels was systematic and whether recent and relevant domains were omitted. As no criteria were found regarding labelling, we were forced to make some assumptions.

There are two different columns in these dictionaries: one containing the abbreviations and the other the domain designations written in full in their respective languages, as shown in Figures 46, 47 and 48.

In typographic terms, academy dictionaries use abbreviations for domain labels. As stressed before, abbreviations are justified by the need to save space in the existing paper editions (cf. Chapter 2, pp. 43–44). The DLPC uses italics, a capital letter and a period; the DLE uses Roman lowercase and a period; and the DAF uses italics, a capital letter and a period. As for the designations written in full, the DLPC and the DAF have uppercase initials, while lowercase initials are used in the DLE (see Table 2).

	Typography
DLPC	abbreviation in italics; full designation in uppercase
DLE	abbreviation in italics; full designation in lowercase
DAF	abbreviation in roman small caps; full designation in uppercase

Table 2: *Comparative typography of domain labels*

After collecting all domain labels included in these dictionaries, the datasets were compiled manually in an Excel sheet. The results are shown in Table 3.

DOMAIN LABELS		
DLPC	DLE	DAF
184	74	132

Table 3: *Domain labels in the three academy dictionaries*

Considering the overall numbers, a certain imbalance in quantitative terms is apparent, which can be explained by the selection of domains, with generic domains coexisting with smaller spectrum domains.

Originally, the DLPC lists 173 domains in the list of abbreviations of the print edition (Annex 4). A closer inspection of the lexicographic articles revealed the presence of labels in the microstructure that were absent from the list of abbreviations: AGRONOMIA [agronomy], BIOQUÍMICA [biochemistry], ECOLOGIA [ecology], ÉTICA [ethics], ETNOLOGIA [ethnology], GINÁSTICA [gymnastics], HISTÓRIA POLÍTICA [political history], MARINHA [navy], METROLOGIA [metrology], PIROTECNIA [pyrotechnics], PSICANÁLISE [psychoanalysis] and TRANSPORTES [transports]. These 11 domains were added to our working list, resulting in 184 domains; however, this number was recalculated after analysing the dictionary's microstructure, since some domains that were listed initially, such as BROMATOLOGIA [bromatology], CIBERNÉTICA [cybernetics], ECONOMIA POLÍTICA [political economy], ESCOLÁSTICA [scholastic], ESPIRITUALISMO [spiritualism], FUTUROLOGIA [futurology], POLÍCIA [police], QUÍMICA BIOLÓGICA [biological chemistry], QUÍMICA ORGÂNICA [organic chemistry], TELEFONIA SEM FIOS [wireless telephony] and VELOCIPEDIA [cycling], were not used in any lexicographic article. We believe that these inconsistencies could be mistakes in the publication of the DLPC. Although these domains did not appear in the printed list, we retained them.

All domains found in the DLPC are displayed in Figure 49.

■ Acúst. Acústica	■ Aeron. Aeronáutica	■ Agr. Agricultura	■ Agron. Agronomia
■ Álg. Álgebra	■ Alveit. Alveitaria	■ Alven. Alvenaria	■ Anat. Anatomia
■ Antr. Antropologia	■ Apic. Apicultura	■ Arit. Aritmética	■ Arm. Armaria
■ Arqueol. Arqueologia	■ Arquít. Arquitetura	■ Artilh. Artilharia	■ Astr. Astronomia
■ Astrol. Astrologia	■ Astronáut. Astronáutica	■ Autom. Automobilismo	■ B.-Art. Belas-Artes
■ Bact. Bacteriologia	■ Balíst. Balística	■ Biol. Biologia (citologia, histologia)	■ Bioquím. Bioquímica
■ Bot. Botânica	■ Bromat. Bromatologia	■ Carn. Carniçaria	■ Carp. Carpintaria
■ Cartog. Cartografia	■ Cerâm. Cerâmica	■ Chapel. Chapelaria	■ Cibern. Cibernética
■ Cinég. Cinégetica	■ Cinem. Cinema, cinematografia	■ Cirurg. Cirurgia	■ Comérc. Comércio
■ Constr. Construção	■ Contab. Contabilidade	■ Corr. Correios	■ Cosmol. Cosmologia
■ Cristallog. Cristalografia	■ Cronol. Cronologia	■ Cul. Culinária	■ Curt. Curtumes
■ Cutel. Cutelaria	■ Des. Desenho	■ Desp. Desporto	■ Diplom. Diplomática
■ Dir. Direito	■ Dir. Can. Direito Canónico	■ Dir. Civil Direito Civil	■ Dir. Comerc. Direito Comercial
■ Dir. Fiscal Direito Fiscal	■ Dir. Intern. Direito Internacional	■ Dir. Marít. Direito Marítimo	■ Eccl. Eclesiástico
■ Ecol. Ecologia	■ Econ. Economia	■ Econ. Pol. Economia Política	■ Electr. Electricidade
■ Electrotéc. Electrotécnica	■ Embr. Embriologia	■ Encad. Encadernação	■ Eng. Engenharia
■ Equit. Equitação	■ Esc. Escolar	■ Escol. Escolástica	■ Escult. Escultura
■ Esgr. Esgrima	■ Espir. Espiritualismo	■ Estat. Estática	■ Ética Ética
■ Etnog. Etnografia	■ Etnol. Etnologia	■ Farm. Farmácia	■ Filol. Filologia
■ Filos. Filosofia	■ Fin. Finanças	■ Fís. Física	■ Fis. Atóm. Física Atómica
■ Fisiol. Fisiologia	■ Fonét. Fonética	■ Fort. Fortificação	■ Fot. Fotografia
■ Fut. Futebol	■ Futur. Futorologia	■ Geneal. Genealogia	■ Genét. Genética
■ Geod. Geodesia	■ Geog. Geografia	■ Geol. Geologia	■ Geom. Geometria
■ Gin. Ginástica	■ Gram. Gramática	■ Heráld. Heráldica	■ Hidrául. Hidráulica
■ Hidrog. Hidrografia	■ Hig. Higiene	■ Hist. História	■ Hist. Nat. História Natural
■ Hist. Polit. História Política	■ Histol. Histologia	■ Hortic. Horticultura	■ Ind. Indústria
■ Inform. Informática	■ Jard. Jardinagem	■ Jur. Jurídico, jurisprudência	■ Ling. Linguística
■ Literat. Literatura	■ Liturg. Liturgia	■ Lóg. Lógica	■ Magn. Magnetismo
■ Mar. Marítima	■ Marc. Marcenaria	■ Mat. Matemática	■ Mecân. Mecânica
■ Med. Medicina	■ Med. Leg. Medicina Legal	■ Metal. Metalurgia	■ Meteor. Meteorologia
■ Métr. Métrica	■ Metrol. Metrologia	■ Mil. Militar	■ Miner. Mineralogia
■ Mitol. Mitologia	■ Mús. Música	■ Náut. Náutica	■ Neur. Neurologia
■ Numism. Numismática	■ Ocult. Ocultismo	■ Olar. Olaria	■ Ópt. Óptica
■ Ouriv. Ourivesaria	■ Paleog. Paleografia	■ Paleont. Paleontologia	■ Parasit. Parasitologia
■ Pat. Patologia	■ Pec. Pecuária	■ Pedag. Pedagogia	■ Pesc. Pesca, pescaria
■ Pint. Pintura	■ Pirotec. Pirotecnia	■ Pisc. Piscicultura	■ Polic. Polícia
■ Polit. Política	■ Pré-hist. Pré-história	■ Psican. Psicanálise	■ Psicofisiol. Psicofisiologia
■ Psicol. Psicologia	■ Psiqu. Psiquiatria	■ Quím. Química	■ Quím. Biol. Química Biológica
■ Quím. Org. Química Orgânica	■ Radiol. Radiologia	■ Rel. Religião	■ Retór. Retórica
■ Sal. Salinas	■ Sapat. Sapataria	■ Seric. Sericicultura	■ Serr. Serralharia
■ Sider. Siderurgia, Siderotecnia	■ Silv. Silvicultura	■ Sociol. Sociologia	■ Tanoar. Tanoaria
■ Taurom. Tauromaquia	■ Teat. Teatro	■ Tecnol. Tecnologia	■ Telecom. Telecomunicações
■ Telef. S. Fios Telefonía sem fios	■ Teleg. Telegrafia	■ Teol. Teologia	■ Teos. Teosofia
■ Terat. Teratologia	■ Tip. Tipografia	■ Topog. Topografia	■ Transp. Transportes
■ Trigon. Trigonometria	■ Veloc. Velocipedia	■ Ven. Venatório	■ Veter. Veterinária
■ Vin. Vinificação	■ Vit. Viticultura	■ Zool. Zoologia	■ Zootec. Zootecnia

Figure 49: Domain labels in the DLPC (184)

Some generic domains and subdomains coexist, including DIREITO [law], DIREITO CANÓNICO [canon law], DIREITO CIVIL [civil law], DIREITO COMERCIAL [commercial law], DIREITO FISCAL [tax law], DIREITO INTERNACIONAL [international law] and DIREITO MARÍTIMO [maritime law]. This also applies to QUÍMICA [chemistry] and QUÍMICA ORGÂNICA [organic chemistry] or MATEMÁTICA [mathematics] and its subdomains GEOMETRIA [geometry], ÁLGEBRA [algebra], ARITMÉTICA [arithmetic] and TRIGONOMETRIA [trigonometry].

In the case of the Spanish dictionary, we extracted the domain labels from the general list of abbreviations (Annex 5). The DLE printed edition lists 72 domains. These domain labels were also worked on during the stay at the RAE's ILEX⁷⁸. During this period, we had access to the *Entorno de Redacción Integrado* (ERI), a computer platform in JAVA and XML that enables the edition of the lexicographic work and allows different kinds of searches. The total number of entries per domain was also obtained and worked out during this stay. After comparing the printed list with the results obtained in ERI, we found some domain labels that were already ignored by the Spanish lexicographers because they had no occurrences (*Cronol.* [chronology]; *Danza* [dance]; *Gen.* [genetics]; *Hist.* [history]; *Náut.* [chronology]); however, we decided to consider two more domain labels that were actually used in their expanded form: ARTE [art] and TEATRO [theatre], bringing the total number of domains to 74.

The domains found in the DLE are depicted in Figure 50.



Figure 50: Domain labels in the DLE (74)

⁷⁸ A research grant sponsored by ELEXIS in November 2018: <https://elex.is/ana-de-castro-salgado/>.

Although the DAF list available online includes 132 domain labels – we also isolated the domain labels from the other labels (Annexe 6) – the total number presented here was obtained by analysing the data provided by the *Académie* itself, to whom we are very grateful for affording us the opportunity to work with the real data contained in their database in 2019 (from letter ‘a’ to ‘savoir’).⁷⁹ We found 12 domain labels on the AF website that were not included in the Excel list provided, but we could not justify their absence: AGRONOMIE [agronomy], CATHOLIQUE [catholic], ESTHÉTIQUE [aesthetics], GRECQUE [greek], HYGIÈNE [hygiene], LÉGISLATION [legislation], OROGRAPHIE [orography], PSYCHOSOCIOLOGIE [psychosociology], RADIOGRAPHIE [radiography], ROMAIN or ROMAINE [roman], SPÉLÉOLOGIE [speleology] and VÉTÉRINAIRE [veterinary]. We assumed their absence must be due to their low frequency of use. If so, it is not clear why these labels are shown on the webpage. In total, 309 domain labels were collected from DAF. However, the contrastive work proceeded, considering only the 132 domains available online, as many questions arose, and we could not find entries illustrating the use of the several domain labels.

The domains available on the DAF webpage are displayed in Figure 51.

⁷⁹ Following a request explaining the scope of this work, the *Académie* shared the list of domain labels for research purposes. We are, therefore, grateful to the academic committee and to Laurent Catach, who was our intermediary during the process and who extracted domain labels with a frequency greater than or equal to five (the others are not representative); the Excel sheet provided contained 297 labels.

■ ACOUST. Acoustique	■ ADM. Administration	■ AÉRON. Aéronautique	■ AGRIC. Agriculture
■ AGRON. Agronomie	■ ALCH. Alchimie	■ ALG. Algèbre	■ ANAT. Anatomie
■ ANTHROPOL. Anthropologie	■ ANTIQ. Antiquité	■ APIC. Apiculture	■ ARBOR. Arboriculture
■ ARCHÉOL. Archéologie	■ ARCHIT. Architecture	■ ARITHM. Arithmétique	■ ARTILL. Artillerie
■ ASTROL. Astrologie	■ ASTRON. Astronomie	■ BALIST. Balistique	■ BÂT. Bâtiment
■ BIOL. Biologie	■ BOT. Botanique	■ BX-ARTS. Beaux-Arts	■ CATHOL. Catholique
■ CH. DE FER. Chemins de fer	■ CHIM. Chimie	■ CHIR. Chirurgie	■ CHORÉGR. Chorégraphie
■ COMPT. Comptabilité	■ CONFIS. Confiserie	■ COURSES Courses de chevaux	■ COUT. Couture
■ CUIS. Cuisine	■ DÉMOGR. Démographie	■ ÉBÉN. Ébénisterie	■ ÉCON. Économie
■ ÉLECTR. Électricité	■ ÉLECTRON. Électronique	■ ENSEIGN. Enseignement	■ ENTOM. Entomologie
■ ÉQUIT. Équitation	■ ESCR. Escrime	■ ESTH. Esthétique	■ ETHNOL. Ethnologie
■ ÉTHOL. Éthologie	■ FÉOD. Féodal ou féodalité	■ FISC. Fiscalité	■ GASTRON. Gastronomie
■ GÉOD. Géodésie	■ GÉOGR. Géographie	■ GÉOL. Géologie	■ GÉOM. Géométrie
■ GRAMM. Grammaire	■ GRAV. Gravure	■ GRECQ. Grecque	■ HÉMAT. Hématologie
■ HÉRALD. Héraldique	■ HIPPOL. Hippologie	■ HIST. Histoire	■ HISTOL. Histologie
■ HORT. Horticulture	■ HYDROGR. Hydrographie	■ HYDROL. Hydrologie	■ HYG. Hygiène
■ INFORM. Informatique	■ JOAILL. Joaillerie	■ LÉGISL. Législation	■ LINGUIST. Linguistique
■ LITURG. Liturgie	■ MATH. Mathématiques	■ MÉCAN. Mécanique	■ MÉD. Médecine
■ MÉTALL. Métallurgie	■ MÉTÉOR. Météorologie	■ MICROBIOL. Microbiologie	■ MILIT. Art militaire ou vie militaire
■ MINÉR. Minéralogie	■ MUS. Musique	■ MYTH. Mythologie	■ NUMISM. Numismatique
■ OCCULT. Occultisme	■ OCÉAN. Océanographie	■ CENOL. Cénologie	■ OPHTALM. Ophtalmologie
■ OPT. Optique	■ ORFÈVR. Orfèvrerie	■ OROGR. Orographie	■ P. ET CH. Ponts et Chaussées
■ PALÉONT. Paléontologie	■ PATHOL. Pathologie.	■ PÂTISS. Pâtisserie	■ PEINT. Peinture
■ PHARM. Pharmacie	■ PHIL. Philosophie	■ PHON. Phonétique	■ PHOT. Photographie
■ PHYS. Physique	■ PHYS. NUCL. Physique nucléaire	■ PHYSIOL. Physiologie	■ POLIT. Politique
■ PRÉHIST. Préhistoire	■ PROS. Prosodie	■ PSYCHAN. Psychanalyse	■ PSYCHIATR. Psychiatrie
■ PSYCHOL. Psychologie	■ PSYCHOPATHOL. Psychopathologie	■ PSYCHOSOCIOL. Psychosociologie	■ RADIOGR. Radiographie
■ RADIOPH. Radiophonie	■ RELIG. Religion	■ RHÉTOR. Rhétorique	■ ROM. Romain ou romaine
■ SC. Sciences	■ SC. NAT. Sciences naturelles	■ SCOLAST. Scolastique	■ SCULPT. Sculpture
■ SOCIOL. Sociologie	■ SPÉLÉOL. Spéléologie	■ STAT. Statistique	■ SYLVIC. Sylviculture
■ TAUROM. Tauromachie	■ TECHN. Technique	■ TÉLÉCOMM. Télécommunications	■ TÉLÉV. Télévision
■ THÉOL. Théologie	■ TRAV. PUBL. Travaux publics	■ TYPOGR. Typographie	■ URBAN. Urbanisme
■ VÈN. Vènerie	■ VÉTÉR. Vétérinaire	■ VITIC. Viticulture	■ ZOOL. Zoologie

Figure 51: Domain labels in the DAF (132)

6.2 Comparison Between Results

The survey of all the domains and the behind-the-scenes work can be accessed on GitHub⁸⁰. As mentioned before, the total of multilingual domain labels collected comprised 184, 74 and 132 domain labels in the DLPC, DLE and DAF, respectively. Clearly, the imbalance in the total number of domains among the three dictionaries is significant. The abundance of subdomains within a general domain indicates a larger number of labels in the DLPC, since there is also a difference of 110 domains vis-a-vis the DLE.

From the analysis, we found that the selection and treatment criteria differ. In the DLE and the DAF, domain labels are used only when the meaning is not considered

⁸⁰ <https://github.com/anacastrosalgado/domain-labelling-in-academy-dictionaries>

to belong to the common lexicon, while in the DLPC, labelling seems to be limited to specifying the domain of a meaning. Take, for example, the first sense of the entries “coração” (DLPC), “corazón” (DLE) and “coeur” (DAF) [heart]. In the DLPC, the domain label for the ANATOMY domain is present, but in the DLE and the DAF, the entries do not have any marking, perhaps because the lexicographers considered them to belong to the general lexicon.⁸¹

To systematise labels and detect any overlaps, when the compiled domain label lists were compared, we found identical abbreviations. However, the abbreviations chosen by the lexicographers behind these dictionaries to represent the same area are not always identical (e.g., *Mús.* is always the abbreviation for the domain of MUSIC, but for the ACOUSTIC domain, we found the abbreviation *Acús.* in the DLE and *Acúst.* in the DLPC). We are indeed aware that our comparison is among three lexicographic resources of different languages, but the proximity of these languages (they are all Romance languages) makes it desirable to propose a homogeneous convention of certain domain labels.

As only DLPC and DLE use abbreviations, Table 4 indicates 18 different abbreviated labels for the same domains in the Portuguese and Spanish dictionaries.

DLPC abbreviation	DLE abbreviation
Acúst.	Acús.
Aeron.	Aer.
Antr.	Antrop.
Arquit.	Arq.
Comérc.	Com.
Desp.	Dep.
Dir.	Der.
Escult.	Esc.
Fonét.	Fon.
Fot.	Fotogr.
Geog.	Geogr.

⁸¹ If we look up the entries ‘vaca’ (‘cow’) and ‘baleia’ (‘whale’) in PRIBERAM and INFOPÉDIA, we will find that they are both identified with the domain ZOOLOGIA [zoology] in INFOPÉDIA, but while the latter has that marker in PRIBERAM, the former has no domain. These types of inconsistencies, unfortunately, are common.

Mecân.	Mec.
Mitol.	Mit.
Psiqu.	Psiquiatr.
Retór.	Ret.
Teat.	Teatro
Telecom.	Telec.
Topog.	Topogr.

Table 4: *Different abbreviations of the same domain labels in the DLPC and DLE*

Table 5 indicates 44 labels and designations that are similarly abbreviated in the Portuguese and Spanish dictionaries.

DLPC abbreviation	DLE abbreviation
Agr.	Agr.
Anat.	Anat.
Arqueol.	Arqueol.
Astrol.	Astrol.
Astr.	Astr.
Biol.	Biol.
Bioquím.	Bioquím.
Bot.	Bot.
Carp.	Carp.
Cineg.	Cineg.
Cinem.	Cinem.
Constr.	Constr.
Ecol.	Ecol.
Econ.	Econ.
Electr.	Electr.
Equit.	Equit.
Esgr.	Esgr.
Fís.	Fís.
Fisiol.	Fisiol.
Geol.	Geol.
Geom.	Geom.
Gram.	Gram.
Heráld.	Heráld.
Inform.	Inform.
Ling.	Ling.
Mar.	Mar.

Mat.	Mat.
Med.	Med.
Meteor.	Meteor.
Métr.	Métr.
Mil.	Mil.
Mús.	Mús.
Numism.	Numism.
Ópt.	Ópt.
Pint.	Pint.
Psicol.	Psicol.
Quím.	Quím.
Rel.	Rel.
Sociol.	Sociol.
Taurom.	Taurom.
Tecnol.	Tecnol.
Transp.	Transp.
Veter.	Veter.
Zool.	Zool.

Table 5: Similar abbreviation labels and domains in the DLPC and DLE

Given that we had quantitative data for Portuguese and Spanish, we were able to detect the seven areas of knowledge with the highest representation (Figure 52).

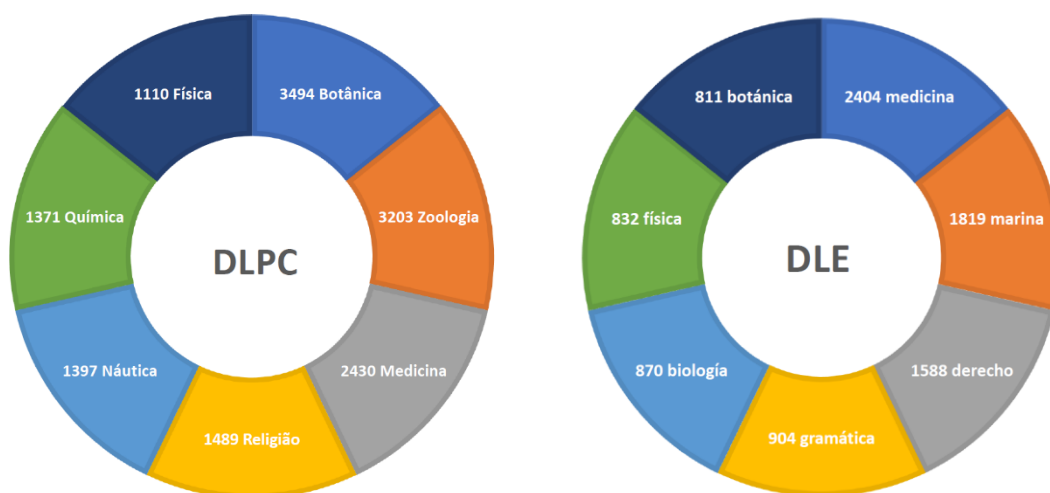


Figure 52: Areas of knowledge with the highest representation in the DLCP and the DLE

The DLPC list was set as the baseline against which the DLE counterpart was compared. Classical domains, such as BOTANY (3494 DLPC vs 811 DLE entries), MEDICINE (2430 DLPC vs 2404 DLE entries) and ZOOLOGY (3203 DLPC entries vs 600 DLE entries), are the most frequent in these dictionaries; they occur in predictable numbers given that these domains have long-standing lexicographic traditions. However, we should question the presence of domains with less representation, such as those with one or two occurrences (see the Portuguese domains in Figure 53 and the DLE's ORTOGRAFÍA [spelling], respectively). Noteworthy are domains with zero occurrences detected in the DLPC (referred to above).

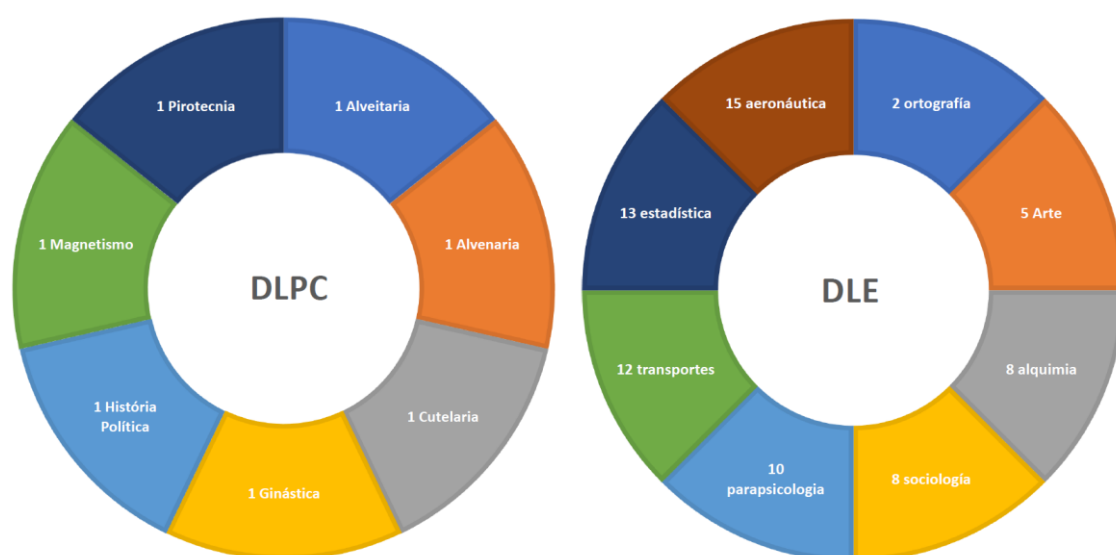


Figure 53: Less frequent domains in the DLPC and the DLE

6.2.1 Mapping Domains

To map domains, we created new metadata to facilitate our analysis, namely a *metatag* (Salgado, Costa & Tasovac, 2021), a tag that identifies the equivalent English designation of the corresponding domain. The English term inserted as metatag corresponds to the domains that will be established in the domain hierarchy (see Chapter 9). This metatag is invisible to the user, but it is handy for search engines and other structures, and specially for our proposal of hierarchical domains. Using this metatag, we were able to build the multilingual domain map.

The domain labels were then manually mapped using semantic properties, such as exact (identify an equivalent domain) and related (points to a generic domain) and none (in cases where we did not find any relation). Our starting point was always the DLPC data, before being confronted with data from the DLE and the DAF.

METALABEL	DLPC	RELATION	DLE	RELATION2	DAF
acoustics	Acústica	EXACT	acústica	EXACT	Acoustique
aeronautics	Aeronáutica	EXACT	aeronáutica	EXACT	Aéronautique
agriculture	Agricultura	EXACT	agricultura	EXACT	Agriculture
anatomy	Anatomia	EXACT	anatomía	EXACT	Anatomie
anthropology	Antropologia	EXACT	antropología	EXACT	Anthropologie
archæology	Arqueologia	EXACT	arqueología	EXACT	Archéologie
architecture	Arquitectura	EXACT	arquitectura	EXACT	Architecture
astrology	Astrologia	EXACT	astrología	EXACT	Astrologie
astronomy	Astronomia	EXACT	astronomía	EXACT	Astronomie
biochemistry	Bioquímica	EXACT	bioquímica	EXACT	biochimie
biology	Biologia (citologia, histologia)	EXACT	biología	EXACT	Biologie
botany	Botânica	EXACT	botánica	EXACT	Botanique
building	Construção	EXACT	construcción	EXACT	Bâtiment
bullfighting	Tauromaquia	EXACT	tauromaquia	EXACT	Tauromachie
carpentry	Carpintaria	EXACT	carpintería	EXACT	charpenterie, charpente.
chemistry	Química	EXACT	química	EXACT	Chimie
cinema	Cinema, cinematografia	EXACT	cinematografía	EXACT	cinéma
commerce	Comércio	EXACT	comercio	EXACT	commerce
computing	Informática	EXACT	informática	EXACT	Informatique
ecology	Ecologia	EXACT	ecología	EXACT	écologie
economics	Economia	EXACT	economía	EXACT	Économie
electricity, electronics	Electricidade	EXACT	electricidad; electrónica	EXACT	Électricité
fencing	Esgrima	EXACT	esgrima	EXACT	Escrime
fine arts	Belas-Artes	EXACT	Arte	EXACT	Beaux-Arts
geography	Geografia	EXACT	geografía	EXACT	Géographie
geology	Geologia	EXACT	geología	EXACT	Géologie
geometry	Geometria	EXACT	geometría	EXACT	Géométrie
grammar	Gramática	EXACT	gramática	EXACT	Grammaire
heraldry	Heráldica	EXACT	heráldica	EXACT	Héraldique
horseriding	Equitação	EXACT	equitación	EXACT	Équitation
law	Direito	EXACT	derecho	EXACT	droit
linguistics	Linguística	EXACT	lingüística	EXACT	Linguistique
mathematics	Matemática	EXACT	matemáticas	EXACT	Mathématiques
mechanics	Mecânica	EXACT	mecánica	EXACT	Mécanique
medicine	Medicina	EXACT	medicina	EXACT	Médecine
meteorology	Meteorologia	EXACT	meteorología	EXACT	Météorologie
military	Militar	EXACT	milicia	EXACT	Art militaire ou vie militaire
music	Música	EXACT	música	EXACT	Musique
mythology	Mitologia	EXACT	mitología	EXACT	Mythologie
navy	Marinha	EXACT	marina	EXACT	marine
numismatics	Numismática	EXACT	numismática	EXACT	Numismatique
optics	Óptica	EXACT	óptica	EXACT	Optique
painting	Pintura	EXACT	pintura	EXACT	Peinture
philosophy	Filosofia	EXACT	filosofía	EXACT	Philosophie
phonetics	Fonética	EXACT	fonética; fonología	EXACT	Phonétique
photography	Fotografia	EXACT	fotografía	EXACT	Photographie
physics	Física	EXACT	física	EXACT	Physique
physiology	Fisiologia	EXACT	fisiología	EXACT	Physiologie
psychiatry	Psiquiatria	EXACT	psiquiatria	EXACT	Psychiatrie
psychology	Psicologia	EXACT	psicología	EXACT	psychologie
religion	Religião	EXACT	religión	EXACT	Religion
rhetoric	Retórica	EXACT	retórica	EXACT	Rhétorique
sculpture	Escultura	EXACT	escultura	EXACT	Sculpture
sociology	Sociologia	EXACT	sociología	EXACT	Sociologie
sport	Desporto	EXACT	deportes	EXACT	sports
telecommunications	Telecomunicações	EXACT	telecomunicación	EXACT	Télécommunications
theatre	Teatro	EXACT	teatro	EXACT	théâtre
topography	Topografia	EXACT	topografía	EXACT	topographie
transport	Transportes	EXACT	transportes	EXACT	transports
veterinary	Veterinária	EXACT	veterinaria	EXACT	Vétérinaire
zoology	Zoologia	EXACT	zoología	EXACT	Zoologie

Table 6: Domains (metalabels) with an exact correspondence (61)

Our analysis revealed that there are currently 61 domains in common in the three dictionaries (Table 6), which we propose to study in the future. These 61 domains

were mapped to an equivalent domain, that is, we assigned an exact property. Classical domains such as BOTANY, MEDICINE, and ZOOLOGY, *inter alia*, were found in these dictionaries, which seem to point to a certain lexicographic tradition.

We used the `related` tag to indicate domains that may share a potential alignment, detecting a possible hierarchical relationship with a generic domain. Table 7 shows some of the domains found.

METALABEL	DUPC	RELATION	DLE	RELATION2	DAF
algebra	Álgebra	RELATED matemáticas		EXACT	Algèbre
ancient physics		RELATED física		RELATED física	Physique ancienne
arithmetic	Aritmética	RELATED matemáticas		EXACT	Arithmétique
atomic physics	Física Atómica	RELATED física		RELATED Physique	
bacteriology	Bacteriologia	RELATED biología		EXACT	bactériologie
cellular biology		RELATED biología		RELATED biología	biologie cellulaire
confectionery		RELATED culinária		RELATED Cuisine	Confiserie
crystallography	Cristalografia	RELATED geologia		RELATED Géologie	
embryology	Embriologia	RELATED biología, medicina		EXACT	embryologie
football	Futebol	RELATED deportes		EXACT	football
genetics	Genética	RELATED biología, medicina		EXACT	génétique
geomorphology		RELATED geologia		RELATED geologia	géomorphologie
gymnastics	Ginástica	RELATED deportes		NONE	
histology	Histologia	RELATED biología, medicina		EXACT	Histologie
magnetism	Magnetismo	RELATED física		RELATED Physique	
mineralogy	Mineralogia	RELATED geologia		EXACT	Minéralogie
nuclear physics		RELATED física		RELATED física	Physique nucléaire
paleontology	Paleontologia	RELATED geologia		EXACT	Paléontologie
parasitology	Parasitologia	RELATED biología, medicina		RELATED Biologie, Médecine	
statics	Estática	RELATED física		RELATED Physique	
statistics		RELATED matemáticas		RELATED Mathématiques	statistique
trigonometry	Trigonometria	RELATED matemáticas		RELATED Mathématiques	

Table 7: A portion of domain labels with a *related* correspondence

We assigned the tag `none` when no match was found, as exemplified in Table 8.

METALABEL	DUPC	RELATION	DLE	RELATION2	DAF
animal healing	Alveitaria	NONE		NONE	
bookbinding	Encadernação	NONE		NONE	
cooperage	Tanoaria	NONE		NONE	
cutlery	Cutelaria	NONE		NONE	
cycling	Velocipedia	NONE		NONE	
fish farming, pisciculture	Piscicultura	NONE		NONE	
futurology	Futurologia	NONE		NONE	
gardening	Jardinagem	NONE		NONE	
hydraulics	Hidráulica	NONE		NONE	
masonry	Alvenaria	NONE		NONE	
nautical	Náutica	NONE		NONE	
pedagogy	Pedagogia	NONE		NONE	
philology	Filologia	NONE		NONE	
police	Polícia	NONE		NONE	
pottery	Olaria	NONE		NONE	
psychophysiology	Psicofisiologia	NONE		NONE	
saltworks	Salinas	NONE		NONE	
school	Escolar	NONE		NONE	
siderurgy	Siderurgia, siderotecnia	NONE		NONE	
tanneries	Curtumes	NONE		NONE	
telegraphy	Telegrafia	NONE		NONE	
winemaking, vinification	Vinificação	NONE		NONE	
wireless telephony	Telefonia Sem Fios	NONE		NONE	
zootechnics	Zootecnia	NONE		NONE	

Table 8: A portion of domain labels without any correspondence, *none*

Not considering the domain label abbreviations that do not match, we accounted for 65 shared domains between the DLPC and the DLE, as shown in Figure 54.

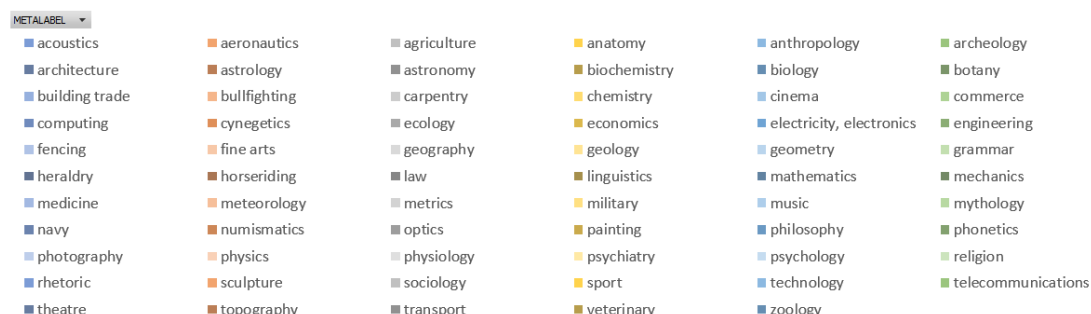


Figure 54: DLPC vs DLE – Correspondence between domain labels in both dictionaries (65)

Between the DLPC and the DAF, the number of domain matches was 136, as shown in Figure 55.



Figure 55: DLPC vs DAF – Correspondence between domain labels in both dictionaries (136)

We compared the domains in the DLE with those in the DAF, revealing 53 shared domains (Figure 56).



Figure 56: DLE vs DAF – Consensus between domain labels in both dictionaries (53)

While the list of abbreviations is ordered alphabetically in a conventional manner, which is a practical resource to determine the location of a particular label, we advocate a prior conceptual organisation of their labels and decoding of their respective values.

6.2.2 Domain Organisation

As stated by Costa (2013), ‘Specialised communication, whether monolingual or multilingual, is not solely a matter of language, it is also a matter of knowledge’ (p. 40). After reviewing the flat domains list, we evaluated whether there was a discernible knowledge organisation. We could only make assumptions in most cases, given the lack of introductory and explanatory texts on the methodology and criteria followed.

As mentioned above, though there is no hierarchical classification of domains, it is possible to detect coexisting generic domains and subdomains. The imbalance referred to can be explained thus: while the DLE has only generic domains (e.g. DEPORTES [sports], GEOLOGÍA [geología]), the DLPC and the DAF register multiple subdomains and even multiple labels for the same or very similar domains (e.g., COURSES DE CHEVAUX and COURSES HIPPIQUES [horse races] in the DAF). Conversely, the high number of domains in the DAF seems to result from a continuous addition of domain labels throughout the successive editions without eliminating outdated markers.

Using the DLPC as the baseline, we noted the case of MATHEMATICS and its subdomains, ALGEBRA (DLPC, DAF), ARITHMETIC (DLPC, DAF), GEOMETRY (DLPC, DLE, DAF) and TRIGONOMETRY (DLPC) or STATISTICS (DLE, DAF). GEOLOGY was also found to have branches considered subdomains of a generic domain. It includes CRYSTALLOGRAPHY (DLPC), MINERALOGY (DLPC and DAF) and PALAEONTOLOGY (DLPC and DAF). The corresponding dictionary definitions for each of these terms (“geology”, “crystallography”, “mineralogy” and “palaeontology”) will be compared to clarify, if possible, the underlying rationale for these subdivisions.

geologia [ʒjuluʒiɐ]. *s. f.* (De *geo-* + gr. *λόγος* ‘estudo’ + suf. *-ia*). **1.** Ciência que estuda a Terra desde a sua origem e formação até à actualidade, a sua estrutura e as sucessivas transformações dos elementos que a compõem. **2.** (com maiúsc.). Disciplina ou curso em que se estuda essa ciência. *Licenciado em Geologia*. **3.** Tratado ou compêndio sobre a ciência geológica. **4.** Características geológicas de uma zona determinada, que são objecto de estudo dessa ciência.

Figure 57: Entry ‘geologia’ [geology] in the DLPC (ACL)

geología

De *geo-* y *-logía*.

1. f. Ciencia que estudia la historia del globo terrestre, así como la naturaleza, formación, evolución y disposición actual de las materias que lo componen.

Figure 58: Entry ‘geología’ [geology] in the DLE (RAE)

GÉOLOGIE nom féminin

XVIII^e siècle. Composé de *géo-* et de *-logie*, tiré du grec *logos*, « discours, traité ».

■ Science qui a pour objet l'étude de l'écorce terrestre, de sa formation, de la nature et de la situation des roches et terrains qui la composent. *Traité de géologie. La minéralogie, la géochimie sont des disciplines de la géologie. Géologie dynamique, voir Géodynamique.*

■ Par métonymie. Se dit des réalités physiques qui constituent l'objet de cette science. *La géologie des Ardennes.*

Figure 59: Entry ‘géologie’ [geology] in the DAF (AF)

In Figures 57, 58 and 59, no domain label can be found. This non-marking is to be expected since the label would be identical to the lemma itself. However, the label could be inserted in the data, thereby not being made available to the user, as we will explain in the Chapter 9. The marking makes it easier for the lexicographer to control terminological data. The usage example from the DAF ‘*La minéralogie, la géochimie sont des disciplines de la géologie.*’ [Mineralogy, geochemistry are disciplines of geology] is notable because MINERALOGY and GEOCHEMISTRY may be considered subdomains of the generic GEOLOGY domain.

cristalografia [kri/ʔɛlɔgrɐfjɐ]. *s. f.* (Do gr. κρύσταλλος ‘cristal’ + suf. *-grafia*). *Miner.* Ciência que estuda e descreve a forma e a estrutura dos cristais, bem como as leis que regem a sua formação.

Figure 60: Entry ‘cristalografia’ [crystallography] in the DLPC (ACL)

cristalografia

Del gr. κρύσταλλος *krýstallōs* ‘cristal’ y *-grafia*.

1. *f. Geol.* Descripción de las formas que toman los cuerpos al cristalizar.

Figure 61: Entry ‘cristalografia’ [crystallography] in the DLE (RAE)

CRISTALLOGRAPHIE nom féminin

XVIII^e siècle. Composé de *cristallo-*, tiré du grec *krustallos*, « cristal », et de *graphie*.

Étude géométrique de la structure des cristaux et des propriétés qui en résultent.

Figure 62: Entry ‘cristallographie’ [crystallography] in the DAF (AF)

The “crystallography” entries (Figures 60, 61 and 62), when compared, are more challenging. The DLPC indicates that it belongs to the MINERALOGY domain. CRYSTALLOGRAPHY can indeed be considered a branch of MINERALOGY; however, the use of

CRYSTALLOGRAPHY as a domain label is questionable. On the other hand, the DLE identifies the term as belonging to the domain of GEOLOGY, more generically; this position will be defended later on. The DAF, however, indicates no domain label.

mineralogia [miniɾɛluʒiɐ]. *s. f.* (De *mineral* + *-logia*).
1. Ciência que trata dos minerais. 2. Obra ou compêndio que trata dos minerais. *Teve que comprar outra mineralogia pois a que tinha estava desatualizada.*

Figure 63: Entry ‘mineralogia’ [mineralogy] in the DLPC (ACL)

mineralogía

De *mineral* y *-logía*.

1. *f.* Ciencia que estudia los minerales.

Figure 64: Entry ‘mineralogía’ [mineralogy] in the DLE (RAE)

MINÉRALOGIE nom féminin

XVII^e siècle. Composé à l’aide de *minéral* et de *-logie*, tiré du grec *logos*, « discours, traité ».

Science qui a pour objet l’étude des minéraux.

Figure 65: Entry ‘mineralogie’ [mineralogy] in the DAF (AF)

The treatment given to the “mineralogy” entries (Figures 63, 64 and 65) is somewhat similar in the three dictionaries, regardless of the unmarked senses.

paleontologia [pɐljõtuɫuʒiɐ]. *s. f.* (De *paleo-* + gr. *ὄν, ὄντος* ‘ser’ + suf. *-logia*). Ciência que se ocupa do estudo dos fósseis animais e vegetais.

Figure 66: Entry ‘paleontologia’ [paleontology] in the DLPC (ACL)

paleontología

De *paleo-*, *onto-* y *-logía*.

1. f. Ciencia que estudia los organismos que han existido en el pasado de la Tierra a partir de sus restos fósiles.

Figure 67: Entry 'paleontología' [paleontology] in the DLE (RAE)

PALÉONTOLOGIE nom féminin

XIX^e siècle. Composé à l'aide de *paléo-*, d'*onto-*, tiré du grec *ôn*, *ontos*, « étant, ce qui est », et de *-logie*, tiré du grec *logos*, « discours, traité ».

Science ayant pour objet l'étude des êtres vivants des temps géologiques, qu'on ne trouve plus qu'à l'état fossile. *Paléontologie humaine, animale, végétale. La paléontologie du Jurassique, du Crétacé. Musée de paléontologie.*

Figure 68: Entry 'paléontologie' [paleontology] in the DAF (AF)

The case of “palaeontology” is similar (Figures 66, 67 and 68). None of the dictionaries use a domain label to mark these entries. When comparing the treatment of the entries “crystallography” and “paleontology”, it seems to mean that the unmarked meanings may be due to the fact that they are defined as sciences, that is, independent domains.

Without any type of marking, the possibility of establishing relationships among the analysed entries is null; such relationships can only be inferred based on the knowledge that the user may have of the domain in question.



We have performed the same analysis for the “football” dictionary entries to check whether any label is used.

futebol [futiból]. *s. m.* (Do ingl. *football*). *Desp.* 1. Modalidade desportiva em que jogam duas equipas, de onze jogadores cada uma, que procuram controlar e introduzir a bola na baliza da equipa adversária. *O futebol é a modalidade desportiva mais popularizada em Portugal. Jogo, desafio de +; clube, equipa, jogador de +; campo de +; jogar +.* **futebol americano**, modalidade desportiva semelhante ao rúgbi, praticada entre duas equipas de onze jogadores cada uma, obtendo pontos aquela que conseguir transportar, com a mão, a bola oval para lá da linha de ensaio adversária. «*O futebol americano é uma modalidade bizarra para quem não está dentro das suas regras*» (DN, 26.1.1997). **futebol de salão**, modalidade desportiva semelhante ao futebol, praticada num campo de menores dimensões, em piso de madeira ou cimento, com uma bola mais pequena, entre duas equipas de cinco jogadores cada uma. «*quinta jornada do Campeonato de Futebol de Salão da I Divisão*» (Record, 22.10.1991). 2. Estilo e técnica de jogar futebol. *Futebol português. Futebol brasileiro.* **não ir em futebolis.** *Fam.*, não se deixar levar; não aderir sem uma justificação séria.

Figure 69: Entry ‘futebol’ [football] in the DLPC (ACL)

fútbol

Tb. **futbol**.

Del ingl. *football*.

1. m. Juego entre dos equipos de once jugadores cada uno, cuyo objetivo es hacer entrar en la portería contraria un balón que no puede ser tocado con las manos ni con los brazos, salvo por el portero en su área de meta.

fútbol, o futbol americano

1. m. Deporte de origen estadounidense, más parecido al *rugby* que al **fútbol**, cuyos jugadores llevan aparatosas protecciones para la cabeza y el cuerpo dada la violencia del juego.

fútbol, o futbol sala

1. m. Modalidad de **fútbol**, con cinco jugadores por equipo, en la que las dimensiones de la cancha y de las porterías son muy similares a las del balonmano.

fútbol, o futbol siete, o fútbol, o futbol 7

1. m. Modalidad de **fútbol**, con siete jugadores por equipo, que se disputa en un terreno de juego cuyas dimensiones son la mitad de las de uno ordinario, y con unas porterías algo más pequeñas que las de este.

futbol

V. **fútbol**.

Figure 70: Entries ‘fútbol/futbol’ [football] in the DLE (RAE)

FOOTBALL (se prononce *foutbol*) **nom masculin**

xvii^e siècle. Mot anglais composé de *foot*, « pied », et *ball*, « balle, ballon ».

Désignait initialement tout jeu de ballon opposant deux équipes. Aujourd’hui, est réservé au jeu mettant aux prises deux équipes de onze joueurs qui doivent faire entrer un ballon, sans le toucher de la main ou du bras, dans les buts défendus par le camp adverse. *Un terrain de football. Un joueur, un club de football. Football amateur, professionnel. La Coupe de France, la Coupe du monde de football.*

Figure 71: Entry ‘football’ [football] in the DAF (AF)

In figures 69, 70 and 71, the DLPC is the only dictionary that indicates the domain label *Desp.* or *SPORT* domain. The other dictionaries do not contain any label.



As a preliminary concluding remarks, despite the undeniable importance of usage labels in lexicographic resources, our analysis of the selected academy dictionaries revealed inconsistencies that can generally be attributed to the absence of an explicitly outlined methodology.

These and many other dictionaries could be improved if they unequivocally explained the lexicographic criteria used in the process of including diasystematic information in entries. In the introductions to all three dictionaries analysed above, the references to the inclusion and processing of this type of information are practically non-existent or too generic. The number of labels selected by the lexicographers for these dictionaries is also unequal. The theoretical background of the choices made by the lexicographers can hardly be extrapolated from a plain list of the abbreviations used.

6.3 Geology and Football Domains: Analysis of Lexicographic Articles

As an exhaustive study of all domain labels is beyond the scope of this thesis, we chose two different domains for the analysis of lexicographic articles: FOOTBALL and GEOLOGY.

6.3.1 Geological Terms

Geological terms were selected to formulate arguments supporting the need for and advantage of establishing conceptual and semantic relationships between lexicographic articles, and to verify the definitions of those terms.

Upon consulting the geological time scale, the term “Phanerozoic”⁸² was selected (Figure 72); however, its French equivalent was not found.

⁸² ‘The uppermost eonothem of the Standard Global Chronostratigraphic Scale. It comprises the Palaeozoic, Mesozoic and Cenozoic erathems, which include rocks with abundant evidence of life. Further, the time during which these rocks were formed, the Phanerozoic Eon, covers the time period between 540 Ma and the present.’ (Neuendorf, Mehl Jr. & Jackson, 2011, p. 486).

fanerozóico¹, **a** [fɐnɐɾɔzjúku, -ɐ]. *adj.* (Do gr. φανερός 'visível' + suf. *-zoico*). *Geol.* Que pertence ou é relativo ao período do Fanerozóico.

Fanerozóico² [fɐnɐɾɔzjúku]. *s. m.* (Do gr. φανερός 'visível' + suf. *-zoico*). *Geol.* Período geológico de vida manifesta no planeta, posterior ao Criptozóico, que abarca os últimos 570 milhões de anos da história da Terra, compreendendo os períodos Paleozóico, Mesozóico, Cenozóico e Antropozóico, que se caracteriza pelo surgimento dos primeiros fósseis e pelo aparecimento do homem.

fanerozoico

Del ingl. *Phanerozoic*, y este del gr. φανερός *phanerós* 'visible', 'evidente', zoo-'zoo-' e -ic-'ico'.

eón fanerozoico

eón fanerozoico

1. m. Geol. eón más reciente, que abarca desde hace 542 millones de años hasta la actualidad.

Figure 72: Entries '*fanerozóico*' and '*fanerozoico*' [*Phanerozoic*] in the DLPC (ACL) and in the DLE (RAE)

In the DLPC, there are two entries belonging to different parts of speech – an adjective (*adj.*) and a masculine noun (*s. m.*). Concerning the entry with superscript number 2, after the domain label, while the definition starts with '*período geológico*' [geological period], there is no reference to the fact that it is an eonothem/eon. The DLE, in turn, has a cross-reference for "eón" [phanerozoic eon], i.e., the definition of this term can be found only in the entry "eón" [eon], as we can see in Figure 72. The lexicographic definition begins with the word '*eón*'. However, Phanerozoic is not described as an eonothem.

We shall now consider some entries related to the geological term "era" (the geochronologic equivalent of an "erathem"⁸³). The following comparative analysis begins with the DLPC (Figure 73), proceeds to the DLE (Figure 74) and finally considers the DAF (Figure 75).

⁸³ See Chapter 9. Chronostratigraphic units. *Stratigraphic Guide*. International Commission on Stratigraphy 'eras carry the same name as their corresponding erathems'. Retrieved from <https://stratigraphy.org/guide/chron>.

era [ére]. 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, TERCIÁRIO.

Figure 73: Fragment of the entry ‘era’ [era] in the DLPC (ACL)

The DLPC defines this geological term as per Sense 4, introduced by the domain label, *Geol.*, from GEOLOGIA [geology]. The lexicographic definition starts with ‘*Cada uma das grandes divisões do tempo geológico*’ [Each of the great geological time divisions]. In the DLPC, after the definition, the four great eras from a paleontological perspective are recorded as polylexical units, or ‘*combinatórias fixas*’ [fixed combinations] which is the term used by the lexicographers of the DLPC. Sorted alphabetically, these eras are: the “era primária” [primary era], “era quaternária” [quaternary era], “era secundária” [secondary era] and “era terciária” [tertiary era]. Each of these areas has a definition followed by synonyms in small capitals: “PALEOZÓICO”, “PRIMÁRIO” [palaeozoic, primary] for the primary era; “ANTROPOZÓICO”, “QUATERNÁRIO” [antropozoic, quaternary] for the quaternary era; “MESOZÓICO”, “SECUNDÁRIO” [Mesozoic, secondary] for the secondary era, and “CENOZÓICO”, “TERCIÁRIO” [Cenozoic, tertiary] for the tertiary era.

era¹ +

3. f. Cada uno de los grandes períodos de la evolución geológica o cósmica. *Era cuaternaria. Era solar.*

Figure 74: Fragment of the entry ‘era’ [era] in the DLE (RAE)

In turn, the DLE does not use the domain label. Following the lexicographic definition ‘*Cada uno de los grandes períodos de la evolución geológica o cósmica*’ [Each

of the great periods of geological or cosmic evolution] – in which the reference to the geological domain can be found in the expression ‘*evolução geológica*’ [geological evolution] –, it presents two examples highlighted using italics and a different colour: ‘*Era cuaternaria. Era solar.*’ [Quaternary era. Solar era.], i.e., while the DLPC registers these polylexical terms as sublemmas, the DLE illustrates their use only as a usage example.

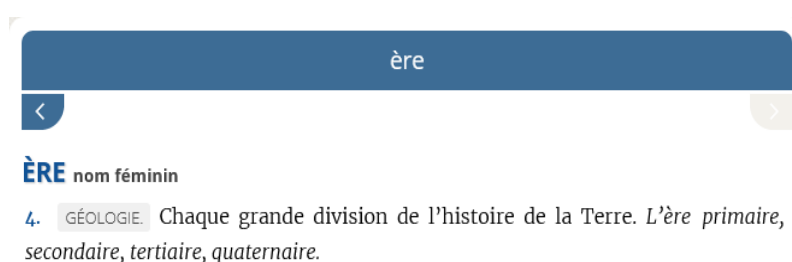


Figure 75: Fragment of the entry ‘ère’ [era] in the DAF (AF)

Finally, the DAF, using the domain label GÉOLOGIE, has the same components as the DLE; it has opted to register the polylexical units as usage examples in italics: ‘*L’ère primaire, secondaire, tertiaire, quaternaire*’ [The primary, secondary, tertiary, quaternary era].

In short, the presence and omission of the GEOLOGY domain label have been verified, and a different way for representing the current polylexical terms has been found, appearing either as a sublemma or as an example. Concerning lexicographic definitions, some reservations concerning scientific precision remain. However, this topic will be explored in the next chapter.

We can proceed to the analysis of “Palaeozoic”⁸⁴, “Mesozoic”⁸⁵ and “Cenozoic”⁸⁶ (i.e., the erathems/eras comprised by the “Phanerozoic”) – Figures 76, 77 and 78.

paleozóico¹, a [paljɔzɔjku, -v]. *adj.* (De *paleo-* + gr. ζῷον 'animal' + suf. *-ico*). Que é relativo à era primária ou ao período geológico do Paleozóico. ≈ PRIMÁRIO.

Paleozóico² [paljɔzɔjku]. *s. m.* (De *paleo-* + gr. ζῷον 'animal' + suf. *-ico*). *Geol.* Divisão cronológica da história da Terra, anterior ao Mesozóico, que abarca os primeiros 345 milhões de anos do éon fanerozóico, compreendendo os períodos Câmbrio, Ordovício, Silúrio, Devonico, Carbonífero e Permiano, que se caracteriza por uma grande diversificação da fauna, com o desenvolvimento dos invertebrados e o aparecimento dos primeiros peixes, batráquios, insectos e répteis. ≈ ERA PRIMÁRIA, PRIMÁRIO.

mesozóico¹, a [mezɔzɔjku, -v]. *adj.* (De *meso-* + gr. ζῷον 'animal' + suf. *-ico*). *Geol.* Que é relativo à era secundária ou ao período geológico do Mesozóico. ≈ SECUNDÁRIO.

Mesozóico² [mezɔzɔjku]. *s. m.* (De *meso-* + gr. ζῷον 'animal' + suf. *-ico*). *Geol.* Divisão cronológica da história da Terra, posterior ao Paleozóico e anterior ao Cenozóico, que engloba cerca de 160 milhões de anos, compreendendo os períodos Cretáceo, Jurássico e Triásico e que se caracteriza pelo aparecimento de grandes répteis, aves e primeiros mamíferos, bem como pelas grandes transformações geológicas que conduziram à distribuição actual dos continentes. ≈ ERA SECUNDÁRIA, SECUNDÁRIO.

cenozóico¹, a [senɔzɔjku, -v]. *adj.* (Do gr. καινός 'novo' + ζῷον 'animal' + suf. *-ico*). *Geol.* Que pertence ou é relativo à era terciária ou ao período geológico do Cenozóico. ≈ TERCIÁRIO, Período +.

Cenozóico² [senɔzɔjku]. *s. m.* (Do gr. καινός 'novo' + ζῷον 'animal' + suf. *-ico*). *Geol.* Divisão cronológica da história da Terra, anterior ao Antropozóico e posterior ao Mesozóico, que engloba cerca de 65 milhões de anos, compreendendo os períodos Neogénico e Paleogénico e que se caracteriza pelo aparecimento dos primeiros primatas, pelo desenvolvimento e crescente domínio dos mamíferos e pelo arrefecimento progressivo do clima. ≈ ERA TERCIÁRIA, TERCIÁRIO.

Figure 76: Entries ‘paleozóico’ [palaeozoic], ‘mesozóico’ [mesozoic], ‘cenozóico’ [cenozoic] in the DLPC (ACL)

In the DLPC, “paleozóico”, “mesozóico” and “cenozóico” have two entries – an adjective (*adj.*) and a masculine noun (*s. m.*). After the domain label, all definitions begin with ‘divisão cronológica da história da Terra’ [chronological division of the Earth’s history] – clearly the lexicographers followed the same strategy – and include the designations of the periods included in the “Palaeozoic” era: ‘compreendendo os períodos’ [comprising the periods]. In the end, synonyms appear in small capital (“ERA

⁸⁴ ‘The lowest erathem of the *Phanerozoic* Eonothem of the Standard Global Chronostratigraphic Scale, above the *Precambrian* and below the *Mesozoic*. Furthermore, the time during which these rocks were formed, the Palaeozoic Era, covers the time period between 540 and 250 Ma.’ (Neuendorf, Mehl Jr. & Jackson, p. 467).

⁸⁵ ‘The middle erathem of the *Phanerozoic* Eonothem of the Standard Global Chronostratigraphic Scale, above the *Palaeozoic* and below the *Cenozoic*. Furthermore, the time during which these rocks were formed, the Mesozoic Era, covers the time period between 250 and 65 Ma.’ (Neuendorf, Mehl Jr. & Jackson, p. 406).

⁸⁶ ‘The upper erathem of the *Phanerozoic* Eonothem of the Standard Global Chronostratigraphic Scale, above the *Palaeozoic* and below the *Cenozoic*. Furthermore, the time during which these rocks were formed, the Cenozoic Era, covers the time period between 65 Ma and the present. It is characterised paleontologically by the evolution and abundance of mammals and angiosperm plants.’ (Neuendorf, Mehl Jr. & Jackson, p. 105).

PRIMÁRIA”, “PRIMÁRIO” [primary era, primary]; “ERA SECUNDÁRIA”; “SECUNDÁRIO” [secondary era, secondary], and “ERA TERCIÁRIA”, “TERCIÁRIO” [tertiary era, tertiary].

paleozoico, ca

De *paleo-*, *zoo-* e *-ico*.
Escr. con may. inicial c. s. en acep. 1.

1. adj. *Geol.* Dicho de una era geológica: Que abarca desde el fin del Precámbrico, hace unos 570 millones de años hasta hace unos 230 millones de años, y comprende sucesivamente los períodos cámbrico, ordovícico, silúrico, devónico, carbonífero y pérmico. U. t. c. s. m.

2. adj. *Geol.* Perteneciente o relativo al **Paleozoico**.

mesozoico, ca

De *meso-*, *zoo-* e *-ico*.
Escr. con may. inicial c. s. en acep. 1.

1. adj. *Geol.* Dicho de una era geológica: Que abarca desde el fin del Paleozoico, hace unos 230 millones de años, hasta hace unos 65 millones de años, y que comprende sucesivamente los períodos triásico, jurásico y cretácico. U. t. c. s. m.

2. adj. *Geol.* Perteneciente o relativo al **Mesozoico**.

cenozoico, ca

Del gr. *καινός kainós* ‘nuevo’, *zoo-* e *-ico*.
Escr. con may. inicial c. s. en acep. 1.

1. adj. *Geol.* Dicho de una era geológica: Que abarca desde el fin del Mesozoico, hace unos 65 millones de años, hasta la actualidad, y comprende los períodos terciario y cuaternario. U. t. c. s. m.

2. adj. *Geol.* Perteneciente o relativo al **Cenozoico**.

Figure 77: Entries ‘paleozoico’ [Palaeozoic], ‘mesozoico’ [Mesozoic], ‘cenozoico’ [Cenozoic] in the DLE (RAE)

The DLE registers each of the “paleozoico”, “mesozoico” and “cenozoico” entries as an adjective with two senses. Senses 1 and 2 are diatechnically marked (*Geol.*). Sense 1 begins with the formula ‘*dicho de una era geológica*’ [said of a geological era], and following the colon, it presents the definition of the said era. At the end, this term is also used as a noun (U. t. c. s. m.). Sense 2, also marked with the domain label ‘*Perteneciente o relativo al Paleozoico*’ [Belonging to or regarding the Palaeozoic], is surprising because it seems to have the same meaning comprehended in Sense 1.

paléozoïque

PALÉOZOÏQUE nom masculin

XIX^e siècle. Composé à l’aide de *paléo-* et du grec *zōikos*, « relatif aux animaux ».

■ **GÉOLOGIE** Ère géologique d’une durée d’environ 340 millions d’années, caractérisée par l’apparition de la vie en milieu terrestre et aérien (on dit aussi Ère primaire). Le Paléozoïque s’étend du Cambrien au Permien.

• Adjectivement. L’ère paléozoïque. Des terrains, des couches, des fossiles paléozoïques.

mésozoïque

* **MÉSOZOÏQUE** adjectif

XIX^e siècle. Composé de *méso-* et de *-zoïque*, tiré du grec *zōikos*, « qui concerne les êtres vivants ».

■ **GÉOLOGIE** Relatif à l’ère secondaire. Terrain, couche mésozoïque. Subst. Le Mésozoïque, l’ère secondaire.

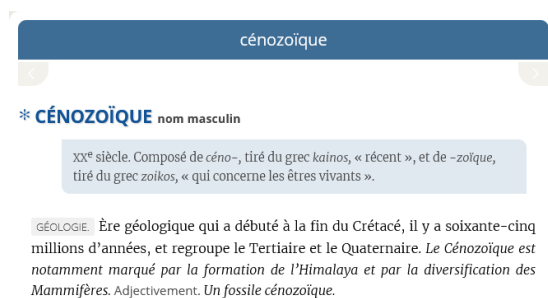


Figure 78: Entries ‘paléozoïque’ [palaeozoic], ‘mésozoïque’ [mésozoïque], ‘cénozoïque’ [Cenozoic] in the DAF (AF)

In the DAF, “paléozoïque” and “cénozoïque” are classified as masculine nouns (*nom masculin*), but “mésozoïque” is classified as an adjective (*adjectif*). The domain label GÉOLOGIE appears in the three entries. After the domain label, the lexicographic definition of the nouns begins with ‘Ère géologique’ [geological era]. The information understood between the curved brackets – (‘on dit aussi Ère primaire’) [(we also say primary era)] – is notable because it functions as a type of synonym but belongs to the lexicographic definition. A usage example appears after this: ‘Le Paléozoïque s’étend du Cambrien au Permien.’ [The Palaeozoic stretches from the Cambrian to the Permian]. The adjectival function of this term, which is not indicated in the usual ‘part of speech’ field, is indicated in the following line and introduced by ‘Adjectivement’ [Adjectively], followed by some examples.

The “Palaeozoic” is divided into six periods: “Cambrian”, “Ordovician”, “Silurian”, “Devonian” and “Carboniferous”. The term “Carboniferous”⁸⁷ was chosen for this analysis.

carbonífero, a [kərbunífɐɾu, -ɐ]. *adj.* (De *carbo*, -*ōnīs* ‘carvão’ + -*fero*). **1.** Que contém ou produz carvão. *Região carbonífera. Florestas carboníferas. Terrenos carboníferos.* **2.** *Geol.* O m. que *carbónico*¹. *Flora e fauna carboníferas. Período carbonífero.*

Figure 79: Entry ‘carbonífero’ [Carboniferous] in the DLPC (ACL)

⁸⁷ ‘A system of the late *Paleozoic* Erathem of the Standard Global Chronostratigraphic Scale, above the *Devonian* and below the *Permian*.’ (Neuendorf, Mehl Jr. & Jackson, p. 98).

“Carbonífero” (Figure 79), in the DLPC, appears only as an adjective; an entry for the term as a noun does not exist (which may have been a slip). The meaning we were interested in – sense 2, marked with the domain label – contains a cross-reference to the “carbónico” (Figure 80) [also Carboniferous in English] entry, introduced by the expression ‘*O m. que*’ [the same as]⁸⁸. Moreover, it is worth noting that one of the examples after the cross-reference is ‘*Período carbonífero*’ [Carboniferous period], which is the term we are seeking.

carbónico¹, **a** [kərbóniku, -ɐ]. *adj.* (Do fr. *carbonique*).
 1. Que se refere a um composto químico que contém carbono. **anidrido**⁺ **carbónico**. **gás**⁺ **carbónico**. **neve**⁺ **carbónica**. 2. *Geol.* Que é relativo ao período geológico do Carbónico. ≈ CARBONÍFERO.
Carbónico² [kərbóniku]. *s. m.* (Do fr. *carbonique*). *Geol.* Período da era primária ou paleozóica que sucede ao devónico e que antecede o pérmico, caracterizando-se pelo aparecimento dos primeiros répteis e insectos alados. *Durante o carbónico desenvolveram-se grandes bosques de fetos.*

Figure 80: Entry ‘carbónico’ [Carboniferous] in the DLPC (ACL)

“Carbónico” has two entries: one as an adjective and the other as a noun. The definition as a noun begins with the word ‘*período*’ [período], illustrated by a usage example in italics: ‘*Durante o carbónico desenvolveram-se grandes bosques de fetos.*’ [During the carbonic period large forests of ferns were developed.]

Meanwhile, the DLE registers “carbonífero” (Figure 81) as an adjective with three senses.

⁸⁸ We will not comment on the preference given to “Carbónico” versus “Carbonífero” because it is beyond the scope of this work. In this regard, we only refer that the current Portuguese official chronological table prefers “Carbónico”, probably because the term “carbónico” is enshrined in national classical geological terminology (e.g., Lima, 1895/98; Teixeira, 1944; Fleury, 1922; Carrington da Costa, 1931; Lemos de Sousa, 1961. In contrast, the “Carbonífero” spelling is equally valid, and more recently adopted by several national and Brazilian geological schools (e.g., Legoinha, 2008; Pais & Rocha, 2010; Pinto de Jesus et al. 2011; Cunha et al. 2012).

carbonífero, ra

Del lat. *carbo*, *-ōnis* 'carbón' y *-fero*.

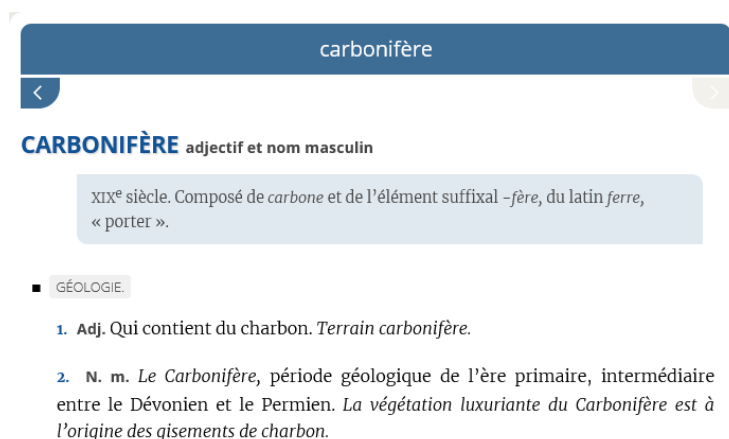
Escr. con may. inicial c. s. en acep. 2.

1. **adj.** Dicho de un terreno: Que contiene carbón mineral.
2. **adj. Geol.** Dicho de un período: Quinto de la era paleozoica, que abarca desde hace 360 millones de años hasta hace 286 millones, y se caracteriza por la aparición de los reptiles y grandes bosques pantanosos que dieron lugar a los yacimientos de carbón mineral. **U. t. c. s. m.**
3. **adj. Geol.** Perteneciente o relativo al **Carbonífero**.

Figure 81: Entry 'carbonífero' [Carboniferous] in the DLE (RAE)

Senses 2 and 3 in the DLE are diatechnically marked (*Geol.*). Sense 2 begins with the formula '*dicho de un periodo*', and after the colon, it presents the definition of the said period. At the end, we found an indication that this term is also used as a noun ('*U. t. c. s. m.*'). Sense 3, also marked with the domain label, '*Perteneciente o relativo al Carbonífero*' [Belonging to or regarding the Carboniferous], is surprising because it seems to have the same meaning as that understood in Sense 2.

In the French lexicographic article, "carbonifère" (Figure 82), the term is classified as an adjective and a noun within a single entry.



carbonifère

CARBONIFÈRE adjectif et nom masculin

XIX^e siècle. Composé de *carbone* et de l'élément suffixal *-fère*, du latin *ferre*, « porter ».

■ GÉOLOGIE

1. **Adj.** Qui contient du charbon. *Terrain carbonifère.*
2. **N. m.** *Le Carbonifère*, période géologique de l'ère primaire, intermédiaire entre le Dévonien et le Permien. *La végétation luxuriante du Carbonifère est à l'origine des gisements de charbon.*

Figure 82: Entry 'carbonifère' [Carboniferous] in the DAF (AF)

The meaning we are interested in is Sense 2, classified as *N. m.* [masculine noun]. Before examining the definition of the term, it is important to note that there is a very intriguing component, '*Le Carbonifère*', in italics. The definition is illustrated below with an example: '*La végétation luxuriante du Carbonifère est à l'origine des gisements de charbon.*' [The lush vegetation of the Carboniferous is the origin of the coal deposits.]

In short, despite having carried out an exercise that included entries corresponding to eras/erathemas, the relationship that can and should be established between them is not visible to the user.

6.3.2 Football Terms

We began with the hypothesis that FOOTBALL is to be integrated into the generic domain of SPORTS. The same understanding can be used for other sports that are included in dictionaries. We can say that SPORTS is a general domain that can be subdivided into different branches (which in turn are domains that function as subdomains within a certain hierarchical organisation).

Aiming to understand if the domain label football is justifiable, we randomly selected some terms related to football. In the DLPC's list of abbreviations, we found the label *Fut.* (FUTEBOL [football] written in full); we identified 120 entries with that label. In the DLE, although the domain FÚTBOL [football] is not listed, the label DEPORTES [sports] does include terms relevant to football. Of the 1915 entries marked with *Dep.*, we selected 147 in which the term "fútbol" appears in the lexicographic article's microstructure (specifically in the lexicographic definition component). In the case of the DAF, since we cannot directly access the diatechnically marked lexicon, we searched for the same units found in the Portuguese and Spanish dictionaries.

First, we examined the position of the labels inside the lexicographic article.

In the DLPC, the existence of domain labels is noteworthy; thus, we decided to choose this dictionary to analyse this topic. We identified different situations:

- (i) the domain label appears after the entry; therefore, all sense components are covered by that label (Figure 83).

águia⁴ [ágjɐ]. *s. m. e f.* (Do lat. *aquila*). *Gír. Fut.* **1.** Clube desportivo Sport Lisboa e Benfica, que tem a águia como insígnia. **2.** O m. que *benfiquista*².

Figure 83: Entry ‘águia’ [eagle; supporter of Sport Lisboa e Benfica sports club] in the DLPC (ACL)

(ii) the domain label appears after a numbered meaning, so it refers only to that specific meaning and explicitly differentiates polysemy cases (Figure 84).

chapéu [ʃɛpɛw]. *s. m.* [...] **13. Fut.** Remate que descreve um arco por cima da cabeça do guarda-redes.

Figure 84: Entry ‘chapéu’ [chip] in the DLPC (ACL)

(iii) the domain label may also be placed before polylexical lexical units, such as the polylexical unit “grande penalidade” (Figure 85), which appears under the dictionary entry ‘grande’.

grande¹ [grɛ̃di]. *adj. m. e f.* [...] **grande penalidade**, *Fut.*, castigo aplicado por falta grave, cometida na grande área e que se traduz por um remate directo à baliza, a partir de onze metros de distância, sendo a bola apenas defendida pelo guarda-redes.

Figure 85: Entry ‘grande penalidade’ [penalty kick] in the DLPC (ACL)

We concluded from the above that the position of the label is not random. The label can cover all the senses of a lexicographic article (e.g., Figure 83) or particular ones (e.g., Figures 84 and 85).

However, labelling is not always regular. We found the same type of articles with and without labels. Thus, within the microstructure of the dictionary, there is no systematic use of the labels. For two good examples, let us look at two lexicographic

articles related to the positions of the football players in the field, which are ‘extremo’ [winger] and ‘lateral’ [back], and compare them in the three dictionaries.

These units, which could be seen as terms, are marked diatechnically. In the DLPC, ‘extremo’, sense 9, has the domain label *Desp.* while ‘lateral’, sense 2, has the domain label *Fut.* (Figure 86).

extremo² [ɛ́/trému ~ ɐj/trému]. *s. m.* [...]

9. *Desp.* Jogador de basquetebol... que actua junto à linha lateral. «*po tituição de Nenê por Torres veio tirar esclarecimento que, que passou a ter um ponta de lança a mais e a mo a menos.*» (DP, 16.6.1971).

lateral² [lɐ'tɾál]. *s. m. e f.* (Do lat. *lateralis*). **1.** O ao lado de alguma coisa. **2.** *Fut.* Jogador que actua da linha lateral do campo. «*Nos primeiros minutos alguma oscilação e nervosismo por parte dos laterais da equipa*» (DN, 27.10.1988). + *direito*,

Figure 86: Entries ‘extremo’ [winger] and ‘lateral’ [back] in the DLPC (ACL)

Comparing these two entries, in “extremo”, the lexicographer does not use the label *Fut.*, probably because the definition details the different types of modalities that use the term (‘*jogador de futebol, basquetebol...*’ [football, basketball player]) while in “lateral” no particular sport is specified. In this case the lexicographic definition exerts an influence on the label assignment.

Thus, within the microstructure of the dictionary, it seems that for the same kind of lexical units, the diatechnical marking differs. Moreover, in some cases, this might be because these entries may have been edited by different lexicographers who eventually did not have a defined methodology to follow.

Let us now examine these two same lexical units in the DLE (Figure 87):

extremo, ma +

14. m. y f. En el fútbol y otros deportes, miembro de la delantera que, en la alineación del equipo, se sitúa más próximo a las bandas derecha o izquierda del campo.

lateral +

3. adj. Dicho de un futbolista o de un jugador de otros deportes: Que actúa junto a las bandas del terreno de juego con funciones generalmente defensivas. *U. t. c. s.*

Figure 87: Entries ‘extremo’ [winger] and ‘lateral’ [back] in the DLE (RAE)

These units are not marked diatechnically. Instead, the lexicographers have chosen another mechanism – the introduction of restrictive expressions in the definition text (Porto Dapena, 2002, p. 308): in “extremo”, ‘*En el fútbol y otros deportes*’ [In football and other sports]]; in “lateral”, ‘*dicho de un futbolista*’ [said of a football player].

We can assume that the lexicographers do not regard these units as properly terminological, since they are used in everyday language, and so they do not indicate any domain label. These units are treated as non-specialised lexical units currently used. There are, however, other entries associated with football that bear the DEPORTES [sports] label. Two examples are “aríete” [striker] and “gol contra” [goal against], which are marked with the DEPORTES label.

In this sense, from the analysed dictionaries we conclude that the choice of the DLE is to mark the units diatechnically only when the meaning belongs explicitly to a specialised context in cases where Spanish speakers will not easily recognise those units.

Let us now consider the equivalent examples in the DAF (Figure 88):



Figure 88: Entries ‘ailier’ [winger] and ‘arrière’ [back] in the DAF (AF)

In terms of markings, what we had observed in the DLPC is repeated in the DAF. “Ailier” does not have any domain label, while “arrière” shows the SPORTS label. The context of sport is indicated in both definitions through the words ‘*sport d’équipe*’ [team sport]. As we do not find any justification for the use of the domain label also in this dictionary, it is not clear why, in the first case, there is no label and in the second, the meaning is marked.

It is clear, from the corpus analysed, that the DLPC distinguishes itself from the DLE and the DAF by using the domain label more frequently to differentiate meanings

or contextualise them by specifying the domain of meaning. We cannot hazard any opinions as to the different criteria. In fact, any criterion can be validated if applied uniformly.

Our analysis confirms that domain labels point to terms, and the three dictionaries use linguistic formulae in the definition, which have the same functions as domain labels. An example is the expression '*jogador de futebol*' [football player], with the introduction of restrictive expressions in the text of the definition through collocations such as '*no jogo do futebol*' [in a football game], '*en el fútbol y otros deportes*' [in football and other sports], '*dans les sports d'équipe*' [in team sports], or of the type '*aplicado a... se aplica a...*' [applied to... applies to...] or '*dicho de un futbolista*' [said of a football player]. There are also cases where more than one mechanism is used simultaneously.

For the end user, the presence of linguistic formulae in the definition is an interesting strategy; however, for the lexicographer the processing of the data may be rendered difficult, as the coherence of the lexicographic resource could be affected. In principle, if the criterion in a given dictionary is to mark domains with a label, we understand that another criterion should not be chosen, including that domain in the definition, without marking it thematically with an appropriate label. On the other hand, computational tools require some coherence so that the lexicographer can properly control this type of information, filtering the dictionary by a domain and exporting all related labels. Thus, one possibility would be to retain these linguistic formulae and mark those meanings with the domain, even with a label that can be invisible to the user.

Continuing to examine the entries related to the FOOTBALL domain, we analysed and compared the behaviour of some units belonging to the conceptual field of 'fan'. This analysis included only the Portuguese and Spanish dictionaries, since we did not find any of the collected French units (e.g., *Les Girondins* (Bordeaux); *Les Canaris* (Nantes); *Les Grenoblois* (Grenoble); *Les Lions* (Sochaux); *Les Merlus* (Lorient); *Les Pailladins* (Montpellier); *Les Bisontins* (Besançon)).

We start with the DLPC, by analysing three different lexicographic articles.

gilista¹ [ʒilífɿtɐ]. *adj. m. e f.* (De *Gil* <*Vicente Futebol Clube*> + suf. *-ista*). **1.** Que pertence ou diz respeito ao clube desportivo Gil Vicente Futebol Clube. «*A equipa profissional sénior gilista voltou, ontem de manhã bem cedo, ao trabalho*» (*A Bola*, 29.3.1997). «*as boas intenções gilistas não incomodavam os sagazes salgueiristas, sempre perigosos no ataque.*» (*Record*, 7.4.1997). **2.** Que é jogador, adepto ou simpatizante desse clube. «*a recarga do capitão gilista foi bem sucedida.*» (*A Bola*, 5.5.1997).

Figure 89: Entry 'gilista' [supporter of Gil Vicente Futebol Clube] in the DLPC (ACL)

In the first example, 'gilista' (Figure 90), there is no domain label.

leão⁴ [liẽw]. *s. m.* (Do lat. *leo*, *-ōnis* <símbolo do Sporting Clube de Portugal>). *Gír. Fut.* O m. que *sportinguista*². «*os leões venceram com grande dificuldade*» (*A Bola*, 5.5.1997). «*há 14 anos que enverga a camisola do Sporting! É um leão de alma e coração!*» (*A Bola*, 29.3.1997).

Figure 90: Entry 'leão' [lion; supporter of Sporting Club de Portugal] in the DLPC (ACL)

In the second example, 'leão' (Figure 90), we find two different labels: *Gír.*, from *Gíria* [jargon] and *Fut.*, from FÚTEBOL. The jargon⁸⁹ label is perhaps justified because the unmarked unit is 'sportinguista', while 'leão' belongs to football jargon. The fact that we have a cross-reference seems to indicate that the DLPC lexicographers preferred the neutral term and not the metaphorical one. This topic also brings us to the question of the language of sport supporters.

⁸⁹ By jargon we mean special lexical units used by a specific social community, group or profession that are difficult for others to understand. It contains diastatic information, referring to socio-cultural group. Cf. Pérez Pascual (2012, p. 192): 'lenguajes sectoriales o jergas profesionales, que utilizan los miembros de un determinado colectivo dedicado' [sectorial or professional language, using members of a specific dedicated collective].

portista¹ [purtíʃtɐ]. *adj. m. e f.* (De <Futebol Clube do> Porto <clube desportivo> + suf. *-ista*). *Fut.* **1.** Que diz respeito ou pertence ao clube desportivo Futebol Clube do Porto. ≈ AZUL-E-BRANCO, DRAGÃO. «*A equipa portista teve, pelo meio, alguns jogos*» (*A Bola*, 16.4.1997). «*a justiça da vitória portista*» (*A Bola*, 5.5.1997). **2.** Que é jogador, adepto ou simpatizante desse clube. ≈ AZUL-E-BRANCO, DRAGÃO. «*tentativas de agressão por parte dos jogadores portistas*» (*Público*, 27.3.1997).

Figure 91: Entry 'portista' [supporter of Futebol Clube do Porto] in the DLPC (ACL)

The third example, 'portista' (Figure 91), has only the domain label.

For the DLE, we also present three selected lexicographic articles related to football team supporters: 'colchonero' (Figure 92), a supporter of Atlético de Madrid, 'culé' (Figure 93), a supporter of Barça, and 'merengue' (Figure 94), a supporter of Real Madrid.

colchonero, ra

- 1.** *adj.* **colchonero**. Perteneciente o relativo a los colchones.
 - 2.** *adj.* **colchonero**. Perteneciente o relativo al Atlético de Madrid, sociedad deportiva española.
 - 3.** *adj.* **colchonero**. Jugador o seguidor del Atlético de Madrid. **U. t. c. s.**
 - 4.** *m. y f.* Persona que tiene por oficio hacer o vender colchones.
- aguja colchonera**

Figure 92: Entry 'colchonero' [supporter of Atlético de Madrid] in the DLE (RAE)

culé

Quizá del cat. *culer* 'trasero, culera'.

- 1.** *adj.* **culé**. Perteneciente o relativo al Fútbol Club Barcelona.
- 2.** *adj.* **culé**. Jugador o seguidor del Fútbol Club Barcelona. **U. t. c. s.**

Figure 93: Entry 'culé' [supporter of Fútbol Club Barcelona] in the DLE (RAE)

merengue

Del fr. *meringue*.

1. *m.* Dulce, por lo común de forma aovada y de color blanco, hecho con claras de huevo y azúcar y cocido al horno.
2. *m.* Baile originario de la República Dominicana, extendido a otros países del Caribe y convertido en baile popular.
3. *m.* Persona de complexión delicada.
4. *m. coloq. Arg., Bol., Par. y Ur.* Lío, desorden, trifulca.
5. *adj. coloq.* Perteneciente o relativo al Real Madrid Club de Fútbol.
6. *adj. coloq.* Jugador o seguidor del Real Madrid Club de Fútbol. *U. t. c. s.*

Figure 94: Entry ‘merengue’ [Real Madrid Club de Fútbol] in the DLE (RAE)

What attracted our attention was the use of the register label, *coloq.* [colloquial] in ‘colchonero’ and ‘merengue’, and its absence among supporters of Barcelona, ‘culé’ (Figure 93), and, for example, ‘periquito’ (another example not illustrated here), a fan of Real Club Deportivo Español de Barcelona. Even so, the treatment of these entries is very systematic. All entries are treated as adjectives (*adj.*) with the indication that they can be used as names (*U.t.c.s.*).



Again, when we compared the DLPC and DLE entries, we found that they are characterised by the absence or presence of the domain label. However, as we have seen, domain labels are useful for the user and the lexicographer and, therefore, it would be important to normalise this treatment. This type of harmonisation will become increasingly important as we move toward linking standards-compliant structured lexical data sets to create accessible and interoperable lexicographic resources.

The entries related to football fans pose another lexicographic issue: the possibility of including encyclopaedic information in general language dictionaries. Why people call the supporters of Futebol Clube do Porto ‘dragões’ [dragons] or the fans of Atlético de Madrid ‘colchoneros’ may be one of the reasons for an end user to look up that entry. This explanation is not found in any of the consulted dictionaries but could be provided in an appropriate field or even in a usage example, as we will demonstrate in the next chapter. We will argue that it makes sense to include this type of information in these lexicographic works, as long as it is properly considered and substantiated.

Having found many entries in the football context that belong to the semantic field of its fans also brings us to the conclusion that in FOOTBALL – such a popular domain – and contrary to GEOLOGY – a highly-specialised domain – there is a strong propensity for another register – jargon.

We will now focus on football terms referring to positions occupied by football players on the field. Table 9 lists some terms in Portuguese related to positions, with their equivalents in Spanish and French⁹⁰. We have marked their presence (✓) or absence (-) in our lexicographic corpus.

Abbrev.	Portuguese	French	Spanish	English [literal translation]	DLPC	DAF	DLE
GUARDA-REDES (gardien de but, portero, goalkeeper)							
GR	guarda-redes	gardien de but	portero	goalkeeper	✓	✓	✓
DEFESA (defense, defensa, defence)							
LD	lateral direito	arrière latéral droit	lateral derecho	right-back	—	—	—
LE	lateral esquerdo	arrière latéral gauche	lateral izquierdo	left-back	—	—	—
DC	defesa central	défenseur central	defensa central	centre-back	—	—	—
LB	líbero	libéro	líbero	sweeper	✓	—	✓
MEIO-CAMPO (centro del campo, milieu du terrain, midfield)							
MD*	médio defensivo	milieu défensif	volante de corte	defensive midfielder	—	—	—
MD	médio direito	milieu droit	volante externo derecho	right midfielder	—	—	—
ME	médio esquerdo	milieu gauche	volante externo izquierdo	wide midfielder	—	—	—
MC	médio-centro	milieu central	volante central	centre midfielder	—	—	—
MO	médio ofensivo	milieu offensif	volante ofensivo	attacking midfielder	—	—	—
ATAQUE (ataque, attaque, attack)							
AV	avariado	avant	delantero	attacker	✓	—	✓
SA	segundo-avariado	deuxième avant	segundo delantero	second striker	—	—	—
PL	ponta de lança	buteur	punta	striker	✓	—	✓
ED	extremo-direito	aillier droit	extremo derecho	right-winger	—	—	—
EE	extremo-esquerdo	aillier gauche	extremo izquierdo	left-winger	—	—	—

Table 9: Terms referring to positions occupied by football players on the field

⁹⁰ The translation into English is used here only for the purpose of making the text clearer.

According to Table 9, only the term “goalkeeper” is recorded in all these dictionaries. Most terms that designate the positions of the players are not recorded in our dictionaries, e.g., “right-back”, “left-back”, “centre-back”, “right-winger” and “left-winger”. We may argue that this is because we are dealing with polylexical units, such as “left-back”, and not just with monolexical units, such as “back” in English, “lateral” in Portuguese and Spanish, and “lateral” in French. Consequently, we decided to search for these units in our lexicographic corpus. The unit “lateral”, when related to football, is included in the DLPC (*‘Fut. Jogador que actua junto da linha lateral do campo’* [Player acting near the sideline] and in the DLE (*‘Dicho de un futbolista o de un jugador de otros deportes: Que actúa junto a las bandas del terreno de juego con funciones generalmente defensivas’* [Said of a football player or a player of other sports: One that acts along the sidelines with generally defensive functions] but is absent from the DAF.

The term “goalkeeper” (included in all these dictionaries) raises some controversial questions. Although the DLPC uses *Fut.* (FOOTBALL) as a domain label listed in the abbreviation list, in the case of “guarda-redes”, the domain label used is *Desp.* (SPORTS) (*‘Desp. Jogador que, no jogo do futebol, andebol, hóquei... ocupa o último posto de defesa, entre os postes da baliza, tentando impedir a marcação de golos’* [Player who, in football, handball, hockey..., occupies the last defense position between the goal posts, trying to prevent the scoring of goals]. This happens because the definition presented above is related not only to the FOOTBALL domain but also to other sports. In the DLE, “portero” is not identified by any label (*‘Jugador que en algunos deportes defiende la portería de su bando’* [Player who, in some sports, defends the goal on their side]. Finally, the DAF uses the SPORTS label (*‘SPORTS. Gardien de but, joueur assurant la défense des buts dans certains jeux de ballon’* [Goalkeeper, player defending goals in certain ball games]).

The DLPC and the DAF distance themselves from the DLE by using the domain label to differentiate meanings or contextualise them, merely specifying the domain of the meaning.

To avoid such inconsistencies, a terminological approach to the domain that dictates a prior organisation of knowledge and establishes relationships between concepts and terms, and, in turn, between different terms would be of major help. As

such, building a concept system by identifying the relations between the concepts that embody the positions occupied by football players would allow the lexicographers to compile all the terms designating them. A conceptual approach to domains, as we will demonstrate in the next chapter, prevents lexicographers from missing essential terms.

6.4 Final Considerations

While the labelling system is a delicate issue within a particular lexicographic resource, the difficulty increases when we compare different resources – comparing labels in different dictionaries, we found that the adopted criteria diverge, making their role unclear (Béjoint, 1988, p. 360). Not everyone endorses the same labels, and their usage is sometimes quite disparate. We must recall and stress our initial premise: ‘there is quite a lot of work involved in putting together a consistent policy on labels in a dictionary’ (Atkins & Rundell, 2008, p. 231). To make matters worse, many dictionaries do not justify the chosen usage labels. The introductory pages of the print editions fail to provide hints or explicit references to the adopted labelling system and/or to any criterion or justification for the usage labels. The application of a labelling system is not always entirely consistent within individual dictionaries and even less so across different lexicographic projects, hindering the tasks of accurately classifying and encoding them. Moreover, this difficulty is composed of the differences and partial incompatibilities found in the lexicographic literature on diasystematic information processing. Ptaszyński (2010), in an article on the causes of the unsatisfactory theoretical treatment of diasystematic information in dictionaries, considers that lexicographers ‘have been searching in vain for an exhaustive and precise answer to the questions of which words to label in what kind of dictionaries and how to do it’ (p. 411). He goes on to state how these problems result from a ‘lack of a firm theoretical basis for the application of diasystematic information (i.e., information about restrictions on usage) in dictionaries’ (Ptaszyński, 2010, p. 411). In many cases, due to the absence of explanations in the introductions, it is challenging to discover the actual value of labels, and it follows that lexicographers, most often than not, simply reproduce them following a certain tradition.

Here, we decided to compare not only the lexicographic data from DLPC but also to compare that data with the DLE and the DAF. Our work on these three dictionaries detected the problematic use of the following:

- domains with multiple labels (e.g., football terms) were found to be classified under the SPORT and FOOTBALL labels in the DLPC (e.g., *libero* [sweeper] in SPORT and *lateral* [back] in FOOTBALL);
- unlabelled equivalent headwords (e.g., *paleozóico* [palaeozoic] adj.) was unlabelled and *primário* [primary] adj., a synonym, appears with a GEOLOGY label;
- combinations of labels referring to closely related domains, such as *antracite* [anthracite] were associated with both MINERALOGY and GEOLOGY or *glaciar* [glacier], associated with both the GEOLOGY and GEOGRAPHY domains;
- despite the similarity of the languages, the abbreviations are not always identical, e.g., the case of the ACOUSTICS domain, which is marked *Acús.* in the DLPC and *Acúst.* in the DLE, or RHETORIC, marked as *Ret.* and *Retór.*, respectively;
- as far as terms referring to football club supporters are concerned, we consider that, besides the domain label, those senses should be marked with the jargon label, i.e., a sociocultural label should be used, identifying the appropriation of a given lexical unit by a particular social group.

Such specificities can lead to numerous issues that complicate data sharing, aligning and linking.

There is an urgent need to review the labelling system, eliminating unnecessary or repetitive labels, as well as those distinctions that, because they are too fine, can sometimes seem arbitrary from the viewpoint of both a lexicographer and a regular dictionary user. Inconsistencies were also observed in using the abbreviated forms, as sometimes they are used but not on other occasions. Other mechanisms also are utilised to mark specialised information, such as the use of formulae present in the definition, sometimes even using more than one mechanism simultaneously.

The consistency of usage labels in dictionaries will significantly improve if every label used is adequately justified, its scope well-delimited in the dictionary outside

matter, and the overall editorial approach to labelling is explained in greater detail than is currently the case.

For a consensus on the best practices towards optimising the labelling process in scholarly dictionaries, it would be desirable for lexicographers to collaborate on the future harmonisation of usage labels across different dictionaries and different languages. This type of harmonisation will become increasingly important as we move towards the mutual linking of standard-compliant structured lexical data sets to create accessible and interoperable lexicographic resources. This research is an early step in that direction.

First of all, including the criteria followed by the lexicographers in making decisions about the specialised lexicon in future editions would help overcome this situation. In the front matter analysed, the reference to the inclusion and treatment of diatechnical information is practically non-existent or too general. The decisions of the lexicographers responsible are not justified and seem to be sustained only by the presentation of a list of abbreviations; nor do the dictionaries give reasons for the use or value of domain labels. At the same time, the number of labels selected by the lexicographers of these dictionaries is uneven. There is also an imbalance in the scope of labels, with the DAF and the DLPC presenting many examples of subdomains that the DLE ignores.

The multilingual domain map constructed in this study can contribute to future standardisation efforts adapted to the required interoperability. The normalisation of the domain labelling process and associated encoding tasks is required to achieve structured, organised, accessible and interoperable lexical resources.

CHAPTER 7

A Terminological Approach for Lexicographic Purposes

This leads us to argue that the term, regardless of its aims, must involve a twofold approach – both its linguistic and conceptual dimensions have to be taken into account.
COSTA (2013)

Our research has strictly lexicographic purposes and aims to employ terminological working methods to contribute to the processing of terms in general language dictionaries and the definition of guidelines. The methodology followed for the systematisation of the study assumes the completion of three essential stages: preparation, processing and publishing; it is structured in ten phases to achieve the proposed objectives based on the theoretical assumptions debated before. The double dimension of terminology governs our entire proposal: we will reconcile iteratively, step by step, both the onomasiological and semasiological approaches. We will propose a methodology that combines harmonised and balanced lexicographic and terminological methods and will show how it can help lexicographers when dealing with terms, especially when it comes to writing definitions. As we will see, the explicit identification of the conceptual relations is the key to writing accurate definitions. Furthermore, there is still no lexicographic resource in Portugal that combines specific lexicographic methodologies with terminological assumptions. We will closely follow the planning already proposed by Silva (2014) and now adapted to general language dictionaries. This proposal will directly apply to the new digital edition of the Portuguese academy dictionary (DLP), for which the database of departure was the DLPC. Further, the proposal will be exemplified by analysing terms with the GEOLOGY and FOOTBALL domain labels.

7.1 Terminological Working Methods for Lexicographic Work

As we aim to apply terminological methods to lexicographic work when terms are at the core of the analysis, we will follow the ISO 704 (2009), ‘Terminology work – Principles and methods’. According to this standard, we must consider three distinct

stages of terminology management: (1) the 'planning'; (2) the 'manipulation of terminological information', that is, the processing of terminological data and (3) the 'decision-making' (ISO 704, 2009, p. v). Accordingly, we will take these three stages into account in the presentation of our methodological proposal and combine them with lexicographic methodologies.

A dictionary plan is crucial to shaping the model of the dictionary to be compiled. Establishing a dictionary plan requires observing the following two main aspects: the organisation plan and the dictionary conceptualisation plan. The first relates to management and logistics. In practical terms and concerning the lexicographic process, Wiegand (1998, p. 151) talks about the 'conceptualisation plan of a dictionary' and divides it into five subdivisions: the general preparation phase (structure, content, format, presentation of the final product); the material acquisition phase (corpus); the material preparation phase (preparation of the collected material); the material processing phase (data to include in the dictionary) and the publishing phase (in print dictionaries, proofreading and final adjustments to the manuscript; in digital dictionaries, layout). We will focus essentially on the material preparation and processing phases.

Following terminological methods helps prioritise the concept. Therefore, concerning the presentation of the results of concept analysis, we will use concept diagrams drawn according to ISO 704 (2009) specifications. Furthermore, and following this same standard, we identified the most relevant activities carried out during these pre-determined moments:

- 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)

Observing these activities, we can identify some tasks that have a purely linguistic nature, such as the analysis of terms as designations of concepts, and other tasks that have a conceptual nature, such as the phase of identification of concepts and the modelling of concept systems. In elaborating our methodological proposal, we will combine the following two dimensions: linguistic analysis and conceptual organisation.

Figure 95 presents the different phases that we established above:

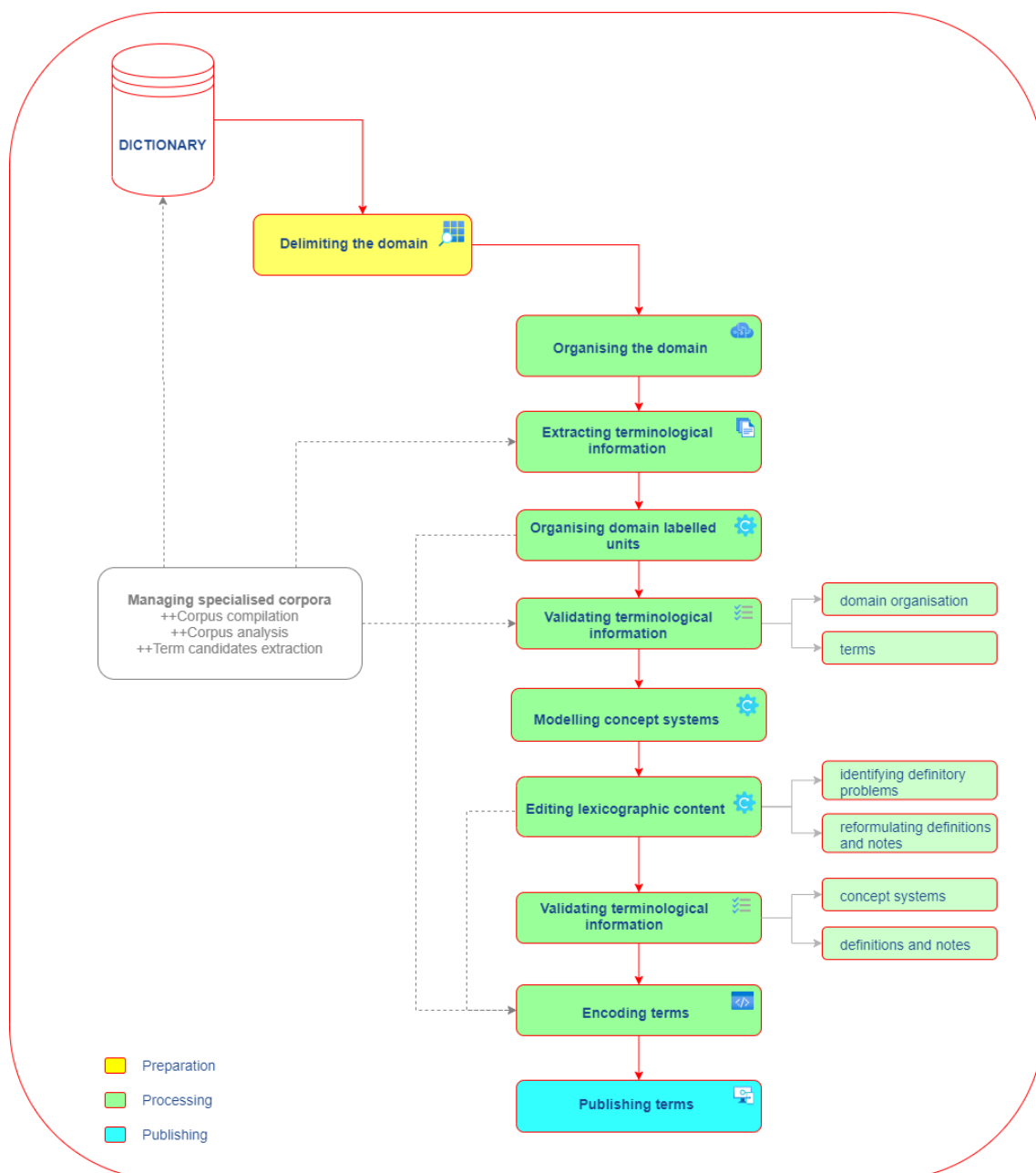


Figure 95: Applying terminological methods when treating terms in general language dictionaries

Figure 95 is based on the reflection made throughout this doctoral research. We highlight in grey a phase that is not addressed here but is essential in current lexicographic work. We refer to the use of corpora since any current dictionary should be based on a reliable corpus. The analysis of specialised corpora is part of the daily lexicographic activity. Computer tools, such as the Sketch Engine⁹¹ software, help lexicographers manage the corpus (compiling, extracting term candidates, annotating, making concordances, queries, etc.) and act as a reference source in extracting usage examples, for instance.

In our research, the selected dictionary – the DLP – will have a double function: it will be both the corpus of analysis and the dictionary that will be improved with our methodological approach. Below, we summarise the ten steps that make up our methodology.

- i) **DELIMITING THE DOMAIN:** The domain should be clearly delimited and cover a specific subject field. Treating all the domains included in a general language dictionary is only feasible with a large team comprising specialists from different areas. In the previous chapter, we saw that the DLPC has 184 domain labels, which would require a solid effort in terms of coordination. Therefore, we recommend to select a domain in advance and work simultaneously on domains directly related to that chosen domain.
- ii) **ORGANISING THE DOMAIN:** Getting to know the domain and subsequently organising it are the two requisite activities for a rapid and systematic identification of the basic concepts, which will result in a better description of the lexicon. In addition to consulting specialised literature, a brief analysis of different existing classification systems (e.g., Dewey Decimal Classification; UNESCO Thesaurus; WordNet Domains Hierarchy) is also recommended. Then, with this acquired knowledge, we suggest proposing the constitution of domain trees keeping in mind the lexicographic purposes. These domain trees should represent *‘una posible organización conceptual*

⁹¹ <https://www.sketchengine.eu/>

de un tema, para fines lexicográficos’ (a possible conceptual organisation of a theme for lexicographic purposes; Guerrero Ramos & Pérez Lagos, 2001, p. 306). Moreover, we recommend the inclusion of this representation in the dictionary, namely in the outside matter, to give the user the possibility of understanding the conceptual scope and the perspective adopted concerning its organisation. Here, we establish a hierarchy: superdomain, domain and subdomain. Also, we argue that this organisation should be shown to the end-user as outside matter.

- iii) **EXTRACTING TERMINOLOGICAL DATA:** In this step, units marked with a domain label must be extracted from the database for a preliminary list analysis. Moreover, the units marked with related domains should be extracted for a joint view of the terminological data. Subsequently, the extracted lists must be analysed, and the lexicographer must organise and structure them (although they can be improved later by the specialists). At this stage, there is a high probability that doubts will arise, such as detecting the lack of a specific unit from a domain as the label was not assigned on unmarked entries or senses (e.g., in Chapter 5, we mentioned the case of the ‘geology’ entries, where no domain label could be found in the three academy dictionaries since the label would be identical to the lemma itself, and, thus, this entry does not appear in the extracted list); querying whether a given term will have a well-assigned domain (e.g., the DLPC indicates that the “cristalografia” [crystallography] entry belongs to the MINERALOGY domain, not the GEOLOGY domain), or even detecting possible candidates for terms in the consulted readings (e.g., the terms “cronostratigráfico” [chronostratigraphic] and “geocronológico” [geochronologic] do not appear in our dictionary). All these cases must be noted for further analysis and future discussion with the specialist.
- iv) **ORGANISING TERMS:** The terminological data extracted can be sorted as alphabetically ordered lists (in the case of the DLP, this is how they are extracted); however, the lexicographer, based on the readings made and the analysis of the lexicographic content, may organise sets of related units for

submission to the specialist. It will be essential to choose some basic concepts and, starting from these, organise all the specialised knowledge – one could say that concepts ‘call for’ one another. Based on the domain tree elaborated for the domain under study, the domain labels should be reviewed, and hierarchical domain labels should be assigned to the terms (lemmas or senses of a particular lemma). The domain hierarchy proposed must be followed. In fact, this task can occur either at this stage or after writing the definitions that correspond to the next stage. Finally, in the last phase, decisions can be made about which domain labels of the hierarchical structure will be visible to the end-user. This decision involves statistical issues and expert proposals.

- v) **VALIDATING TERMINOLOGICAL DATA:** Any validation process⁹² can and should be *‘adaptado às realidades em causa e aos objetivos pretendidos com o ato de validação’* (adapted to the realities in question and the objectives intended with the act of validation; Silva, 2014, p. 159). This step comprises two different activities: validating domain organisation and validating terms. The proposed domain tree must be validated by the team composed of the specialist(s) on the subject field, the terminologist and the lexicographer. Next, the collected terms must be validated/approved by the specialist(s). In this process, the specialist(s) frequently propose additional terms not represented in the extracted list(s) or even call the lexicographers’ attention to poorly assigned domain labels.
- vi) **MODELLING CONCEPT SYSTEMS:** After validating terms, it is necessary to identify the concepts and then model the terminological data collected, establishing relationships among concepts and pointing them to the terms. Once the relationships are correctly identified, lexicographers can start writing the definitions.
- vii) **EDITING LEXICOGRAPHIC CONTENT:** The lexicographic content is edited throughout all the tasks. In this phase, meanings are explained; in other words, the

⁹² In terms of validation processes, for a more detailed description, see Silva (2014), pp. 159–180.

lexicographer proposes a linguistic description of the concept designated by a term. The concept–term equation must be considered. In this sense, a definition establishes a relation between the concept identified and the term in which the *definiendum* is the term. The terminological definition is adapted to general language dictionaries. Existing definitions may have to be reformulated in cases where defining problems are identified. Meanwhile, the lexicographer can propose new definitions based on the previously established concept relations. Additional information can be inserted as notes. As the definitions are drafted, it might be necessary to define other terms whose concepts are connected during the modelling process.

- viii) **VALIDATING TERMINOLOGICAL DATA:** Together with the lexicographer, the expert(s) perform a second task in the validation process. This validation process comprises two activities: validating concept systems and validating the new and reformulated definitions and the notes.
- ix) **ENCODING TERMS:** In the editing process, all the information must be encoded and annotated in an interoperable format that must be defined in the general preparation phase. Generally, lexicographers use computational tools available to support dictionary writing. Another method of dictionary writing uses markup languages, such as XML⁹³, to insert, organise and edit data. This task cuts across the entire process and directly relates to the editing process.
- x) **PUBLISHING TERMS:** In this phase, the validated terms are ready to be made available to the end-user.

Next, we will describe all the above-listed steps by applying the principles and methodology that we follow in the DLP.

7.2. Establishing the Lexicographic Source Corpus (dictionary)

Our base lexicographic corpus is the DLPC from ACL published in 2001, which gave rise to the DLP, an updated version corresponding to the first Portuguese digital academy

⁹³ In DLP, we use the Oxygen XML Editor: <https://www.oxygenxml.com/>

dictionary. Thus, our database includes part of the DLPC material that is being reformulated and will soon be updated on the web.

7.3 Delimiting the Domain

Our starting point was the set of domain labels included in the DLPC. We chose GEOLOGY and FOOTBALL as the domain labels with which to test the proposal for a set of methodological guidelines regarding the lexicographic treatment of terms. This choice was justified in the Introduction section of this work (see pp. 12, 13).

To become familiar with these topics, as we are not specialists in those subject fields, the delimitation of the domains took into account the following procedures: we collected and consulted documentary sources such as textbooks, specialised texts, international glossaries, terminological dictionaries, scientific publications and reference web pages; we also consulted some existing classification systems, as we will discuss further on; we proposed domain trees, which, not intending to be exhaustive, would allow us to identify and establish related subdomains quickly. Meanwhile, the lexicographer/terminologist and professionals in the corresponding areas established constant contact and collaboration.

Regarding GEOLOGY, the participation in a workshop on Sequential Stratigraphy⁹⁴ promoted by the ACL in 2018 enabled further familiarisation with the specialised discourse of stratigraphy. Concerning FOOTBALL, the constant consultation of members of the Portuguese Football Federation⁹⁵ proved to be advantageous.

For our purpose and to restrict the terms under analysis, we asked the specialist to select some terms from the GEOLOGY domain, especially stratigraphical terms. Additionally, we decided to select terms related to the position of football players on the field and some related to supporters in the context of the FOOTBALL domain.

⁹⁴ https://www.facebook.com/events/991915247621994/?active_tab=about

⁹⁵ <https://www.fpf.pt/pt/>

7.3.1 The Geology Domain as a Case Study

Geology is the study of the Earth (from the Greek *geo*, ‘earth’ + *logy* ‘study’). According to the *Glossary of Geology* (Neuendorf, Mehl Jr. & Jackson, 2011), geology is defined as ‘The study of the planet Earth, the materials of which it is made, the processes that act on these materials, the products formed, and the history of the planet and its life forms since its origins’ (p. 267). More precisely, geology is one of the earth sciences that represent ‘*o conjunto das ciências que estudam as fases sólida, líquida e gasosa presentes no planeta Terra*’ [the set of sciences that study the solid, liquid and gaseous phases present on the planet Earth] (Lemos de Sousa, Antunes & Salgado 2015, p. 4).

It is important to note that the terms “earth sciences” or “geosciences” are sometimes used as a synonym for “geology” or “geological sciences”. However, this usage should be avoided, as the concept of earth sciences is much broader than that of geology, which refers to the fields of science dealing with the planet Earth. Probably, this also happens because the term “geology” is older than “earth sciences”. “Earth sciences” are the ‘*ciência que estuda a história do planeta Terra e da vida que nele se desenvolveu: origem, estrutura, composição, evolução, causas e processos que originaram o seu estado atual*’ [science that studies the history of the planet Earth and the life that developed on it: origin, structure, composition, evolution, causes and processes that gave rise to its current state] (ibidem).

The motivation for choosing the domain of geology is derived from the familiarisation with this area within the scope of a collaboration with the Research Unit on Energy, Environment and Health (FP-ENAS) of the University Fernando Pessoa⁹⁶, to create a Glossary of Chronostratigraphic/Geochronologic Units and the ACL work developed around the edition of the *Thesaurus de Ciências da Terra* [Earth Sciences Thesaurus]⁹⁷.

⁹⁶ <http://international.ufp.pt/research/rd-centers/fp-enas/>

⁹⁷ See <https://volp-acl.pt/index.php/publicacoes-do-illp>. The team includes various specialists in Earth sciences, such as Manuel João Lemos de Sousa and Cristina Fernanda Alves Rodrigues, and Ana Salgado as a linguist. The relevance of the work is also justified by the fact that inconsistencies (variants, use of loans, malformed transfers, poorly written definitions in general language dictionaries) have been verified during this research.

The examples related to geology belong to stratigraphic terminology, defined as ‘the total of unit-terms used in stratigraphic classification’⁹⁸. Stratigraphy is the branch of earth sciences that deals with stratified rocks. The OED defines it as ‘the branch of geology concerned with the order and relative position of strata and their relationship to the geological timescale’. Saying ‘the branch of’ immediately conveys the idea of subordination to something. The OED definition allows us to say that stratigraphy is a subordinate concept of geology. At the same time, we prefer to consider it a conceptual branch of earth sciences, as we will discuss.

Stratified rocks are found in the strata, i.e., in the layers of the Earth. They can be rocks of any class, but with a distinctive character and individuality distinguishing them from the rocks of the adjacent layers. The scope of stratigraphy is vast and, through the description of the strata and their relative ages, extends the knowledge of such characteristics and attributes of stratified rocks to their distribution, lithological composition, paleontological content, geochemical and geophysical properties, as well as their genetic interpretation – the how and where they were formed – and geological history.

Working within the above framework, respected authors of the North American school (Krumbein & Sloss, 1963) considered that the study of stratigraphy encompasses the subjects of sedimentary petrology and sedimentology. However, this is not our current understanding. Today, sedimentary petrology and sedimentology are autonomous branches of earth sciences owing to their distinct objectives and, above all, study methods. Currently, stratigraphy is confined, on the one hand, to the study of the geological cycle and sedimentation media and, on the other hand, to space-time relationships in the context of the meanings of LITHOSTRATIGRAPHY, BIOSTRATIGRAPHY and CHRONOSTRATIGRAPHY.

The International Commission on Stratigraphy (ICS), founded in 1961, is the oldest constituent scientific body in the International Union of Geological Sciences (IUGS). Its primary objective is to precisely define global units (systems, series and stages) of the *International Chronostratigraphic Chart*⁹⁹ that, in turn, are the basis for

⁹⁸ <https://stratigraphy.org/guide/defs>

⁹⁹ <https://stratigraphy.org/chart>

the corresponding units (periods, epochs and ages of the *International Geological Time Scale*), thus setting global standards for the fundamental scale for expressing the history of the Earth.

The *International Stratigraphic Guide*¹⁰⁰ was developed 'to promote international agreement on principles of stratigraphic classification and to develop an internationally acceptable stratigraphic terminology and rules of procedure in the interest of improved accuracy and precision in international communication, coordination and understanding'¹⁰¹.

The *International Chronostratigraphic Chart* describes the geological time in which the history of the Earth is inscribed. The different versions are subject to continuous adjustments. For the last English version (v2021/07), see Figure 96.

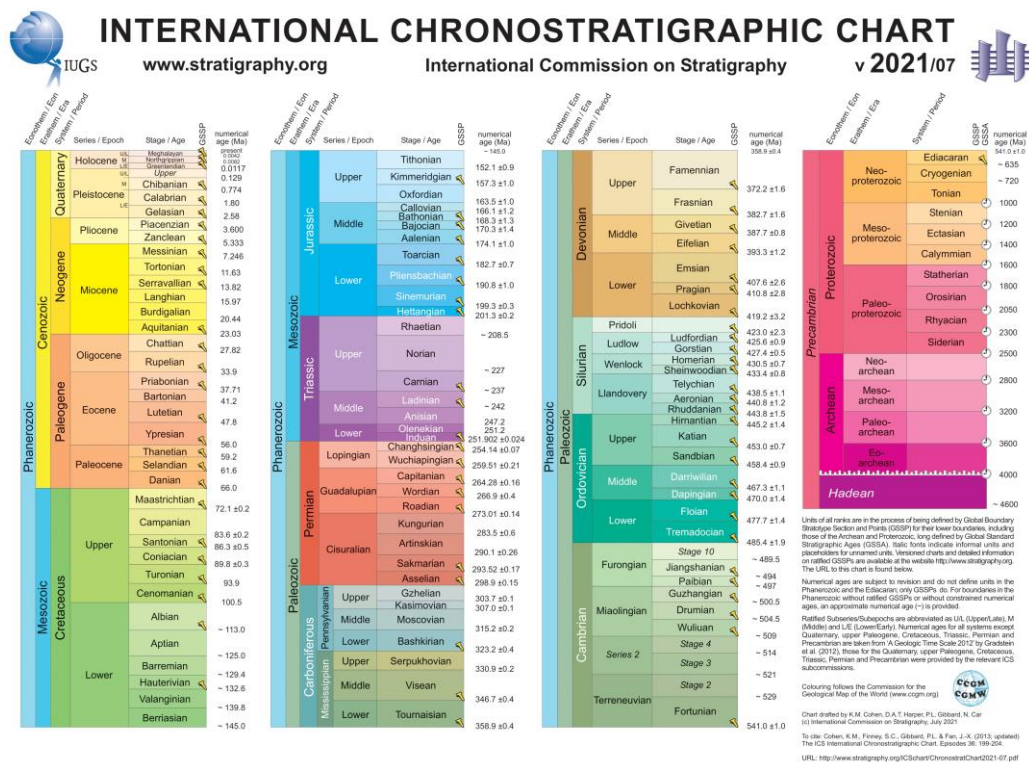


Figure 96: International Chronostratigraphic Chart (Cohen et al., 2021)

The existing Portuguese version (Cohen et al., 2017) dates from 2017 and was made by the Laboratório Nacional de Energia e Geologia (LNEG/IGCP – UNESCO).

¹⁰⁰ The Abridged Version of the International Stratigraphic Guide can be found at: <https://stratigraphy.org/guide/>

¹⁰¹ <https://stratigraphy.org/guide/intro>

This chart combines a numerical absolute time scale that uses as unit a million years (chronometric scale) and a scale in relative time units (chronostratigraphic scale) established by convention. The chronostratigraphic scale is based on the International Standardised System of stratigraphical units (e.g., ‘Jurassic’, ‘Paleocene’). This system, regulated by the ICS UNESCO/United Nations, describes the relative divisions of geological time (eons, eras and their subdivisions), establishes the limits of the units and calibrates them with the chronometric scale, attributing to them the corresponding absolute ages.

The lower boundaries of all units (stages, series, systems and erathems) are currently in the process of being defined by means of sections and points, as *Global Stratotype Section and Boundary Points* (GSSP). The official GSSP are marked in the chart with the Golden Spike symbol, which is also placed on the ground. Finally, the colour code is according to the Commission for the Geological Map of the World (CCGM-IUGS).

The *International Stratigraphic Guide* recommends the following chronostratigraphic terms and geochronologic equivalents to express units of different rank or time scope (Table 10):

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.

Table 10: Conventional hierarchy of the chronostratigraphic/geochronologic units

The chronostratigraphic units are tangible stratigraphic units in the field because they comprise a set of strata consisting of all the rocks, layered or unlayered, formed during a specified interval of geologic time. The units of geologic time during which chronostratigraphic units were formed are called geochronologic units.

The categories within the stratigraphic classification correspond to the rocks of the Earth’s crust. Each category, however, is related to a different property or attribute of the rocks and a different interval of Earth history.

As far as general dictionaries are concerned, geology can be located within classical domains in the lexicographic tradition – as a domain label, it has been present in different dictionaries for centuries. The first point to note is that the terms “cronostratigráfico” [chronostratigraphic] and “geocronológico” [geochronologic] do not appear in any of the dictionaries under analysis. We first consulted the *Guide* to see how the specialists defined these terms. The chronostratigraphic units are understood as ‘bodies of rocks, layered or unlayered, that were formed during a specified interval of geologic time’¹⁰², the geochronologic units as ‘a subdivision of geologic time’¹⁰³. The units of geologic time during which chronostratigraphic units were formed are called geochronologic units.

Geological time is described in two different ways: a quantitative chronology based on absolute ages expressed in millions of years and established by means of radiometric measurements; and using an event chronology based on stratigraphic scales.

Having chosen a highly-specialised domain, as is the case of GEOLOGY, the option for another subject field should guarantee, from the outset, the application of our methodological proposal. For this, we decide to choose a domain that was very distant from pure sciences. We chose the FOOTBALL domain.

7.3.2 The Football Domain as a Case Study

Our interest in football arises from the fact that it has been the most popular sport on the planet since the end of the 19th century, with worldwide expansion via different societies on every continent. It is estimated that 250 million people are directly involved in football and that 1.4 billion people in the world have some interest in football (Morris, 1985). Moreover, its presence as a media event is unquestionable.

We start from Bourdieu, Dauncey and Hare’s (1998) premise: ‘talking about sport scientifically is difficult because it is too easy in one sense: everyone has their own ideas on the subject, and feels able to say something intelligent about it’ (p. 15). Additionally,

¹⁰² <https://stratigraphy.org/guide/chron>

¹⁰³ <https://stratigraphy.org/guide/defs>

(1) there are those who know the world of sport very well in practice but do not know how to talk about it; (2) there are those who, not knowing extensively about the world of sport, can talk about it and dare to do so; and (3) there are others who do so without proper ownership. According to Lipoński (2009, p. 25), ‘the language of sport has been existing since antiquity’. Taborek (2012, p. 237) argues that we cannot speak about the language of sport, as it is only possible to refer to its technical or professional vocabulary ‘inserted’ into the general language.

Football is often referred to as 11-player football because it is played between two teams of 11 players each, as seen in the definitions of the term in the three academy dictionaries (Figure 97): ‘*onze jogadores*’ (DLPC), ‘*onze joueurs*’ (DAF) and ‘*once jugadores*’ (DLE).

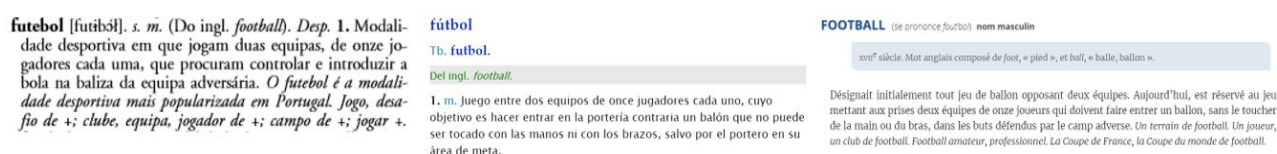


Figure 97: Entries ‘futebol/football/fútbol’ (DLPC, DLE, DAF)

The 11 football players occupy specific positions on the field, which relate to specific terms (Figure 98).



Figure 98: Football players occupy different positions on the field (Salgado & Costa, 2020)

For quick identification of all the possible positions of football players on the field, we created an illustration (Figure 99):



Caption

GUARDA-REDES (goalkeeper)
GR guarda-redes (goalkeeper)

DEFESA (defence)
LD lateral direito (right-back)
LE lateral esquerdo (left-back)
DC defesa central (centre-back)
LB líbero (sweeper)

MEIO-CAMPO (midfield)
MD* médio defensivo (defensive midfielder)
MD médio direito (right midfielder)
ME médio esquerdo (left midfielder)
MC médio-centro (centre midfielder)
MO médio ofensivo (attacking midfielder)

ATAQUE (attack)
AV avançado (attacker)
SA segundo-avanzado (second striker)
PL ponta de lança (striker)
ED extremo-direito (right-winger)
EE extremo-esquerdo (left-winger)

Figure 99: Positions of football players on the field

The positions of the players indicate their specific function on the field; they are typically associated with the tactical scheme used and can be divided into four fundamental positions: (1) goalkeeper (GR); (2) defence (LD, LE, DC, LB); (3) midfielder (MD*, MD, ME, MC, MO); and (4) attack (AV, SA, PL, ED, EE).

7.4 Organising the Domain

This step is part of an extralinguistic level. In the absence of an explanation of their labelling system in the introductory pages of academy dictionaries (cf. Chapter 6), we decided to compare how other existing domain labelling classification systems organise their descriptors¹⁰⁴ to establish analogies.

7.4.1 Comparing Classification Systems

Many library classification systems were developed in the 19th and 20th centuries as an answer to increasing collections of data, and they continue to be used in the systematic physical arrangement of documents in various institutions and the organisation of digital catalogues. These types of classifications improve the traditional alphabetical order of, for example, traditional dictionaries.

Dahlberg developed the notion of knowledge organisation in the 1970s: the German term *Wissensordnung* (knowledge ordering) was used to refer to the conceptual and systematic organisation of human knowledge (Dahlberg, 1974). In English, this term was then translated into ‘knowledge organisation’ and later adopted internationally. Thus, knowledge organisation systems (KOS) are mechanisms for organising information and include classification schemes.

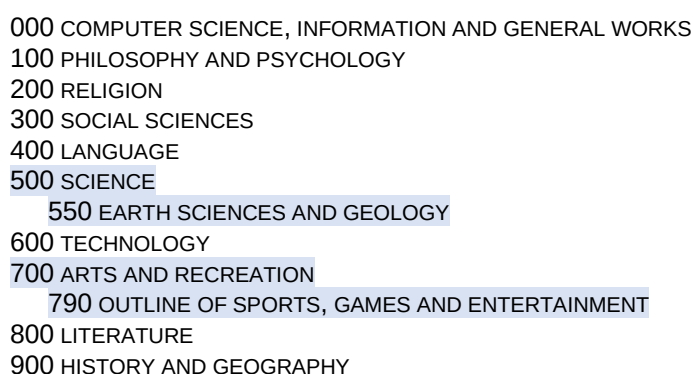
In the present investigation, we considered the following classification systems: the Dewey Decimal Classification (DCC), the Universal Decimal Classification (UDC), the UNESCO Thesaurus, EuroVoc and the WordNet Domains Hierarchy.

The different classification proposals present hierarchical models between domains and subdomains. After looking into the different classification systems, we

¹⁰⁴ A descriptor is a ‘term used to represent a concept when indexing’ (ISO 25964, p. 9).

chose to locate the domains under study to find out their location and respective organisation. We start with EARTH SCIENCES/GEOLOGY and move on to SPORTS/FOOTBALL.

Dewey Decimal Classification (DDC). The Dewey Decimal Classification (DDC)¹⁰⁵ was conceived by Melvil Dewey (1851–1931) in 1873 and first published in 1876. The DDC is published by the OCLC Online Computer Library Center, Inc. The DDC is a closed hierarchical system for library organisation purposes based on the division of fields of study into ten classes with decimal extensions. The classification structure is hierarchical and the annotation follows the same hierarchy. The ten main classes are (Figure 100):



```
graph TD; 000[000 COMPUTER SCIENCE, INFORMATION AND GENERAL WORKS] --> 100[100 PHILOSOPHY AND PSYCHOLOGY]; 000 --> 200[200 RELIGION]; 000 --> 300[300 SOCIAL SCIENCES]; 000 --> 400[400 LANGUAGE]; 000 --> 500[500 SCIENCE]; 000 --> 600[600 TECHNOLOGY]; 000 --> 700[700 ARTS AND RECREATION]; 000 --> 800[800 LITERATURE]; 000 --> 900[900 HISTORY AND GEOGRAPHY]; 500 --> 550[550 EARTH SCIENCES AND GEOLOGY]; 700 --> 790[790 OUTLINE OF SPORTS, GAMES AND ENTERTAINMENT];
```

000 COMPUTER SCIENCE, INFORMATION AND GENERAL WORKS
100 PHILOSOPHY AND PSYCHOLOGY
200 RELIGION
300 SOCIAL SCIENCES
400 LANGUAGE
500 SCIENCE
 550 EARTH SCIENCES AND GEOLOGY
600 TECHNOLOGY
700 ARTS AND RECREATION
 790 OUTLINE OF SPORTS, GAMES AND ENTERTAINMENT
800 LITERATURE
900 HISTORY AND GEOGRAPHY

Figure 100: *Dewey Decimal Classification System*

Each class is separated into ten divisions numbered 0–9. Class 000 is the broader class and is used for works that are not limited to a particular discipline. EARTH SCIENCES is included in class 500, which is devoted to the broader class of SCIENCE. Specifically, EARTH SCIENCES is found in class 550, and is a kind of a catchall for all the sciences that explore the Earth, such as GEOLOGY, located in class 551 (GEOLOGY, HYDROLOGY, METEOROLOGY) or PETROLOGY in class 552. Class 700 covers ARTS AND RECREATION, which includes SPORTS in class 790.

Universal Decimal Classification (UDC). One of the most widely used classification schemes, based on the Dewey system but extended, is the Universal

¹⁰⁵ <https://www.oclc.org/en/dewey.html>

Decimal Classification (UDC)¹⁰⁶. The UDC scheme is a bibliographic and library classification created by Paul Otlet (1868–1944) and Henri La Fontaine (1853–1943) that intended to develop a universal bibliography, *Manuel du Répertoire de Bibliographie Universelle*, also called *Classification de Bruxelles*, to carry out the bibliographic control of all bibliographies that were known and registered to date. Eugen Wüster, for instance, used the UDC system to plan the domains and subdomains on which the definition of terms depended in his systematic dictionary entitled *The Machine Tool* (Wüster, 1968). In Figure 101, we show the main classes:

```

0 SCIENCE AND KNOWLEDGE. ORGANISATION. COMPUTER SCIENCE.
  INFORMATION SCIENCE. DOCUMENTATION. LIBRARIANSHIP.
  INSTITUTIONS. PUBLICATIONS
1 PHILOSOPHY. PSYCHOLOGY
2 RELIGION. THEOLOGY
3 SOCIAL SCIENCES
4 VACANT
5 MATHEMATICS. NATURAL SCIENCES
  55 EARTH SCIENCES. GEOLOGICAL SCIENCE
    550 ANCILLARY SCIENCES OF GEOLOGY
      551 GENERAL GEOLOGY. METEOROLOGY. CLIMATOLOGY.
        HISTORICAL GEOLOGY. STRATIGRAPHY. PALEOGEOGRAPHY
      552 PETROLOGY. PETROGRAPHY
      553 ECONOMIC GEOLOGY. MINERAL DEPOSITS
      556 HYDROSPHERE. WATER IN GENERAL. HYDROLOGY
    56 PALAEONTOLOGY
6 APPLIED SCIENCES. MEDICINE. TECHNOLOGY
7 THE ARTS. ENTERTAINMENT. SPORT
  796 SPORT. GAMES. PHYSICAL EXERCISES
    796.3 BALL GAMES
8 LANGUAGE. LINGUISTICS. LITERATURE
9 GEOGRAPHY. BIOGRAPHY. HISTORY

```

Figure 101: *Universal Decimal Classification System*

GEOLOGY is found in class 5 dedicated to MATHEMATICS and NATURAL SCIENCES – more precisely in class 55, EARTH SCIENCES. GEOLOGICAL SCIENCES is a kind of catchall for other related domains, such as the fields that we found in subclasses 550, 551, 552, 553 and 556. What caught our attention in this classification was the fact that PALAEONTOLOGY is independent of other geological domains. Football belongs to SPORT, class 7, more

¹⁰⁶ <http://www.udcsummary.info/php/index.php>

precisely to subclass 796.3, ball games: 'Ball games in which the ball is played with foot and hand / Including: Football (soccer, rugby etc.)'.¹⁰⁷

In addition to these classification systems, other resources can facilitate the organisation of domains.

UNESCO Thesaurus. The UNESCO Thesaurus¹⁰⁸ is a controlled vocabulary developed by the United Nations Educational, Scientific and Cultural Organisation that includes subject terms for the following areas of knowledge: education, culture, natural sciences, social and human sciences, communication and information. The UNESCO Thesaurus is mainly used for indexing and searching resources in UNESCO's document repository. The first edition of the Thesaurus was released in English in 1977, with French and Spanish translations in 1983 and 1984. The second revised and restructured version was released in 1995. Today, the Thesaurus is available in English, French, Spanish, Russian (since 2005) and Arabic (since 2020).

The UNESCO Thesaurus was the first vocabulary to be published in Simple Knowledge Organisation System (SKOS) format. Concepts are grouped into seven broad subject areas, which are broken down into thesauri. The UNESCO Thesaurus complies with the ISO 25964-1 (2011) standard¹⁰⁹. We found seven major subject fields (Figure 102):

- 1 EDUCATION
- 2 SCIENCE
 - 2.35 EARTH SCIENCES
- 3 CULTURE
 - 3.65 LEISURE
- 4 SOCIAL AND HUMAN SCIENCES
- 5 INFORMATION AND COMMUNICATION
- 6 POLITICS, LAW AND ECONOMICS
- 7 COUNTRIES AND COUNTRY GROUPINGS

Figure 102: UNESCO Thesaurus Classification System

¹⁰⁷ <https://udcsummary.info/php/index.php?id=64676&lang=en>

¹⁰⁸ <http://vocabularies.unesco.org/browser/thesaurus/fr/>

¹⁰⁹ <https://www.iso.org/standard/53657.html>

Within class 2, we found EARTH SCIENCES (2.35). Here, we found 61 descriptors such as GEOPHYSICS, MINERALOGY, PALAEONTOLOGY and many others. Among these descriptors, there is also earth sciences, which was perhaps not to be expected since it is the designation that superordinates the rest. Searching for sports, we found it in class 3 (CULTURE), in subclass 3.65 (LEISURE).

EuroVoc. EuroVoc¹¹⁰ is the most useful controlled vocabulary for optimising access to the subject matter in EU and national legal data. It has also been compiled following the requirements of the ISO 25964-1 (2011) standard. EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU. It contains terms in 23 European languages. This resource is managed by the Publications Office, which moved forward to ontology-based thesaurus management and semantic web technologies conformant to W3C recommendations as well as the latest trends in thesaurus standards. This thesaurus was also applied in the establishment of the Inter-Active Terminology for Europe (IATE).

EuroVoc is divided into 21 domains, composed by 127 subdomains and more than 6,700 detailed descriptors. The domains include INTERNATIONAL ORGANISATION; GEOGRAPHY; INDUSTRY; ENERGY; PRODUCTION, TECHNOLOGY AND RESEARCH; AGRI-FOODSTUFFS; AGRICULTURE, FORESTRY AND FISHERIES; ENVIRONMENT; TRANSPORT; EMPLOYMENT AND WORKING CONDITIONS; BUSINESS AND COMPETITION; SCIENCE; EDUCATION AND COMMUNICATIONS; SOCIAL QUESTIONS; FINANCE; TRADE; ECONOMICS; LAW; EUROPEAN UNION; INTERNATIONAL RELATIONS; POLITICS. The depth of content differs among the 21 domains, with the domains aligned with the interests of the European Union having more elaborate content than other domains.

We highlight some domains registered in EuroVoc because they are important today and they are not registered in dictionaries, such as ENVIRONMENT, TECHNOLOGY, ENERGY, INDUSTRY and EDUCATION. None of these domains is present in the language dictionaries under study. Why is a domain like ENVIRONMENT missing? Is it because of the intersection of the terminology of that area with other domains, such as ECOLOGY or BIOLOGY? TECHNOLOGY represents another such case; we proceed with the hypothesis that

¹¹⁰ <https://eur-lex.europa.eu/browse/eurovoc.html>

this domain is not included in the dictionaries studied because its terminological units are considered already common in the general lexicon since it is ordinary nowadays for a Portuguese, French or Spanish speaker to integrate words such as ‘GPS’ or ‘wi-fi’ into their daily discourse.

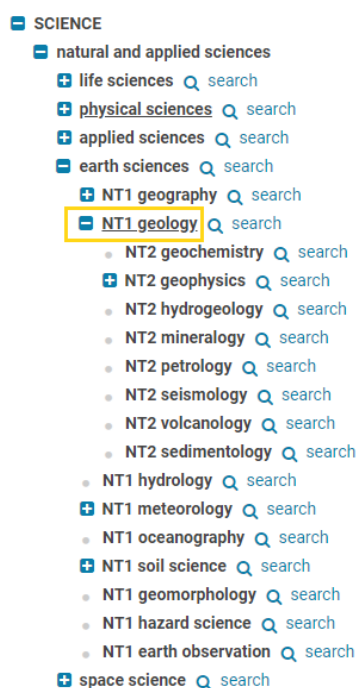


Figure 103: EuroVoc Classification System

Under the descriptor SCIENCE, we found EARTH SCIENCES subdivided into GEOGRAPHY, GEOLOGY, HYDROLOGY and METEOROLOGY. We did not find SPORTS. The Publications Office uses VocBench¹¹¹ for the maintenance of EuroVoc. VocBench is a web-based open-source collaborative platform for managing multilingual controlled vocabularies that uses semantic technologies and complies with the SKOS and SKOS-XL standards. It is particularly suitable for managing large thesauri in RDF format.

WordNet Domains Hierarchy. Wordnet Domains Hierarchy¹¹² (WDH) is a language-independent resource composed of 200 domain labels organised in a

¹¹¹ <http://vocbench.uniroma2.it/>

¹¹² <https://wndomains.fbk.eu/labels.html>

hierarchical structure. Issues concerning the semantics, completeness, balancing among each domain coverage and the granularity of domain distinctions have been addressed regarding the Dewey Decimal Classification (Figure 104).



Figure 104: WordNet Domains Hierarchy

We found GEOLOGY in PURE_SCIENCE/EARTH while FOOTBALL belongs to the class called FREE_TIME under SPORT.

7.4.2 Hierarchising domain labels

As Atkins and Rundell (2008) argue, instead of conceiving a ‘totally *flat* (non-hierarchical list of domains) [...] it is more practicable to try to build a domain list with a certain hierarchical structure, so that instead of “physics”, “chemistry”, etc., you have

“science: physics”, “science: chemistry”, and so on’ (p. 184). We agree with this argument and find it advantageous to apply a previously organised structure in both the composition and the editing phases of a lexicographic resource since it facilitates the lexicographer’s control over terminology. Thus, it will be possible to ensure that no ‘glaring omissions’ are present and to ‘mark vocabulary items more accurately’ (*ibidem*).

We accord with Dubois (1990) that highly specialised disciplines will need identification of ‘*grands domaines*’ (pp. 1583–1584) or superordinate domains. Although we understand Dubois’ reason for referring to only highly specialised domains, we see advantages in selecting superdomains, even when it is explicitly not about specialised knowledge. The terms that are common to multiple domains will receive the ‘top-level’ domain marker (Atkins & Rundell, pp. 184, 185). We have adopted the term superdomain¹¹³.

In the organisation of domains (Figure 105), we consider the existence of three possible levels: superdomain, domain and subdomain.

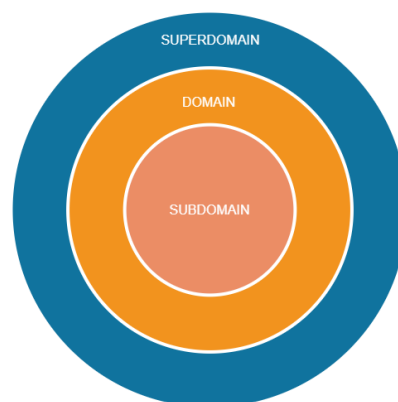


Figure 105: Domain hierarchy

The superdomain corresponds to the broadest taxonomic grouping followed by a domain, whereas the subdomain is part of a broader domain.

For knowledge and lexicographic content organisation, we believe it will be helpful to establish a hierarchical structure in general language dictionaries for two main

¹¹³ Costa (1993), for example, used the term ‘*macrodomínio*’ (macrodomain).

reasons: 1) to organise an increasing amount of terminological information included in lexicographic resources and 2) to provide the lexicographers greater control over specialised content in order to be able to detect inconsistencies and control their work more efficiently, even if they are invisible to the user, justifying our recommendation for a better organisation of the set of terms in general language dictionaries. As Silva (2014) states, *‘quanto melhor estiver organizado um sistema conceptual, mais fácil se torna, também, a gestão da terminologia’* (the better a concept system is organised, the easier it is to manage terminology also; p. 135), both at the level of decision-making on the inclusion/exclusion of concepts and concerning the drafting of definitions.

The methodology adopted involves the validation of the above-mentioned superdomains (EARTH SCIENCES and SPORTS) and the identification of the domains and the various related subdomains. The lexicographer – who is generally not an expert in the fields in which they have to work – can draft a domain tree or a conceptual scheme that will be subsequently validated by the specialist and whose representation will aim at structuring knowledge for the scope of dictionaries; this means that the representation may not fully correspond to the conception the specialists might have from their area of intervention.

Since we have found four domain labels related to the concept of earth sciences in the DLPC and DAF (CRYSTALLOGRAPHY, GEOLOGY, MINERALOGY and PALAEONTOLOGY) and one (GEOLOGY) in the DLE, the GEOLOGY domain was subjected to a detailed analysis. The next step was to collect those domain labels from the academy dictionaries under study that were possibly associated with the superdomain of EARTH SCIENCES. We used metalabels (see Chapter 6), that is, the English equivalent of the different domains. Subsequently, we compared the location of these labels in the classification systems consulted. The results are presented in Table 11:

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/Crystallography
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

Table 11: Comparison of academy dictionaries domain labels and classification systems (Salgado, Costa, & Tasovac, 2021)

The first point to highlight is the similarity of the labels between the Portuguese and French dictionaries: there are four identical labels. At the same time, Spanish has only one, a generic label – a subject already discussed in Chapter 6.

The second point is that while observing this table, we found that domains that could all be included in GEOLOGY in a general language dictionary (as in the DLE) appear, after all, to be associated with other specialised areas in the classification systems. Taking the DDC as an example, EARTH SCIENCES appears in class 550. Thus, it is a kind of catchall for all the sciences that explore the Earth. Further, we found GEOLOGY in class 551, followed by HYDROLOGY and METEOROLOGY. However, the domains of CRYSTALLOGRAPHY and MINERALOGY are indexed to class 540, which covers the area of CHEMISTRY and other divisions related to the mineralogical sciences represented in class 548 (CRYSTALLOGRAPHY) and class 549 (MINERALOGY). In turn, PALAEONTOLOGY figures in an independent class, 560, and is associated with PALEOZOLOGY. The UDC follows this same line. Concerning EuroVoc, the editors have preferred to include EARTH SCIENCES in SCIENCE/NATURAL AND APPLIED SCIENCES. Another proposal, which we are tempted to approach, is that of UNESCO's Thesaurus. GEOLOGY, in this case, is in class 2 SCIENCE/2.35 EARTH SCIENCES, where MINERALOGY and PALAEONTOLOGY are also included. We agree with this approach, except for the insertion of CRYSTALLOGRAPHY in PHYSICAL SCIENCES (2.20).

All these classification proposals are valid and reveal the complexity of the topic. The fact that, for example, MINERALOGY is associated with CHEMISTRY, not with GEOLOGY, is acceptable since much of the subject actually falls into the CHEMISTRY domain; however, it cannot be neglected that the subject is also directly related to GEOLOGY. Thus, the notion of interdisciplinarity is a central point in several sciences; in terms of domain organisation, we must always bear in mind the possible multidisciplinary nature of many domains. As we will see, highly specialised domains share their knowledge with other branches of knowledge (e.g., geology intersects with other areas, such as CHEMISTRY, GEOGRAPHY and BIOLOGY). The complexity of a generic domain, such as EARTH SCIENCES, with its frequent interdisciplinarity with other domains, makes GEOLOGY's delimitation as a domain for analysis even more important.

Another point that we must pay attention to is the nature of the lexicographic works in question (general language dictionaries, not terminological dictionaries) and the target audience to whom they are addressed. In principle, a greater degree of specialisation of a domain requires more effort of interpretation from the end-user. Conversely, we can say that the lexicographer must understand a specific concept very well and know how to establish the relationships among concepts. When defining that specific term, it must be comprehensible for the end user. The organisation and subsequent segmentation of a domain as vast as that of EARTH SCIENCES, in general, or GEOLOGY, in particular, thus carries advantages for both the lexicographer and the end-user.

After comparing the different classification systems, we now present our proposal to represent domains associated with EARTH SCIENCES in general language dictionaries (Figure 106), which the specialist validated. Since the present scheme has been drawn up for the specific purpose of this thesis (domain labelling in dictionaries), it must be analysed and understood while taking that purpose into account. We also use some anchors that play the role of lexical markers.

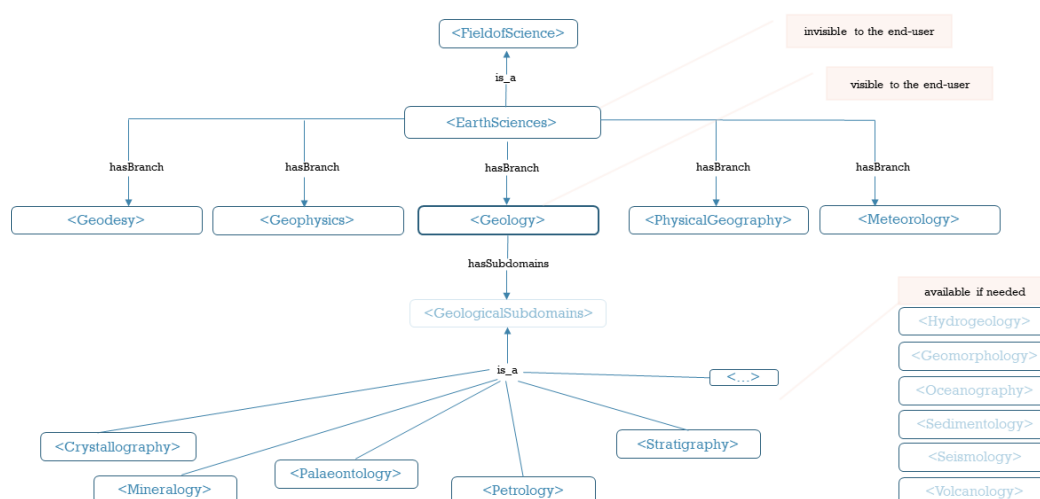


Figure 106: Domain labels within the EARTH SCIENCES superdomain showing GEOLOGY as domain and identifying its subdomains

In our proposal (Figure 106), EARTH SCIENCES represents a broad subject area (superdomain) that can be broken down into narrower subject branches (GEOLOGY, GEODESY, GEOPHYSICS, PHYSICAL GEOGRAPHY, METEOROLOGY). In turn, the narrower subject branch of GEOLOGY has various subdomains (CRYSTALLOGRAPHY, MINERALOGY, PALAEONTOLOGY, PETROLOGY, STRATIGRAPHY).

Even though GEOLOGY as a domain label in general language dictionaries is part of a certain lexicographic tradition, we argue that EARTH SCIENCES should be placed at the top level.

Concerning the visibility of lexicographic content, not all information from a lexicographic resource needs to be visible in the final product. Some mechanisms allow the insertion of tags whose visibility will be null for the end-user. At the moment of writing this thesis, we have an invisible¹¹⁴ superdomain (metalabel: EARTHSCIENCES), a visible domain (metalabel: GEOLOGY) and five potentially visible subdomains (metalabels: CRYSTALLOGRAPHY, MINERALOGY, PALAEONTOLOGY, PETROLOGY, STRATIGRAPHY). Concerning geological subdomains, the information is invisible to the end-user, as we will explain in

¹¹⁴ Here, the domain visibility is mentioned in the context of the end user.

Chapter 9. This point will have to be further discussed before making the dictionary available online; it involves other issues that are not directly related to the topic of this thesis and that will have to be approved by the Dictionary Committee¹¹⁵ and geology collaborators.

However, if there is a need to include other subdomains, they are already foreseen (see Figure 106, ‘available if needed’): HYDROGEOLOGY, GEOMORPHOLOGY, OCEANOGRAPHY, SEDIMENTOLOGY, SEISMOLOGY, VOLCANOLOGY. These labels are thus available to lexicographers, and their use can be re-evaluated if discussed again with the specialist. This is a point that we consider advantageous since it avoids the multiplication of labels and different designations.

Finally, it is self-evident that concerning GEOLOGY, only the elaboration of concept systems will allow us to have a more concrete notion of the subdomains that should be conveyed and identify the many various concepts shared among the multiple subdomains. Although there is no consensus, the analysis of classification systems has allowed us to validate our starting hypothesis of including GEOLOGY in the EARTH SCIENCES superdomain in a generalised way. The same happens with FOOTBALL, which is, as explained below, integrated into the superdomain of SPORT.



Concerning FOOTBALL, we find advantages to using SPORTS as a superdomain so that all sports are linked in general language dictionaries. However, we did not find any advantage to establishing a higher level, as, for example, the UDC does with LEISURE. We insinuate that the sport classification and the possibility of contextualising a given note after the definition in the football context are sufficient in language dictionaries.

The domain label FOOTBALL is recent in general language dictionaries¹¹⁶, and the SPORTS label has often been used to identify football terms. The question now arises as to whether there is any advantage in adopting the domain label FOOTBALL (abbreviated

¹¹⁵ These decisions have been discussed with the ACL Dictionary Committee and may be amended. The relevance of assigning a given domain label can be evaluated considering the quantitative data, that is, the number of entries that can be classified in these subdomains.

¹¹⁶ Rull (2008) states, for instance, that the label *Dep.* was introduced in the 1970 edition of DLE.

form: 'Fut.') in general language dictionaries. The reference to this sport is often identified in the definition through thematic indications. For instance, Nomdedeu Rull (2001) proposes that the label *Fút.* should be applied to terms used only in football (e.g., 'hooligan', 'líbero', 'volante') and the label *Dep. esp. Fút.* (sports especially in football) to terms used in sports in general (e.g., 'club', 'equipo', 'fútbol', 'medio'), especially in football. Instead, we will endorse the use of FOOTBALL as a subdomain of the SPORTS superdomain and the TEAM SPORTS domain and that the referred indication proposed by Nomdedeu Rull (2001) be provided in the note field, as we will demonstrate going forward. We also argue that assigning the FOOTBALL label will only make sense if labels are created for all sports terms, with a frequency of occurrence in general dictionaries attested to be high. We are far from considering having a label for all sports. Take, for example, the list of Olympic sports (Figure 107).

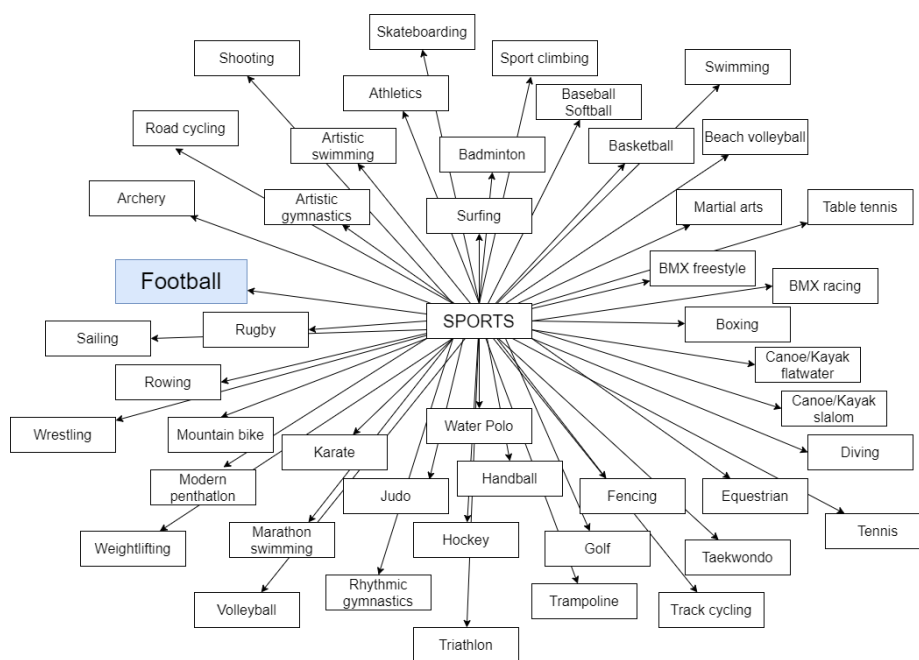


Figure 107: Olympic sports

Right from the start, our lexicographic experience makes us reject the prospect of presenting a domain such as BMX FREESTYLE, BMX RACING, MOUNTAIN BIKING, ROAD CYCLING, TRACK CYCLING. However, we found some advantages in being able to hypothesise the inclusion of a domain like CYCLING that integrates the associated modalities. The same

can be said about water sports. Establishing WATER SPORTS as a domain, we can include sports-related terms such as CANOEING, DIVING, ROWING, SAILING, SURFING, SWIMMING and WATER POLO. For example, looking at sports such as JUDO or KARATE, a MARTIAL ARTS label seems fit. We can see a lot of work to be done in the future concerning the organisation of other sports labels. Each case is different, and each sport should be analysed in the future.

Regarding granularity, we believe that its level does not need to be very detailed for general language dictionaries; a finer granularity may allow for a more significant number of combinations.

Considering the above-mentioned points, we decided to integrate FOOTBALL into TEAM SPORTS whose related terms can act as subordinates to the SPORTS label.

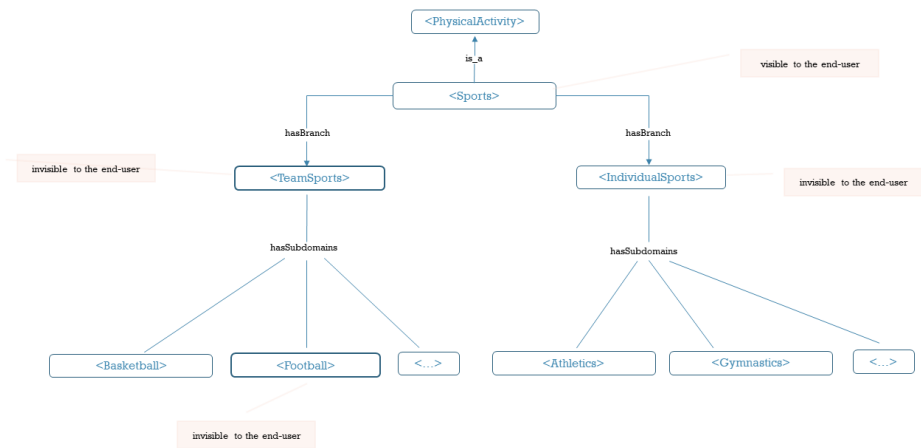


Figure 108: Domain labels within the SPORTS superdomain showing TEAM SPORTS, INDIVIDUAL SPORTS as domains and FOOTBALL as a subdomain

In Figure 108, we present a possible structure of the SPORTS superdomain. However, we recognise that much work must be done to establish the related domains better. For this work, we aimed to include only FOOTBALL in the hierarchy and determined that the scope of our work is about football; now we have only created the labels TEAM SPORTS and INDIVIDUAL SPORTS. Here, TEAM SPORTS includes all the sports that involve competition between teams of players, such as BASEBALL, BASKETBALL, CRICKET, FOOTBALL, HANDBALL, HOCKEY, RUGBY, VOLLEYBALL and WATER POLO. Thus, the following is the hierarchy:

SPORTS is a superdomain, TEAM SPORTS is a domain and FOOTBALL is a subdomain. The INDIVIDUAL SPORTS includes sports in which, generally¹¹⁷, participants compete as individuals. In this way, we can subordinate other modalities to this category, such as MARTIAL ARTS, ATHLETICS, CYCLING, HORSEBACK RIDING, FENCING, GYMNASTICS, GOLF, RACING, SWIMMING, SQUASH, TENNIS, certain WATER SPORTS, COMBAT SPORTS and WINTER SPORTS.

Similar to geological subdomains, the information about sports subdomains will be invisible to the end-user in the case of the FOOTBALL label, showing only the superdomain. As explained further, this specific information will be provided in the definition or as a note. Nonetheless, we are aware that when the content of the dictionary is completely revised and updated, some options now taken may be subject to further debate; for example, and as we have been insisting, the number of occurrences of terms in a given domain can justify the use of a given label.

The structuring of the domains that we have just completed has more to do with the organisation and structuring of knowledge and lexical data of a specialised nature. This organisation will allow for advanced research in the future.

The annotation of the superdomain, the domain and the subdomains will be made using TEI (Chapter 9) in the new edition of the ACL dictionary (DLP), and their visibility to the public, or not, will be discussed when the dictionary is ready to be made available.

Since the domain (labels) under study have been organised, we will now describe our methodological steps for the treatment of terms.

7.5 Extracting Terminological Data

In this level, we followed a semasiological approach, i.e., we analysed terms as a verbal designation of a concept. We collected all the terms tagged with the domains under study in the DLP and randomly selected some of those terms.

¹¹⁷ We used the adverb 'generally' as a safeguard, since among the sports mentioned, sometimes there may be team competitions. For example, in horse riding, the competitions can be among individuals, pairs or teams, or in canoeing, one can participate individually or as a member of a club.

7.6 Organising Terms

The labels covered should be organised in a hierarchy and not just listed. It is important, both for lexicographers and end-users, to see the relations among them. The attribution of domain labels must consider the previously established organisation of the domains and the lists of terms per domain that must be extracted for later presentation and request of validation from the specialist. During this phase, the lexicographer should fit the domain in question into the established hierarchy of labels. When presenting the documentation to the specialist, they will only have to validate the lexicographer's proposal. There will possibly be terms that share domains.

7.7 Validating Terminological Data

Throughout the entire process, contact with specialists is key to validating information. A data validation process must ensure the quality of lexicographic data (cf. Silva, 2014). Whenever possible, validation should consider both linguistic and conceptual components. Within the scope of the work developed at the DLP, meetings are scheduled to clarify doubts. The meetings are always prepared by the responsible lexicographer who selects and organises the data that must be subject to validation.

7.7.1 Domain organisation

The draft of an initial domain tree must be shown to the specialist(s) of the subject field and discussed with them. In this case, we always make specialists aware of the fact that it is an organisation with application in general language dictionaries, and not exactly an organisation of in-depth specialised knowledge.

7.7.2 Terms

This is the phase in which the extracted and analysed terms must be validated by the specialist(s). For this purpose, we created a validation grid in Excel with the following structure: Entry, Source, Domain, Yes, No, I don't know and Notes. The column Entry contains the units extracted from the dictionary in alphabetical order. The Source is

only important in the case of polylexical units, as it informs the specialist in which entry that unit is registered – we use the ID of the entry. Concerning the Domain, this cell contains the diatechnical information included in the dictionary. We explain to the specialist that the information must also be validated. If there is some inconsistency or even errors, we ask them to leave a note. With regard to the Yes, No and I don't know columns, the specialist is expected to express their opinion regarding the inclusion of the terms in question. The Notes column is provided in case the specialist needs to make a comment that they consider relevant for the indicated answer. The specialist(s) also often detect(s) the need to propose more terms related to the listing first presented or poorly assigned domain labels.

ENTRY	SOURCE	DOMAIN	YES	NO	I DON'T KNOW	NOTES
andar	xml:id="DLP-andar"					
éon	xml:id="DLP-eon"	Geol.				
eonotema	xml:id="DLP-eonotema"	Geol.				
época	xml:id="DLP-epoca"	Geol.				
era	xml:id="DLP-era"	Geol.				
era primária	xml:id="DLP-era"	Geol.			X	<i>Preciso de ver a definição incluída no dicionário.</i>
era quaternária	xml:id="DLP-era"	Geol.			X	<i>Preciso de ver a definição incluída no dicionário.</i>
era secundária	xml:id="DLP-era"	Geol.			X	<i>Preciso de ver a definição incluída no dicionário.</i>
era terciária	xml:id="DLP-era"	Geol.			X	<i>Preciso de ver a definição incluída no dicionário.</i>
eratema						<i>Incluir termo.</i>
idade						<i>Incluir sentido geológico.</i>
período	xml:id="DLP-periodo"					<i>Incluir sentido geológico.</i>
série						<i>Incluir sentido geológico.</i>
sistema	xml:id="DLP-sistema"	Geol.				

Proposta de etiqueta de domínio: Ciências da Terra/Geologia/Estratigrafia						Podemos continuar a usar o domínio Geologia, mas recomendo integrar todos estes termos no subdomínio Estratigrafia.
---	--	--	--	--	--	---

Figure 109: Validation grid template (DLP)

In Figure 109, we show an example of the validation grid template. In this case, we previously discussed with the specialist the need to systematically include all chronostratigraphic and geochronological units in the dictionary. We showed them the dictionary content concerning these types of units and asked them to propose a domain classification for these terms.

7.8 Modelling Concept Systems

A concept system is intended to represent the knowledge of a domain by using a set of structured concepts and the respective relationships established between them. To build the concept systems, our references were the concept relations and the graphic representations in the UML (Unified Modelling Language) notation proposed by the ISO 704 (2009) standard through concept diagrams¹¹⁸.

After understanding the fundamental notions of the subject fields, we extracted and collected an unstructured set of concepts from the DLPC, and updated these specialised meanings to DLP. We chose the Portuguese examples for illustrative purposes and scrutinised how the ISO 704 (2009) standard treats concept relation types. Based on these examples, we form concept systems, subject to some adjustments after submission to the specialist. Concept systems are classified according to the types of relations among the concepts. We identify hierarchical relations – generic and partitive – and associative relations to model the concept systems.

¹¹⁸ A concept diagram is a ‘graphic representation of a concept system’ (ISO 1087:2019, p. 7).

We start with hierarchical relations, where we have superordinate and subordinate concepts in relation to each other in a nested hierarchy. As mentioned above, there are two types:

a) **Generic relations:** '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). The superordinate concept is called the generic concept, and the subordinate concept is the specific concept. In other words, the generic concept is a *parent* that imposes its characteristics on a *child* or the specific concept, and possible coordinate concepts are *siblings* – following the principle of inheritance (ibidem, p. 9). The first two elements are usually referred to as the *genus* and the *differentia*. In this type of relation, the subordinate concept must be a kind of tying concept.

Below, we represent a generic concept relation using the concept of <GeochronologicUnit> and employing a tree diagram as established in ISO 704 (2009).

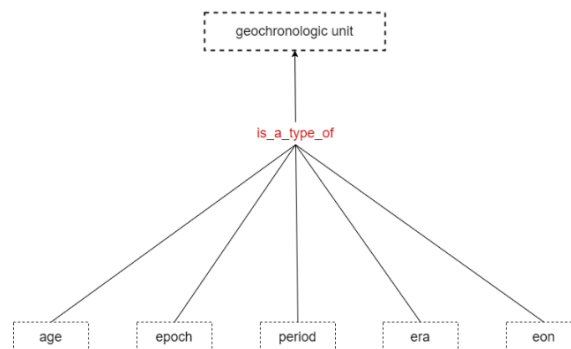


Figure 110: Representation of a generic relation using the concept of <GeochronologicUnit>

In this concept diagram (Figure 110), <GeochronologicUnit> is the generic or superordinate concept and <Age>, <Epoch>, <Period>, <Era> and <Eon> are the specific or subordinate concepts. The generic relation can be expressed by the formulae

- *X is a type of A.*

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

In other words: *<Age>, <Epoch> and <Period> are types of Geochronologic Units.*

In these types of relations, the specific concepts inherit a set of characteristics from their generic superordinate concept, i.e., the superordinate concept includes the subordinate concepts. The extension of the subordinate concept is smaller than that of the superordinate concept. The type of conceptual relation was made explicit using the marker *is_a_type_of*, which structures the generic/specific type relation.

Regarding the semasiological approach, this marker also gives us the possibility of detecting semantic relations¹¹⁹ such as hypernym-hyponym relations, where “idade”, “época”, “período”, “era” and “éon” (specific terms) are the hyponym of the hypernym “unidade geocronológica” (generic term). Hyperonymy establishes a one-way implication relationship between two terms: *If X is an eon, then X is a geochronological unit.* Nevertheless, we cannot reverse the equation and say: *If X is a geochronological unit, then X is an eon.* Thus, a hyponym X is a type of hyperonym X.

Two points require our attention:

(1) The fact that the term “unidade geocronológica”, which is associated with the superordinate concept *<GeochronologicUnit>*, is not defined in the DLPC – it does not appear in the dictionary entry “unidade”, and “cronostratigráfico” does not even appear as a headword.

(2) The subordination established between different concepts is not mirrored in the DLPC. These subordinate concepts shown in Figure 110 constitute different entries in general language dictionaries. To the best of the authors’ knowledge, their relations are not identified in Portuguese dictionaries, except in the definitions themselves, for example, in the case of the PRIBERAM or the INFOPÉDIA, as we demonstrate below. The lexicographic article “era” in the PRIBERAM¹²⁰ is defined as *‘Divisão da escala de tempo*

¹¹⁹ The relationships can be of two types in the paradigmatic axis: hierarchical and inclusion relationships, and equivalence and opposition relationships. The former help structure terms, dependencies of the hyperonymy/hyponymy or homonymy/meronymy established between them, and the latter establish synonymy, antonymy and co-hyponymy relationships.

¹²⁰ ‘era’, in *Dicionário Priberam da Língua Portuguesa* [online], 2008-2021, <https://dicionario.priberam.org/era> [2021-10-28].

geológico, superior ao período e inferior ao éon [Division of geological time scale, greater than period and less than eon] (PRIBERAM [emphasis added]). In the INFOPÉDIA¹²¹, the lexicographic definition is *‘unidade de divisão de tempo geológico, hierarquicamente inferior ao éon e superior ao período, definida por critérios paleontológicos e litológicos’* [unit of geological time division, hierarchically lower than the éon and higher than the período, defined by paleontological and lithological criteria] (INFOPÉDIA, 2021 [emphasis added]). On the contrary, and since we are modelling a concept system, we do not propose including this feature in the definition because the information given is not essential to define the given concept but may help understand it. Instead, we will recommend a note to provide additional information, and our diagrams could be made visible to the end-users so that they understand how the terms are interlinked and can visualise the relations between concepts, which are generally found isolated in these types of lexicographic works because they generally follow the alphabetical order. As explained above, this information need not be expressed in the definition because it is not a delimiting characteristic of the concept. One of the possible ways to represent these relations – which already follow terminological methods – is to annotate them in TEI (see Chapter 9, 9.3.3 *Encoding Semantic Relations*, p. 303). Users will better understand them because they can see the visual representations of these relations that will appear associated with the geological sense of “era”.

Concerning the definition of “unidade geocronológica”, not included in the DLPC, we will propose a definition considering the information retrieved from the following diagram (Figure 111):

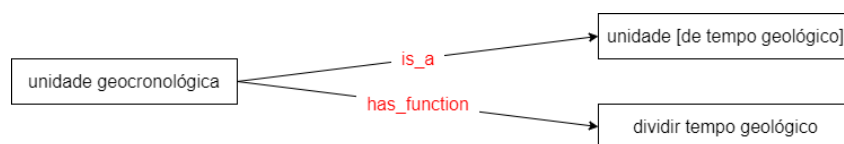


Figure 111: Representation of the relation the conceptual markers *is_a* and *has_function* established from <GeochronologicUnit>

¹²¹ Porto Editora – ‘era’, in *Dicionário Infopédia da Língua Portuguesa* [online]. Porto: Porto Editora. [2021-10-28]. Available at <https://www.infopedia.pt/dicionarios/lingua-portuguesa/era>.

The conceptual relation marker `is_a` establishes a hierarchical relation of subsumption. The conceptual marker `has_function` indicates the functionality of the unit. We assume that we are in the presence of the so-called complex relationships (Sager, 1990, pp. 34–35), which are domain- and application- dependent – this is an associative conceptual relation. Thus, we propose the following definition for “*unidade geocronológica*” in DLP: ‘*unidade que divide o tempo geológico; subdivisão do tempo geológico*’ [unit that divides geological time; geological time subdivision]. Returning to Figure 110, the reference to related subordinate concepts could be included in an additional note, as we shall see.

b) **Partitive relations:** We have a partitive relation ‘when the *superordinate concept* represents a whole, while the *subordinate concept* represents parts of that whole’ (ISO 704, 2009, p. 13). The parts together form the whole. The superordinate concept in a partitive relation is called the comprehensive concept, representing the whole. The subordinate concept is called the partitive concept, which represents a part of this whole.

To illustrate a partitive concept relation, we again use the geological concepts that correspond to the concept of `<GeochronologicUnit>`: `<Age>`, `<Epoch>`, `<Period>`, `<Era>` and `<Eon>`. As seen above, the terms designating these concepts denote time relations in all rocks, precisely when they were formed, whether stratified or non-stratified. As mentioned in Chapter 6, the primary means by which geological time information is conveyed is through the Geological Time Scale and its units. Thus, all these units are part of the `<GeologicalTimeScale>`. This is represented in the following rake diagram.

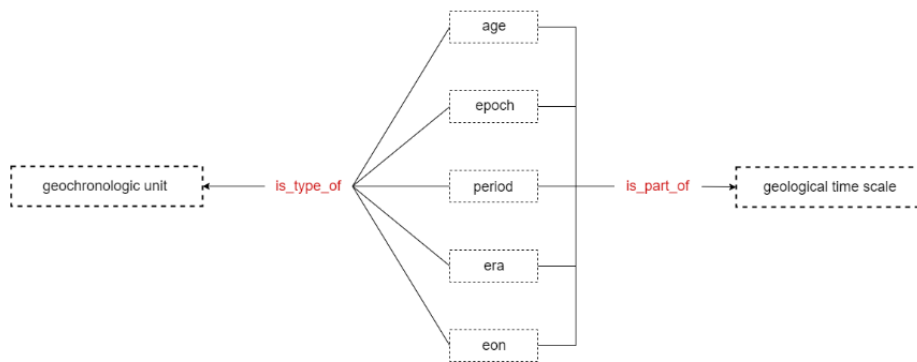


Figure 112: Representation of a partitive relation using the concepts of <GeochronologicUnit> and <GeologicalTimeScale>

Partitive relation can be expressed by the formulae

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

In other words, *The concepts <Age>, <Epoch> and <Period> are constituent parts of the Geological Time Scale.*

The conceptual relationship between the broader concept and its parts was made explicit through the conceptual marker `part_of`. Contrary to what was observed in generic relations, the principle of inheritance does not apply here, i.e., the concepts in a partitive relation do not inherit the characteristics of the superordinate concepts, but do inherit their parts. The <GeologicalTimeScale> is a comprehensive concept. All identified subordinate concepts – <Age>, <Epoch>, <Period>, <Era> and <Eon> represent parts of a whole, but they have distinctive characteristics concerning the related comprehensive concept. In this task, detecting the essential characteristics (see Chapter 5, p. 128) to identify a concept is crucial to defining a given concept by delimiting its position concerning other concepts as one or a set of characteristics that delimits it. Thus, to differentiate the subordinate concepts above, we have to identify the delimiting characteristic.

In the lexical-semantic field, this means listing the characteristics that distinguish or differentiate a sense from its hyperonym and co-hyponyms. As we shall see later, following that list, we will be able to formulate concept definitions or, in Aristotle's

words, *differentia*. It is also important to note that the marker *part_of* designates a part-whole holonymy/meronymy lexical-semantic relation.

In Chapter 6, while presenting the DLPC lexicographic article “era”, we had observed that the polylexical terms, “era primária” [primary era], “era quaternária” [quaternary era], “era secundária” [secondary era], and “era terciária” [tertiary era] appeared as sublemmas. Comparing the Portuguese lexicographic article with the DLE and the DAF, we also observed that these dictionaries include these polylexical terms in examples. In the DLPC, we found each polylexical term presents a definition, followed by synonyms in small capitals: “PALEOZÓICO”, “PRIMERO” [palaeozoic, primary] for the primary era; “ANTROPOZÓICO”, “QUATERNÁRIO” [antropozoic, quaternary] for the quaternary era; “MESOZÓICO”, “SECUNDÁRIO” [mesozoic, secondary] for the secondary era, and “CENOZÓICO”, “TERCIÁRIO” [cenozoic, tertiary] for the tertiary era.

These distinctions are classic designations that fell out of favour in the 20th century. The stratigraphic terminology emerged gradually as the rock bodies were studied. As the terminological variation increased from author to author, the creation of the ICS was highly relevant. From a paleontological point of view, the time after the Pre-Cambrian was divided into these four great eras, each of which is defined by the dominant forms of life. The concept of <Quaternary>, as knowledge evolved, underwent a conceptual change. Moreover, today, we no longer speak of the <QuaternaryEra> since it is an anachronistic concept from the original subdivision of rocks. The term “quaternary” was reintroduced in the *International Chronostratigraphic Chart* as <Period> or <System> of the <Cenozoic> or <Tertiary>. In other words, after this analysis, we confirm that all academy dictionaries are outdated regarding the treatment of the terms “quaternário” or “era quaternária”.

The links between all these concepts related to the concept of <GeologicalEra> can be represented through a tree.

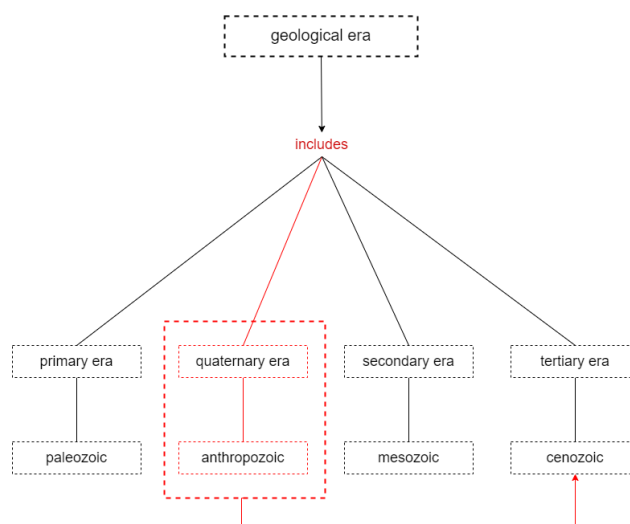


Figure 113: Representation of a generic relation using the concept of <GeologicalEra>

In this concept diagram (Figure 113), <GeologicalEra> is the generic or superordinate concept and <PrimaryEra>, <SecondaryEra> and <TertiaryEra> are the specific or subordinate concepts. As mentioned above, the concept of <QuaternaryEra> has undergone a conceptual change and is now referred to as <Period> of the <TertiaryEra> or, more commonly, of the <Cenozoic>.

The type of concept relation was made explicit using the linguistic marker *includes*, which structures a generic/specific type relation. The specific concepts here inherit characteristics from their generic superordinate concept. The subordination relation is also represented in the dictionary itself, since “era” is a lemma; in contrast, “era primária” or “era terciária” are polylexical terms that appear as sublemmas within the lexicographic article “era”. We do not see this representation in online dictionaries as problematic, since end-users, when searching for “era primária”, for example, could be automatically directed to the polylexical term they are searching for without having to read the entire lexicographic article linearly in search of the desired expression. The concept of <Era> represented here corresponds to the following definition in the geological context: *‘unidade de divisão do tempo geológico (unidade geocronológica), que integra vários períodos’* [geological time division unit (geochronologic unit), integrating several periods]. Although we could have chosen to use these terms in the definition itself, we recognised that they are specialised and decided not to use them in the definition’s wording to avoid circularity. Keeping it, the user, if they want, may

consult the related terms to familiarise themselves with unknown concepts in the geological domain. As notes, we include the following information retrieved from our analysis: ‘1) *A era corresponde ao intervalo de tempo geológico durante o qual se depositou um eratema (unidade cronostratigráfica).*’ [The era corresponds to the geological time interval during which an erathem (chronostratigraphic unit) was deposited.]; ‘2) *Na escala do tempo geológico, a era é hierarquicamente superior ao período e inferior ao éon.*’ [On the geological time scale, the era is hierarchically superior to the period and inferior to the eon].

Concerning the terms “primary era”, “quaternary era” and “secondary era”, pursuant to a discussion with the specialist Professor M. J. Lemos de Sousa, the Dictionary Committee decided to furnish information concerning the old senses, since they are highly likely to be found in literature or using a search engine like Google (see the end of this chapter, Figure 118). To make end-users aware of this update, we concluded that it is worth keeping this sense as marked, using a temporal usage label: ‘*obsoleto*’ [obsolete]. In the case of the conceptual change of “quaternary era”, we have decided to make a cross-reference to “quaternário” [quaternary] as period/system while signalling the new and more current usage.



We will now show some concept relations from the FOOTBALL domain. We can do the same exercise using the concept of <FootballPlayer>, the concepts that refer to the positions of football players in the field, and the concepts of <Back> and <Winger>.

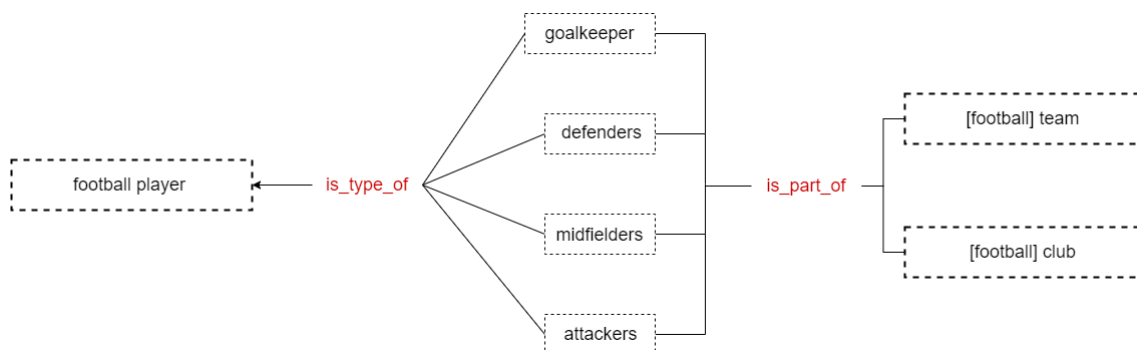


Figure 114: Representation of a mixed concept system with the concepts of <Back> and <Winger>

In Figure 114, we have a mixed concept system (ISO 704, 2009, p. 19), i.e., one ‘constructed using a combination of concept relations’ (*ibidem*, p. 19).

Here, we use a rake diagram as established. The concept of <FootballPlayer> is part_of a <FootballTeam> and <FootballClub>, as we demonstrate on the right-side as a partitive relation. On the left side, a relationship of the generic/specific type is demonstrated, where <FootballPlayer> is the superordinate concept, and <Goalkeeper>, <Defenders>, <Midfielders>, and <Attackers> are the specific or subordinate concepts. In Portuguese general language dictionaries, the term “futebolista” has been defined only as a football player (cf. INFOPÉDIA, PRIBERAM, HOUAISS). However, in FOOTBALL, the concept always refers to a /professional athlete/ (and not an amateur footballer or one who practices for pleasure). To refer to this professional activity (i.e., a /player hired by football club/), it is important to introduce this characteristic into the definition. Thus, the concept of <FootballPlayer> represented here corresponds to the following definition: ‘*atleta profissional que joga futebol; jogador de futebol*’ [professional athlete who plays football; football player]. A person can play football every day but that does not make them a football player. The terminological tasks enable the lexicographer to delimit the concept well and specify its definition.

We will now include two other football terms shown in Chapter 6, “extremo” [winger] and “lateral” [back] in our analysis to explain their inclusion in this concept system.

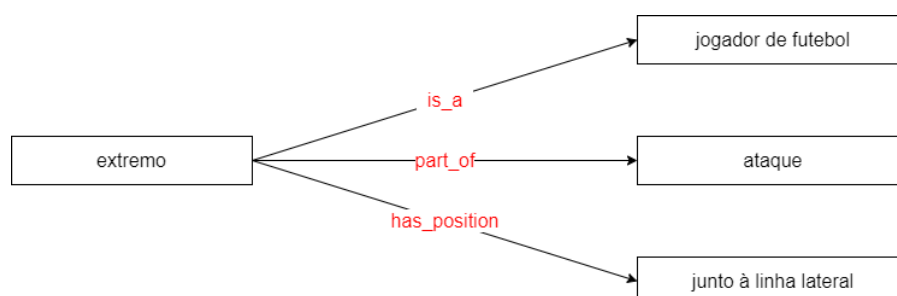


Figure 115: Representation of the relation of the conceptual markers *is_a*, *part_of*, and *has_position* established from <Winger>

As we can see in Figure 115, <Winger> is_a <FootballPlayer> who is part_of the <Attack> and has_position on one of the sidelines, i.e., acts on one of the sidelines. In this case, when drafting a definition, the lexicographer must consider that this is not just a football term but also occurs in other sports, such as basketball¹²². For example, the DLPC defines the term as *'jogador de futebol, basquetebol... que actua junto à linha lateral'* [football player, basketball player... who plays by the sideline]. These definitional formulae are traditional in lexicography; however, we see no need to introduce this information here. We will not use the FOOTBALL label, but rather SPORT and, in a note, we will introduce the following information: *'Nota: Termo recorrente em desportos coletivos, designadamente no futebol e no basquetebol.'* [N.B.: Recurring term in certain team sports, namely football and basketball.] The new definition of the DLP is: *'jogador que faz parte do ataque de uma equipa e que atua num dos lados do campo, junto à linha lateral'* [player who is part of a team's attack and acts on one side of the field by the sideline]. As of writing this thesis, the definition and note have not yet been concluded, as they require validation in other team sports. Depending on whether it acts along the right or left sideline, the terms "extremo-direito" [right-winger] and "extremo-esquerdo" [left-winger] appear. This information can also be included in a note as a cross-reference to these other two terms: *'2) Cf. extremo-direito; extremo-esquerdo.'* [2) Cf. right-winger; left-winger.].

The conceptual markers also led us to identify some lexical-semantic relations. The marker part_of designates a part-whole holonymy/meronymy relation. *If X is a constituent part or a member of Y, X is a meronym of Y, and Y is a holonym of X, i.e., "attack" is a holonym of "winger"*.

We will now analyse the "lateral" [back] term. The definition of the DLPC, *'jogador que actua junto da linha lateral do campo'* [player who acts by the sideline of the field], did not permit distinguishing the concept of <Back> from <Winger>. Therefore, we decided to analyse the concept.

¹²² It should be noted that in a multilingual work, the same concept may correspond to different equivalents. Just for the sake of being precise, in English, the term "winger" is used in used in hockey, but in basketball "wing" is used more often.

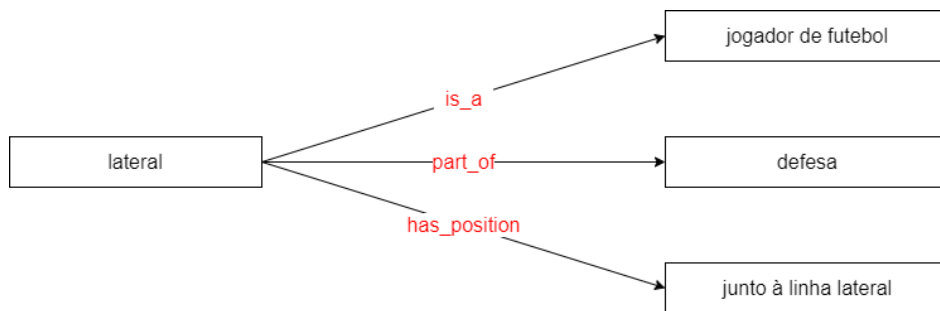


Figure 116: Representation of the relation of the conceptual markers *is_a*, *part_of* and *has_position* established from <Winger>

As we see in Figure 116, a <Back> *is_a* <FootballPlayer> that is *part_of* the <Defence> and *has_position* in one of the sidelines, i.e., acts on one of the sidelines – hence the polylexical term “wing-back”, which is also common in Portuguese, “defesa lateral”. When comparing the two concepts <Winger> and <Back>, the delimiting characteristic, i.e., the characteristic that truly determines a differentiation between the two concepts is the identified partitive relationship: while the former is an attacker, the latter is a defender. To avoid a circular definition, we do not use the term “linha lateral” [sideline] in the definition when defining the concept of <Back>. Therefore, we propose the following definition: *‘jogador que geralmente faz parte da defesa e que atua junto a uma das linhas que delimitam o campo em comprimento (linha lateral)’* [player who is usually part of the defence and who acts along one of the lines that delimit the field in length (sideline)]. The ‘generally’ was introduced in the definition at the specialist’s request since some game tactics require the player to move off defence. This happens in offensive schemes where the backs have the duty of supporting the attacking plays.

The conceptual markers also led us to identify some lexical-semantic relations. The marker *part_of* designates a part-whole holonymy/meronymy relation. In this case, “defence” is a holonym of “back”.

The relationships established between the concepts emerge gradually in the definitions. Taking this last example, we verify that in the definition of the term “lateral”, the concepts of <Attack>, <Defence>, and <SideLine> appear, all of which will be defined in the DLP.

Combined, all these concepts are part_of a <FootballTeam> and a <Football Club>.

We have been constantly and progressively testing the inclusion of new concepts in systems and validating their introduction.



We will now refer to non-hierarchical relations, i.e., associative relations.

c) **Associative relations:** ‘An associative relation exists when a thematic connection can be established between concepts’ (ISO 704, 2009, p. 17). These types of concepts are not hierarchically related but have a robust semantic or pragmatic connection. Some examples of associative relations cited by the standard are marked with dichotomic labels such as cause-effect, matter–substance–property, quantity–unit.

To illustrate an associative concept relation, we continue with the concept of <GeochronologicUnit>. To understand this concept, the concepts of <Time> and <Geochronology>¹²³ are crucial. These, in turn, necessarily call for the related concepts of <Rock>, <Chronostratigraphy>¹²⁴ and <ChronostratigraphicUnit>. Geochronology expresses the timing or age of events in Earth’s history. However, it can also qualify rock bodies, stratified or unstratified, concerning the time intervals at which they formed. At the same time, chronostratigraphic units are ranked according to the length of time they record. In other words, we could say that the chronostratigraphic units used to designate rock bodies that formed contemporaneously correspond to the geochronologic units used to designate the intervals at which they formed.

To clarify the definition of <ChronostratigraphicUnit>, we repeated the exercise we did for <GeochronologicUnit>.

¹²³ The *Stratigraphic Guide* defines geochronology as ‘The science of dating and determining the time sequence of the events in the history of the Earth.’ See: <https://stratigraphy.org/guide/chron>.

¹²⁴ The *Stratigraphic Guide* defines chronostratigraphy as ‘The element of stratigraphy that deals with the relative time relations and ages of rock bodies.’ See: <https://stratigraphy.org/guide/chron>.

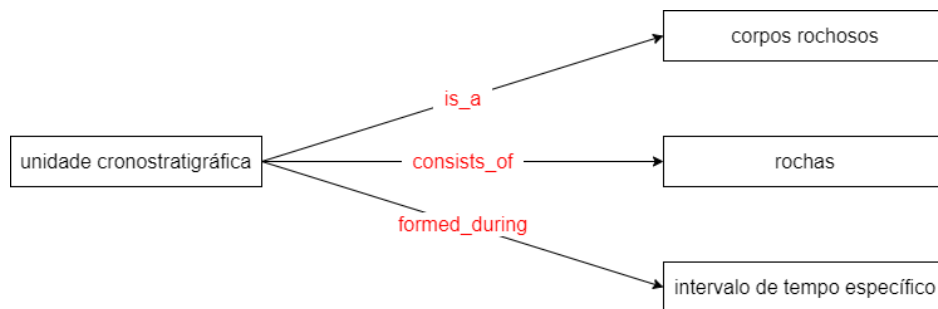


Figure 117: Representation of the relation between the conceptual markers *is_a*, *consists_of* and *formed_during* established from *<ChronostratigraphicUnit>*

In Figure 117, we highlight the conceptual relation marker *consists_of*. It indicates the compositional structure of the concept *<ChronostratigraphicUnit>*. The next lexical marker establishes a temporal relation identified by the lexical marker *formed_during* – again, a complex relationship to be further explored. Our definition: *‘corpos rochosos que incluem as rochas formadas durante um intervalo específico de tempo geológico’* [a set of rocks that includes all rocks that were formed during a specific interval of geologic time].

Here, the two concepts *<ChronostratigraphicUnit>* and *<GeochronologicUnit>* interrelate in a non-hierarchical associative relation since they depend on a certain pragmatic aspect (in this case, based on the dichotomy material-time criterion). The following (Figure 118) diagram presents a line with an arrowhead at each end.

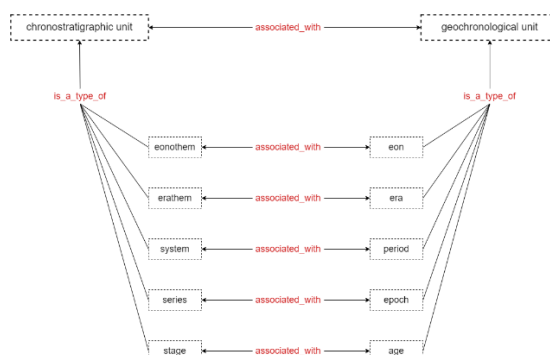


Figure 118: Representation of an associative relationship with the concepts of *<ChronostratigraphicUnit>* and *<GeochronologicUnit>* with generic and partitive relations – a mixed concept system

Associative relations, in terminology work, are always bidirectional. In this case, we have a non-hierarchical relation: material–time. As mentioned above, the concept <ChronostratigraphicUnit> is related to <GeochronologicUnit> – this is a material relation. If one wishes to allude to the <time> – a relationship of temporal dependency – when these strata were deposited, then the concept of <ChronostratigraphicUnit> is replaced by that of <GeochronologicUnit>. First, we identified the highest genus, i.e., a subsumption relation – a hierarchical relation in which a given generic concept (genus) subsumes specific concepts (species). In the next diagram (Figure 119), the associations between <Eonothem>–<Eon>, <Erathem>–<Era>, <System>–<Period>, <Serie>–<Epoch>, <Stage>–<Age> are visible.

Before presenting a last, more elaborate diagram, we emphasise two other key concepts for interpreting the *International Chronostratigraphic Chart*: <RelativeDating> and <AbsoluteDating>. The former consists of a dating process that enables us to assess the age of a particular geological formation using stratigraphic indicators such as the fossil record. The latter determines the age of geological formations or certain events, referred to in numerical values, usually millions of years (M.a.), using specific techniques like radiometric dating.

In the following diagram (Figure 119), we present a sample of the elaborate system of the concept of <Phanerozoic>.

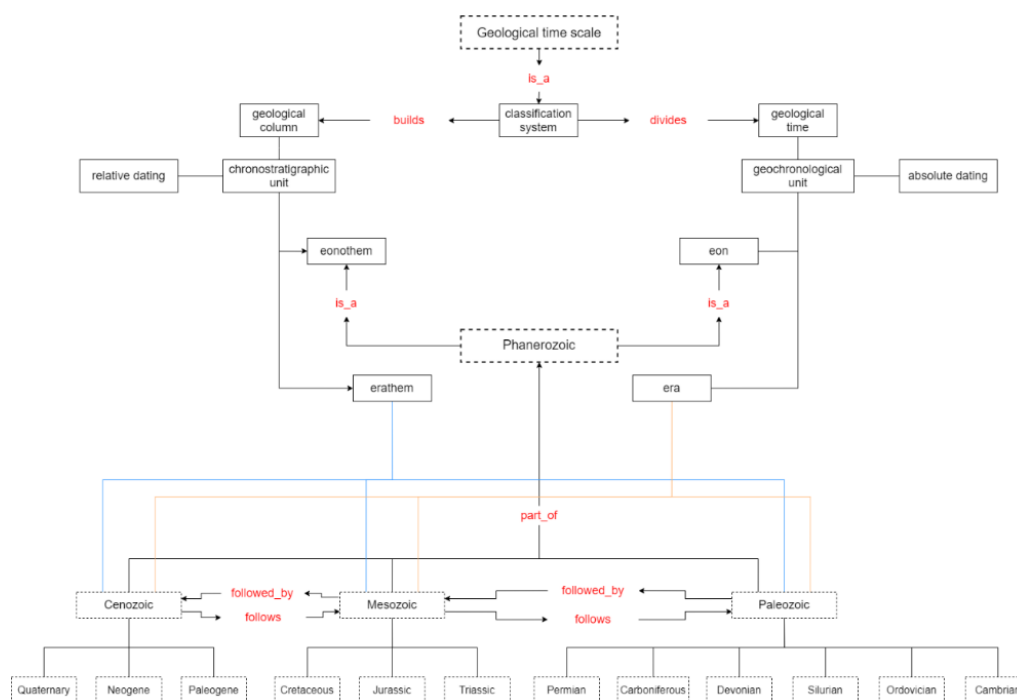


Figure 119: Conceptualising <Phanerozoic>

The degree of specificity becomes higher, and the intension of the concept becomes narrower. As we can see above, <Cenozoic>, <Mesozoic> and <Palaeozoic> are more specific than <Phanerozoic>. These relations correspond to the so-called hyponymy-hypernymy relationships in the lexical-semantic field and are always symmetric. Whenever *x* is a hyponym of *y*, *y* is a hypernym of *x* and *vice versa*. In other words, “Cenozoico”, “Mesozoico” and “Paleozoico” are hyponyms of “Cenozoico”, which is a hypernym of the former. Generally, a hypernym has more than one hyponym term. Thus, all terms that designate geological systems/periods are hyponyms of geological erathems/eras.

We leave only the following final note: all the terms included in the *International Chronostratigraphic Chart* will be included in the dictionary as well, except stage designations – information that we consider possible to reserve for specialised or terminological dictionaries. In addition, a concept system clarifies the relations between concepts in a subject field, facilitating the formulation of definitions that reflect the concept system.

7.9 Editing Lexicographic Content

Before lexicographers start editing the content, the related concepts and the type of *inter se* relations must be already identified. Concerning the definitions in the DLPC, we need:

(1) to reformulate existing definitions because they are outdated or lack scientific reasoning.

(2) to formulate new definitions based on the concept systems.

Thus, we have identified two different tasks inherent to this activity: (i) the identification of definitory problems and (ii) the proposal for definitions and notes (cf. Silva, 2014, pp. 147–149).

7.9.1 Identifying Definitory Problems

DLPC definitions do not follow a structured lexicographic definition model and, unfortunately, it is easy to find inconsistencies. Looking at two related terms that have been explored in the previous sections, “era” and “época” [epoch], we do not find any relationship between them. Marked with the GEOLOGIA domain label, “era” is defined as *‘cada uma das grandes divisões do tempo geológico, cujos limites estão marcados por mudanças geológicas ou paleontológicas e que abrange vários períodos’* [each of the great divisions of geological time, whose boundaries are marked by geological or paleontological changes and which span several periods] while “época” is defined as *‘intervalo de tempo, nas divisões estratigráficas, que é relativo às formações de uma série ou conjunto de terrenos; subdivisão do período’* [time span, in the stratigraphic divisions, which is relative to the formation of a series or set of terrains; period subdivision]. These two concepts, <Era> and <Epoch> are characterised as being /time span/, but this characteristic is not delimited in the definition of “era”. This highlights that these entries were, in all probability, written by different lexicographers and that there was no systematic harmonisation afterwards. To define concepts consistently, we recommend analysing definitions by terms whose concepts are directly related (e.g., defining together the entries that refer to geochronologic units or, in the other example, football player positions).

Generally, the most frequent problems in this type of exercise were (see ISO 704, 2009, p. 30):

- a) definitions that do not refer to the concept being defined;
- b) definitions that contain unnecessary characteristics (it is crucial to separate the conceptual characteristics essential to the definition from the secondary characteristics that can be a note);
- c) definitions that include the term to be defined;
- d) definitions that are too long.

7.9.2 Reformulation Definitions and Notes

As indicated in Chapter 5, the ISO standards (ISO 704, 2009; ISO 1087, 2009) distinguish between intensional definition and extensional definition. The former consists of listing the immediate superordinate concept and delimiting the characteristics of the defined concept; the latter comprises listing its subordinate or partitive concepts. The definition by analysis or *genus-differentia* (Sager, 1990) corresponds to the intensional definition of ISO standards.

The intensional definition does not contain features belonging to other superordinate or subordinate concepts: it (1) clarifies only the class to which the defined concept belongs; (2) specifies what distinguishes it from other concepts situated in the same class; and (3) lists all its essential features.

Intensional definitions based on generic associations include the superordinate concept, followed by the delimiting characteristics within a concept system (e.g., <Era> among <GeologicalTimeSpan>). The superordinate concept's characteristics (that make up the intension) are assumed in the definition, which is the inheritance principle. Establishing conceptual relations facilitates the lexicographer's work, imparting greater consistency and ensuring good data harmonisation and standardisation. It also enables the creation of a definitory model, e.g., <GeochronologicUnit> [superordinate concept] + *formed_during* [subordinate concepts].

To illustrate this, Table 12 presents five different terms extracted from the DLPC and compares them with the definitions of the DLP written by us after modelling the concept systems. All of them define a type of <GeochronologicUnit>:

HEADWORD	DLPC (2001)	DLP (2021)
éon [eon]	<i>Geol.</i> 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]	<i>Geol.</i> 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]	<i>Geol.</i> 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 (unidade geocronológica) durante o qual se formou um andar (unidade cronostratigráfica) 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 crono.

Table 12: Comparison of definitions ‘éon’, ‘era’, ‘período’, ‘época’, ‘idade’ in DLPC (2001) and DLP (2021)

If we observe the proposed definitions, the consistency and systematisation in the treatment of terms are remarkable, compared to the lack of systematisation evident in the previous edition.

A terminologist may find it weird to use curved parentheses in definitions relating to chronostratigraphic and geochronological units. However, its use is purposeful and a lexicographic principle adopted in some Portuguese dictionaries (e.g., Houaiss, 2015, cf. ‘*remissão discreta*’ [discrete cross-reference]). The inclusion of parenthetical information is a way of suggesting to the end-user the consultation of other dictionary terms for further clarification. Incidentally, the same terms could have been used in the definitions themselves. However, we avoided them as we considered them quite specialised and difficult to grasp for an ordinary user. Finally, the specialist considered the introduction of parentheses in these cases essential for a good understanding.

These terms are geochronologic units, which are hyponyms (specific meaning), while the geochronologic unit is a hypernym (generic), which is established by using the conceptual marker *is_a* in our modelling. Then, we have a lexical-semantic relation of holonomy-meronymy. This relation was established through the conceptual marker *part_of*.

The definition of <Age> corresponds to a literal or strict sense of the term, instead of the common generic sense relating to the elapsed time. Most

misunderstandings in the definitions are due to the confusion between the following two very distinct entities: the rocks present in <Rock> (chronostratigraphic units) and <Time> corresponding to their genesis (geochronologic units). There are more misunderstandings regarding the definition of <Age> when taken in a narrow sense (time of formation of a stage) or a broad sense, i.e., the latter when it refers to chronological time in general. Further, the subject fits into the general rules of systematics, in which it is essential, in the domain of taxonomy, to define not only the base unit (in this case, the stage) but also the hierarchy (ascending or descending) of the different divisions, as indicated in the diagram by the arrows.

The same methodology was applied to terms relating to chronostratigraphic units. The position of the individual unit within the geological hierarchy is decided by the time interval represented by each unit:

HEADWORD	DLPC (2001)	DLP (2021)
eonotema [eonothem]	—	conjunto de rochas (unidade cronostratigráfica) formadas durante um éon (unidade geocronológica) Nota: Na escala cronostratigráfica, o eonotema é a categoria hierárquica mais elevada.
eratema [erathem]	—	conjunto de rochas (unidade cronostratigráfica) formadas durante uma era geológica (unidade geocronológica) Nota: Na escala cronostratigráfica, o eratema é hierarquicamente superior ao sistema e inferior ao eonotema.
sistema [system]	<i>Geol.</i> período geológico que se caracteriza pela fauna, flora e mutações próprias	conjunto de rochas (unidade cronostratigráfica) formadas durante um período geológico (unidade geocronológica) Nota: Na escala cronostratigráfica, o sistema é hierarquicamente superior à série e inferior ao eratema.
série [serie]	—	conjunto de rochas (unidade cronostratigráfica) formadas durante uma época geológica (unidade geocronológica)

		Nota: Na escala cronostratigráfica, a série é hierarquicamente superior ao andar e inferior ao sistema.
andar [stage]	<i>Geol.</i> conjunto dos terrenos ou das camadas geológicas correspondentes a uma idade <i>O andar é definido pelos seus fósseis característicos.</i>	conjunto de rochas (unidade cronostratigráfica) formadas durante uma idade geológica (unidade geocronológica) 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 subandar e cronozona.

Table 13: Comparison of definitions ‘eonothem’, ‘erathem’, ‘system’, ‘series’, ‘stage’ in the DLPC (2001) and the DLP (2021)

In formulating these definitions, we followed the concept systems previously modelled and were also particular about writing definitions that will be useful to the intended user. As we can see, most of the terms are not included in the 2001 edition (DLPC). The terms “eonothem” and “erathem”, for example, do not figure in current Portuguese dictionaries. We cannot understand why some were included and others were not, and can only attribute it to a lapse. Their introduction is justified in methodological terms and because those units are included in geology textbooks. Following the presented methodology will avoid this type of lapse in the future since we defend the treatment of terms by the relationship they establish with each other and not precisely by planning a dictionary revision based on alphabetical ordering.

In addition to the definitions, we aim to comment on the use of notes in the two tables presented above. Our proposed definitions contain only the characteristics that are necessary to identify the concepts. Any additional information is included as a note. Lexicographers would have to add explanations, contexts, notes, encyclopaedic information, or even some representation in other media. This is especially relevant in the football context. For example, to illustrate what a “trivela” is in the football context, a link to a *YouTube* video could be provided – a link¹²⁵ showing, for example, Quaresma (a Portuguese football player) curling the ball.

Though the extent of a note as it is given on “andar” [stage] can be awkward, we are working on an academy dictionary with slightly different aims than purely commercial dictionaries.

Now, to illustrate the definition of a partitive relation, we resorted to erathems, and eras comprised by the <Phanerozoic>: <Palaeozoic>, <Mesozoic> and <Cenozoic>. All these concepts are defined as a part_of the most generic concept of which they are a part (Phanerozoic period). In the new definition, we could have used a lexical marker such as ‘that is part of’ but we prefer to use the formula “of the Phanerozoic” instead.

HEADWORD	DLPC (2001)	DLP (2021)
cenozoico [cenozoic]	<i>Geol.</i> divisão cronológica da história da Terra, anterior ao Antropozóico e posterior ao Mesozóico, que engloba cerca de 65 milhões de anos, compreendendo os períodos Neogénico e Paleogénico e que se caracteriza pelo aparecimento dos primeiros primatas, pelo desenvolvimento e crescente domínio dos mamíferos e pelo arrefecimento progressivo do clima; era terciária; terciário	1) designação do eratema superior (<i>unidade cronostratigráfica</i>) do eonotema Fanerozoico, correspondente ao conjunto de rochas formadas durante a era (<i>unidade geocronológica</i>) respetiva 2) designação da era tardia (<i>unidade geocronológica</i>) do eón Fanerozoico, correspondente ao intervalo de tempo durante o qual se formaram as rochas do respetivo eratema (<i>unidade cronostratigráfica</i>), entre 66 milhões de anos até à atualidade SINÓNIMOS: terciário

¹²⁵ https://www.youtube.com/watch?v=3yCL8vpmX18&t=49s&ab_channel=Canal11

		Nota: O sistema/período cenozoico integra as séries/épocas: Paleogénico, Neogénico e Quaternário. Nota: Como nome, escreve-se com inicial maiúscula.
mesozoico [mesozoic]	<i>Geol.</i> divisão cronológica da história da Terra, posterior ao Paleozóico e anterior ao Cenozóico, que engloba cerca de 160 milhões de anos, compreendendo os períodos Cretáceo, Jurássico e Triásico e que se caracteriza pelo aparecimento de grandes répteis, aves e primeiros mamíferos, bem como pelas grandes transformações geológicas que conduziram à distribuição actual dos continentes; era secundária; secundário	1) designação do eratema médio (unidade cronostratigráfica) do eonotema Fanerozoico, correspondente ao conjunto de rochas formadas durante a era respetiva (unidade geocronológica) 2) designação da era intermédia (unidade geocronológica) do eón Fanerozoico, correspondente ao intervalo de tempo durante o qual se formaram as rochas do respetivo eratema (unidade cronostratigráfica), entre 251 e 66 milhões de anos SINÓNIMOS: secundário Nota: O sistema/período mesozoico integra as séries/épocas: Triássico, Jurássico e Cretácico. Nota: Como nome, escreve-se com inicial maiúscula.
paleozoico [paleozoic]	<i>Geol.</i> divisão cronológica da história da Terra, anterior ao Mesozóico, que abarca os primeiros 345 milhões de anos do éon fanerozóico, compreendendo os períodos Câmbrico, Ordovícico, Silúrico, Devónico, Carbónico e Pérmico, que se caracteriza por uma grande diversificação da fauna, com o desenvolvimento dos invertebrados e o aparecimento dos primeiros peixes, batráquios, insectos e répteis	1) designação do eratema inferior (unidade cronostratigráfica) do eonotema Fanerozoico, correspondente ao conjunto de rochas formadas durante a era respetiva (unidade geocronológica) 2) designação da era inicial (unidade geocronológica) do eón Fanerozoico, correspondente ao intervalo de tempo durante o qual se formaram as rochas do respetivo eratema (unidade cronostratigráfica), entre 541 e 251 milhões de anos SINÓNIMOS: secundário Nota: O sistema/período paleozoico integra as séries/épocas: Câmbrico, Ordovícico, Silúrico, Devónico, Carbonífero e Pérmico.

		Nota: Como nome, escreve-se com inicial maiúscula.
--	--	--

Table 14: Comparison of ‘cenozoico’, ‘mesozoico’, ‘paleozoico’ definitions in the DLPC (2001) and the DLP (2021)

As seen above, concepts can be grouped into categories, considering their distinctive characteristics. All these units are a designation of an <Era> (geochronologic unit) or an <Erathem> (chronostratigraphic unit) of the <Phanerozoic>, depending on whether one considers a geological time interval or the rocks deposited during that interval. To distinguish one concept from another within the same concept system, the delimiting characteristics of each concept in Table 14 were instrumental in creating the concept systems and consequently for writing definitions. Even further back in time, with the divisions of geological time looser and more insecure, mainly due to the lack of fossil data in good condition in the rocks of the past when life was still simple and not so diversified, the establishment of time boundaries was a distinctive feature.

Chronostratigraphic units are usually defined based on selected type sections that include the entire unit (stratotypes). In contrast, a geochronologic unit is distinguished based on a rock succession and defined by a division of time expressed by a specific number of years. It is also necessary to consider the principle of superposition, according to which the deposition of the strata (sedimentation) always occurs in chronological order from the bottom to the top of the stratigraphic column. This is expressed by the lexical markers ‘hierarquicamente,’ ‘superior,’ ‘inferior’ in the notes. In this way, in a succession of strata whose order has not been altered, each stratum is older than the one that covers it and more recent than the one that serves as its base.

We also detected another type of semantic relationship: intralinguistic equivalence relationships, i.e., synonymous relationships between two or more terms, such as between “Primário” [primary] and “Paleozoico” [palaeozoic] (Table 14). Finally, we just need to explain that the synonyms given are valid for both meanings (senses 1 and 2).

According to ISO 704 (2009), ‘synonyms should never be used in place of a definition in the way they often are in general language dictionaries’ (p. 22). Indeed, general language dictionaries often consist of one or more synonyms. Moreover, ‘a synonym definition is only really acceptable when the *definiendum* and the synonym are semantically identical’ (Atkins & Rundell, 2008, p. 421). We thus agree that synonyms can have a valuable complementary role when supporting a definition.

As we have seen in the preceding chapter, <Palaeozoic> is divided into six periods: <Cambrian>, <Ordovician>, <Silurian>, <Devonian> and <Carboniferous>. We chose to explore the concept of <Carboniferous>.

HEADWORD	DLPC (2001)	DLP (2021)
carbonífero/carbónico [carboniferous]	<i>Geol.</i> período da era primária ou paleozóica que sucede ao devónico e que antecede o pérmico, caracterizando-se pelo aparecimento dos primeiros répteis e insectos alados	1) sistema do eratema Paleozoico e do eonotema Fanerozoico 2) intervalo de tempo geológico (período) durante o qual as rochas desse sistema foram formadas Notas: 1) Na escala cronostratigráfica, o Carbonífero sucede ao Devónico e é anterior ao Pérmico. 2) Como nome, escreve-se com inicial maiúscula.

Table 15: Comparison of definitions of the concepts designated by the terms ‘carbónico’ and ‘carbonífero’ in the DLPC (2001) and the DLP (2021)

As we can see, once the concepts and their relationships are well identified, the methodological steps are iteratively repeated. Above all, we define a concept concerning its place in the knowledge system. The delimiting characteristics determine or differentiate a given concept from others and play a crucial role in defining terms. The sum of these characteristics is the intension of the concept. On the other hand, we need to consider the distinctive characteristics that allow us to differentiate a concept from others close to it.

Throughout this process, the lexicographer must identify the concept to be defined, locate it within the concept system, distinguish it from other concepts, establish

relations between concepts and know how to identify and describe/define the characteristics of the concept. Once all these phases have been completed, the lexicographer will be able to write a definition, avoiding circularity, inaccuracies, or non-essential information, define every lexical unit used in a definition, comply with the replaceability principle and avoid ambiguity and definitions in the negative, all the while following guidelines for good lexicographic practices (cf. ISO 704, 2009, pp. 30–34). Finally, definitions must be intelligible, concise, and a precise statement of what the concept is. The language used should be appropriate for the target audience.

Finally, it is important to remember that the concept systems presented here were subject to validation by specialists.

The same methodology can be duplicated and applied to defining football terms. The result of our analysis is shown in Table 16.

HEADWORD	DLPC 2001	DLP
ataque [attack]	<i>Desp.</i> acto ofensivo com o objectivo de marcar golos ou pontos e de um modo geral de derrotar o adversário	DESPORTO conjunto de jogadores que fazem parte de uma equipa e cuja função principal é atacar a baliza da equipa adversária com o objetivo de marcar golos ou pontos
defesa [defence]	<i>Desp.</i> Conjunto de jogadores que têm como função contrariar o ataque do adversário, actuando na parte recuada do meio campo da sua equipa.	DESPORTO conjunto de jogadores que fazem parte de uma equipa e cuja função principal é proteger a sua baliza
meio-campo [midfield]	—	DESPORTO conjunto dos jogadores que fazem parte de uma equipa e que atuam na zona central do campo

Table 16: Comparison of the definitions of the terms ‘ataque’, ‘defesa’, ‘meio-campo’ in the DLPC (2001) and the DLP (2021)

Finally, we have written new definitions for the concept related to the position in the field.

HEADWORD	DLPC 2001	DLP
guarda-redes [goalkeeper]	<i>Desp.</i> jogador que, no jogo do futebol, andebol, hóquei... ocupa	DESPORTO jogador de uma equipa que atua na baliza, cuja função é

	o último posto de defesa, entre os postes da baliza, tentando impedir a marcação de golos sinónimos arqueiro; (Bras.) goleiro	impedir a entrada da bola na sua baliza com o objetivo de evitar que a equipa adversária marque golos ou pontos SINÓNIMOS: arqueiro (Brasil); goleiro (Brasil) Nota: Termo recorrente em desportos coletivos, designadamente no futebol, andebol, hóquei, etc.
avanzado [attacker]	<i>Desp.</i> jogador que, em certas modalidades, nomeadamente no futebol, se encontra na linha de ataque da sua equipa ≠ defesa.	DESPORTO jogador de uma equipa que faz parte do ataque, cuja função é atacar a baliza adversária com o objetivo de marcar golos ou pontos SINÓNIMO: atacante (Brasil) Nota: Termo recorrente em desportos coletivos, designadamente no futebol, andebol, hóquei, etc.
extremo [winger]	<i>Desp.</i> Jogador de futebol, basquetebol... que actua junto à linha lateral	DESPORTO jogador que faz parte do ataque de uma equipa e que atua num dos lados do campo, junto à linha lateral
lateral [back]	<i>Fut.</i> Jogador que actua junto da linha lateral do campo	DESPORTO jogador de uma equipa que geralmente faz parte da defesa e que atua junto a uma das linhas que delimitam o campo em comprimento (<i>linha lateral</i>), cuja função é estabelecer a ligação entre a defesa e o meio-campo
libero [sweeper]	<i>Desp.</i> Jogador mais recuado de uma equipa de futebol, que tem como função colmatar as brechas provocadas na defesa pela equipa adversária.	FUTEBOL jogador de uma equipa, em posição recuada relativamente aos defesas centrais, cuja função é defender sem a posse de bola e de auxiliar o ataque quando recupera a posse da bola Nota: A designação da posição vem do italiano <i>libero</i> , que significa 'livre', uma vez que para controlar possíveis falhas dos colegas a sua posição tem necessariamente de ser livre.
defesa [defender]	<i>Desp.</i> Jogador de futebol ou de outros desportos, que actua na parte recuada do meio campo da sua equipa.	DESPORTO jogador de uma equipa que faz parte da defesa, cuja função é impedir que a equipa adversária marque golos ou pontos
médio [midfielder]	—	DESPORTO jogador de uma equipa que atua no meio-campo, cuja função principal é fazer a ligação entre a defesa e o ataque

ponta de lança [striker]	<i>Fut.</i> Elemento avançado de uma equipa, geralmente marcado pelos defesas centrais da equipa adversária. avançado.	FUTEBOL jogador mais avançado de uma equipa que atua no meio da defesa adversária, cuja função principal é finalizar as jogadas, marcando golo
------------------------------------	--	--

Table 17: Comparison of the definitions of the terms ‘guarda-redes’, ‘avançado’, ‘extremo’, ‘lateral’, ‘líbero’, ‘defesa’, ‘médio’, ‘ponta de lança’ in the DLPC (2001) and the DLP (2021)

One case in Table 17 caught our attention. Although the DLPC uses the domain label FOOTBALL, in the term “líbero” [sweeper], the label is SPORT, which is incomprehensible since it is an exclusive term in the football context.

Another term we noticed was “guarda-redes” [goalkeeper]. We found a curious detail in the HOUAISS dictionary that enables us to explain what a specific characteristic of a concept is, which can be dispensed with in general language dictionaries. The goalkeeper is the only one who has the right to touch the ball with their hand, as long as they do it in the wide area of his field, and that detail was defined. This particularity is described in the definition itself: *‘jogador que atua na baliza e é o único a ter o direito de tocar na bola com a mão, desde que o faça na grande área do seu campo’* (HOUAISS) [player who plays in goal and is the only one who has the right to touch the ball with their hand, provided they do so in the wide area of the field]. In our case, and considering the templates created for writing the definitions, we did not integrate this feature in the definition of goalkeepers.

In sum, we can argue that the analysis of the relations among concepts is very useful for the successful writing of definitions.

7.10 Validating Terminological Data

Together with the lexicographer, the specialist(s) perform a second moment of validation. In this phase, we schedule new meetings with the specialists to validate our conceptual work concerning the concept systems model and show the linguistic work around the drafting of definitions. This validation process comprises two activities: validating concept systems and validating the new and reformulated definitions and the notes.

7.10.1 Concept Systems

We show our diagrams to the specialist to debate our proposals. All the correlations among concepts must be validated.

7.10.2 Definitions and Notes

The pre-validation treatment consists of bringing together the proposals for the definitions. The definitions are extracted from the database through a list of terms to be presented. A post-validation treatment step follows. At this stage, the lexicographer/terminologist must analyse the results obtained after validation by the experts as well as their possible comments. In the final phase, it may be necessary to arrange a final meeting with the experts (cf. Validation with mediation; Silva, 2014, p. 172). Here, the terminologist/lexicographer must play the role of mediator in order to reach a final consensus.

In this study, in the validation process, the following elements were taken into account: the definition must describe the concept being defined; the definition must be concise and clear, without losing the complexity inherent to the concept; the essential and intrinsic characteristics of the concepts must be identified; the definition must take into account the level of language suitable for the objectives it sets out to achieve; it must take into account the types of audiences it is intended for (Silva, 2014, pp. 176–177). The indication of domains was constantly done via labels. Still following Silva (*ibidem*), as to the form, the definition should avoid using in its text the term that is being defined, opting for the affirmative form and avoiding a paraphrasis.

The discussion with specialists was preferably centred on the conceptual level. We also felt the need to introduce terms in the definition, and we always checked if they were included in the dictionary. The adverb ‘generally’ was used since it was essential, but we should keep its use to a minimum.

7.11 Encoding Terms

The encoding of the work done will be exemplified in Chapter 9.

Currently, the DLP database has an entry tagging system that is built on LeXmart¹²⁶. We have the following statuses: edited, revised and validated. In terms of search, the following possibilities are implemented: simple, reverse and advanced search. The simple search allows searching for a term in its canonical form. The reverse search allows searching for a term by one of its components. Finally, an advanced search allows end-users to see related terms.

Lastly, we demonstrate the result of a lexicographic article after applying the traditional lexicographic methods and adding terminological principles. We created a lexicographic/terminological form template structured as follows (Table 18).

Lexicographic/terminological component	Content	Type
Headword	Term ID (identification of the lexicographic article) Type of lexical unit Lemma (term or concept designation) Pronunciation *Orthographic variants (forms that coexist in parallel in the Portuguese language and cases in which the term has undergone spelling changes)	Text editor
POS	POS (grammatical category) Gender (gender of names and adjectives) Number (number of nouns and adjectives)	Dropdown
SENSE	Usage information > hierarchical domain labels (identification of the domain to which the term belongs) Definition (concept definition) Semantic relations Synonym Hypernym Hyponym Cross-reference (unit that points to related terms) Co-occurrent Usage examples Examples (bibliographical references: title, source of publication)	Dropdown Text editor Text editor
Sub-headword [++ sense]	Term ID (identification of the lexicographic article) Lemma (polylexical term or concept designation)	
ETYMOLOGY	Term origin	Text editor
IMAGE	Images that complement the definition of the term	Link

¹²⁶ <http://lexmart.eu/>

NOTE	General notes, usage notes, encyclopaedic nature, and external links (e.g., media)	Text editor
MANAGEMENT		
LEXICOGRAPHER [in charge]	Who edited the lexicographic article	Dropdown
STATUS	Status (new, revised, edited, needs validation, validated by the expert) Date (term creation/revision date)	Dropdown
COMMENTS	Internal lexicographer/terminologist/expert comments	Text editor

Table 18: *Lexicographic/Terminological form of a term in a general language dictionary*

Fixed combinations can also appear in a given lexicographic article. They may also feature privileged co-occurents. The relations that are established among concepts are annotated in TEI, as mentioned above, and will be demonstrated in Chapter 9.

7.12 Publishing Terms

The best result we can come up with is a complete and finished DLP dictionary entry. We selected a geological term and a football term: ‘era’ from GEOLOGY and ‘defence’ from FOOTBALL. They were edited in LeXmart and are presented in Figure 120 and Figure 121.

- era** [ˈɛɾɐ]
- nome feminino*
- 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
 - era cristã
a que se conta a partir do nascimento de Jesus Cristo
 - era de Cristo
o mesmo que era cristã
 - era hispânica
a que inicia 38 anos da era cristã
 - era muçulmana
a que inicia o calendário islâmico e que corresponde ao ano 622 da era cristã, assinalando a data da fuga do profeta Maomé de Meca para Medina (Hégira)
 - início de uma nova ordem de coisas

ℳ Entrámos numa nova era de maior justiça e liberdade.

 - era atómica era digital; era do fogo; era dos computadores; era espacial
 - qualquer período cronológico demarcado

SINÓNIMOS época ; tempo

 - era dos afonsinhos
tempos antiquados, desatualizados
SINÓNIMOS tempo dos afonsinhos

ℳ Por alusão irónica aos tempos em que reinaram os primeiros Afonsos de Portugal.
 - GEOLOGIA** intervalo de tempo geológico (unidade geocronológica) durante o qual se formou um eretema (unidade cronostratigráfica)

ℳ 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
 - GEOLOGIA, OBSOLETO** designação da era paleozoica
SINÓNIMOS primário
 - GEOLOGIA** o mesmo que paleozoico
 - era quaternária
 - GEOLOGIA, OBSOLETO** denominação da era antropozóica
SINÓNIMOS antropozóico ; quaternário
 - GEOLOGIA** o mesmo que cenozoico
 - era secundária
 - GEOLOGIA, OBSOLETO** designação da era mesozoica
SINÓNIMOS secundário
 - GEOLOGIA** o mesmo que mesozoico
 - era terciária
 - GEOLOGIA, OBSOLETO** designação da era cenozoica
SINÓNIMOS terciário
 - GEOLOGIA** o mesmo que cenozoico
 - numeração que indica um ano ou um certo número de anos

ℳ a era de 74 (1974); a era dos anos 90 (1990)
 - BRASIL** idade; número de anos

ℳ Estes miúdos são da mesma era.
- ETIMOLOGIA** Do latim *aera*

Figure 120: Entry ‘era’ [era] updated in the DLP (2021)

Entrada editada

nome feminino

1. ação de proteger ou proteger-se; ato ou efeito de defender ou de defender-se

☞ *instinto de defesa; reflexo de defesa*☞ **defesa pessoal**

técnica de defesa sem armas, cujo objetivo é neutralizar um ataque surpresa

☞ **em defesa de***locução prepositiva*

em apoio de; em favor de

! O Governo tomou medidas em defesa do meio ambiente.

! O sindicato agiu em defesa dos direitos dos trabalhadores.

☞ **jogar à defesa**

não se expor

! Não consegue ser espontânea por estar permanentemente a jogar à defesa.

2. o que serve para proteger, defender

! O guarda-chuva era a sua única defesa contra o aguaceiro.

! A trincheira era a única defesa dos soldados.

! Este forte já foi a principal defesa da ilha.

☞ *arma de defesa; mecanismo de defesa; meio de defesa*☞ **defesa nacional**

conjunto de meios destinados a assegurar a integridade territorial de um país contra possíveis ataques estrangeiros

3. ato de sustentar ou de defender (uma ideia, uma causa, proposição, etc.)

SINÓNIMOS *alegação; justificação*☞ **defesa de tese**

ato solene em que um candidato a determinado grau académico apresenta um trabalho inédito e argumenta em sua defesa perante um júri

☞ **não ter defesa (possível)**

INFORMAL

não ter qualquer justificação para a prática de um ato considerado errado e reprovável

4. DIREITO advogado que defende um réu em juízo

! O juiz deu a palavra à defesa.

! A defesa apresentou as alegações finais.

☞ *advogado de defesa*

5. DIREITO ato de justificar, defender ou defender-se um réu em tribunal contra uma acusação, expondo factos e produzindo provas em seu favor

SINÓNIMOS *alegação*

! A defesa e a acusação.

6. DIREITO conjunto de procedimentos adoptados pelo réu ou pelo seu advogado de defesa para refutar as acusações

! O advogado preparou cuidadosamente a defesa do arguido.

☞ **em legítima defesa***locução adverbial*

em reação a uma agressão ilícita, atual ou iminente

! Disparou em legítima defesa.

! Matou em legítima defesa.

7. ato de impedir, de proibir

SINÓNIMOS *impedimento; interdição; proibição*

8. ZOOLOGIA chifres de animais

SINÓNIMOS *chifre; cornu; haste*

! Usado no plural.

9. ZOOLOGIA dente muito saliente em alguns animais

SINÓNIMOS *presa*

! Usado no plural.

! As defesas do elefante. As defesas do javali.

! As morsas defrontam-se, utilizando as defesas.

10. FISIOLOGIA resistência natural do organismo contra agressões externas

11. DESPORTO jogada ou tática defensiva

! O guarda-redes fez uma defesa incrível.

12. DESPORTO ação de contrariar o ataque, a investida do adversário

SINÓNIMOS *ataque*☞ **jogar à defesa**

FUTEBOL procurar defender a sua baliza, sem atacar, sem procurar marcar golos

«Errado é jogar à defesa.» (DN,26.10.1988)

13. DESPORTO conjunto de jogadores que fazem parte de uma equipa e cuja função principal é proteger a sua baliza

! A equipa adversária deparou com uma defesa intransponível.

14. EQUITACÃO resistência oposta por um cavalo à transposição de um obstáculo ou para se subtrair à ação do cavaleiro

15. ANTIGO preservativo

16. REGIONALISMO (ALENTEJO) grande herdade ou propriedade rústica

nome masculino e feminino

DESPORTO jogador de uma equipa de futebol que faz parte da defesa e que atua junto a uma das linhas que delimitam o campo em comprimento (linha lateral)

«ficou na frente, não apenas para dar continuidade aos lances mas também para 'prender' os defesas contrários.» (DN,27.10.1988)

ETIMOLOGIA Do latim, *defensa*

Figure 121: Entry 'defesa' [defence] updated in DLP (2021)

All these tasks involved iterative work in both the linguistic and conceptual dimensions. The results obtained are immensely satisfactory, ensuring better lexical organisation and greater definition accuracy, consequently improving the overall quality of the lexicographic articles.

PART III

ENCODING AND MODELLING DICTIONARIES

CHAPTER 8

Standards for Structured Lexicographic Resources

We should acknowledge that machine readable dictionaries as well as terminological databases, even if conceived to fulfil other types of requirements, should not be seen as completely separated resources which would deserve unconnected standardisation activities.

ROMARY (2013, p. 1266)

This chapter provides an overview of the most well-known and widely used formal representations and standardised models within the lexicographic universe (Tiberius et al., 2020) aimed at creating lexicographic resources as a result of legacy print or born-digital resources. The description we provide aims to (1) trace a broad framework of models for the representation of language, and (2) reflect on specific problems related to the representation of lexicographic content. The TEI is especially important in the context of this research as we chose to apply a serialisation of the TEI to our research data. These guidelines are a long-standing tradition with an excellent reputation in scholarly dictionary projects. After contextualising and describing the main features of some standards applied to structured lexicographic resources, emphasising their strengths and limitations, our focus is on the TEI Lex-0, a new baseline encoding and a target format for lexicographic data.

The complexity and heterogeneity of lexicographic resources have been recognised by the scientific community (Müller-Spitzer, 2008; Romary & Wegstein, 2012; Pilehvar & Navigli, 2014; McCracken, 2016; McCrae et al., 2019; Salgado et al., 2019, among others) owing to the diversity of their structural components and the numerous resources that obey various criteria for representing and processing lexicographic data with different levels of information (e.g., orthographic, morphological, phonetic, semantic, syntagmatic, etymological).

With respect to lexicography, standards establish specifications and procedures, provide a common and consistent language, guarantee the material's reliability and interoperability, and try to facilitate the representation of lexicographic data. A survey of user needs (Kallas et al., 2019) was carried out in the context of the ongoing ELEXIS

project. The survey results concerned data formats and standards and showed that although many lexicographic projects use XML or databases, there are still projects working with unstructured data and text formats. The authors (Kallas et al., 2019) outlined two main trends: 'a) a transition from non-structured data or text format to structured data format, b) still insufficient use of (standardised) structured formats enabling reliable re-use and linking of dictionary data' (pp. 54–55).

8.1 ISO Standards for Lexicography

The International Organization for Standardization (ISO)¹²⁷ is an international non-governmental organisation composed of several national standardisation bodies that develop and publish a wide range of standards. The international standards are the result of the work carried out through ISO Technical Committees, generally composed of specialists and governmental and non-governmental international organisations.

The standards that are of interest for this work are the ones developed by the TC37, 'Language and terminology', namely the Subcommittee 2 (SC2)¹²⁸, 'Terminology workflow and language coding', and Subcommittee 4 (SC4)¹²⁹, 'Language and resource management'.

Regarding lexicographic standardisation, the third edition of ISO 1511 (2007), 'Presentation/representation of entries in dictionaries – requirements, recommendations and information', under the direct responsibility of ISO/TC 37/SC 2, is of particular interest to us. This standard was first published in 1973, entitled 'Lexicographical symbols particularly for use in classified defining vocabularies' (ISO 1511, 1973), which, dealing with the variety of codes used in printed dictionaries, gave rise to a revised standard entitled 'Lexicographical symbols and typographical conventions for use in terminography' (ISO 1511, 1997). This second edition in 1997 cancelled and replaced the first edition from 1973. As the title indicates, the original scope of the ISO 1511 (1997) was the harmonisation of the layout of print dictionaries – 'the use of lexicographical symbols and typographical conventions in terminological

¹²⁷ <https://www.iso.org/home.html>

¹²⁸ <https://www.iso.org/committee/48124.html>

¹²⁹ <https://www.iso.org/committee/297592.html>

entries in specialized dictionaries in general, and standardized vocabularies in particular' (ISO 1951, 1997, p. iv) – and 'did not address the actual needs of dictionary making' (Derouin & Le Meur, 2008, p. 754), especially when dictionaries started to be published in electronic format before they became genuinely digital. There was not, in fact, any concern for the structure, reusability and exchange of data.

Then, in 2000, a new revision of this standard began (Derouin & Le Meur, 2002, p. 932) with the sending of a questionnaire to lexicographers, terminological experts, dictionary authors, publishers, terminology departments of industrial companies and national or international bodies. The results of the feasibility study showed that ISO 1951 (1997) did not meet the current needs in lexicography (Derouin & Le Meur, 2002, p. 932), indicating the need for a new standard in the field of lexicography.

The completely revised ISO Standard 1951 (2007) was published again in 2007, covering all lexicographic, monolingual and multilingual products, as well as general and specialised dictionaries. The review process aimed to

a) support the creation and management of various types of dictionaries, b) allow dictionary content to be reused in different and electronic formats, c) facilitate necessary production, exchange and management procedures, d) propose a specific model based on current best professional practices (Derouin & Le Meur, 2008, p. 755).

The ISO 1951 (2007) focused on encoding the representation of lexicographic data in dictionaries via a system called X_mLex¹³⁰ (formerly called LEX_{ml}), an abstract model. This formal model proposed a way of presenting entries in both printed and electronic dictionaries. Following a lemma-oriented approach, the relationship between the formal structure and the presentation of entries used by editors and consulted by users is explained in the examples of XML encoding provided in the informative annexes of this standard. It 'specifies a formal generic structure, independent of the publishing

¹³⁰ Cf. Lexicographical Markup Language, See Report on the Revision of the Lexicographical Standard ISO 1951 Presentation/Representation of Entries in Dictionaries (<http://www.lrec-conf.org/proceedings/lrec2002/pdf/344.pdf>).

media, and an extensible list of constituents ('data elements') (Derouin & Le Meur, 2006, p. 3).

ISO 151 (2007) considers dictionary entries to be 'comments' about 'topics', which are lexical units. Thus, an entry has a main topic (the headword) and may contain other topics (e.g., variants, translations), called 'related topics'. Topics and comments are data elements. Each data element has a content model.

According to ISO 151 (2007), the information contained in each dictionary entry is organised following three mechanisms ('compositional elements'):

(1) *containers* or 'compositional element used to supply additional information about one single specific data element by the means of other elements' (ISO 151, 2007, p. 2) (e.g., a headword container is used for giving the pronunciation);

(2) *blocks* or 'compositional element used to factorize elements that are shared as refiners by many instances of a specific element' (*ibidem*) (e.g., a punctuation such as comma or semicolon to separate meanings, square brackets for contexts);

(3) *groups* or 'compositional element used to aggregate several independent elements' (*ibidem*) (e.g., a sense is described by a group of elements such as definition, usage labels).

ISO 151 (2007) suffers from conceptual deficiencies that have never been corrected. Lemnitzer, Romary and Witt (2013), arguing that 'the next revision of the standard should integrate the TEI tagset as the reference for implementing the proposed model' (p. 19), found that ISO 151 (2007) 'does not actually provide a useful encoding scheme for the representation of print dictionaries' (p. 17), and that this standard should be explored as a starting point to provide 'a real generic model for dictionary representation' (*ibidem*). We also agree with these researchers (Lemnitzer, Romary & Witt, 2013) when they state that 'ISO 151 (2007) suffers from an incomplete design which makes it hardly usable in concrete applications' (p. 19). To the best of our knowledge, few have applied this standard (we know only of isolated cases, such as that of the Langenscheidt publishing house in Munich); at the time of writing this thesis, the possibility of revising this standard is still being debated. We are of the view that, if the

review work proceeds, it should be adjusted considering the efforts involving other recently revised standards, such as the serialisation of the ISO 24613 standard, and above all must reflect current dictionary practices. Furthermore, ISO 24613 supports not only human-readable dictionaries such as the ISO 1951 but also automatic language processing dictionaries intended for use by computer programs. In order to develop a standard that establishes the model for the presentation of entries in dictionaries – a point we consider important in terms of homogeneity and, consequently, interoperability –, it is necessary to carry out an exhaustive study of the variety of layouts for presenting data beforehand, covering a wide range of lexicographic resources. Thus, although we consulted this standard, we have only taken advantage of some definitions and have chosen not to use it to represent our lexical data. There are, however, other ISO standards developed as high-level specifications that are also relevant to our work as follows:

- ISO 639 (ISO 639-1, 2002; ISO 639-2, 1998); ISO 639-3, 2007) provides internationally accepted codes for the representation of names of languages;
- ISO 24613 (ISO 24613-1, 2019; ISO 24613-2, 2020; ISO 24613-3, 2021; ISO 24613-4, 2021; ISO 24613-5, 2021), which will be explored in a subsequent section;
- ISO 1087 (2019), ‘Terminology Work – Vocabulary – Part 1: Theory and Application’, which was used in Chapter 7. This standard defines the fundamental concepts of terminology work, also emphasising the meaning of concept relations and concept systems;
- ISO 704 (2009), ‘Terminology Work – Vocabulary – Principles and Methods. This standard was very useful for our work, as we have seen in Chapter 7. It establishes the basic principles and methods for preparing and compiling terminologies and describes the terminological representation that we adopted in this research.

8.2 Simple Knowledge Organisation System

Simple Knowledge Organisation System¹³¹ (SKOS) is a model for sharing and linking KOS, such as thesauri, taxonomies, classification schemes and other structured and controlled vocabularies available on the web.

SKOS is part of a series of developments and research projects focused on developing and improving web resources at the turn of the millennium (Baker et al., 2013). SKOS answered the need for a common RDF schema for modelling thesauri, a type of knowledge organisation system, and defining inter-vocabulary mappings. It became a W3C recommendation in 2009 (Miles & Bechhofer, 2009). The information science community widely uses SKOS to publish vocabularies on the semantic web. Some notable examples include the European Union's Vocabularies¹³² and the Art & Architecture Thesaurus.¹³³

The model is expressed as an ontology in Web Ontology Language (OWL), which enables the modelling of controlled vocabularies as RDF graphs, as well as their mapping to external resources and integration in the Linguistic Linked Open Data (LLOD) cloud. Among other possibilities, the model allows concepts to be identified with Uniform Resource Identifier (URI), lexicalised with multilingual labels, documented with notes, linked to other concepts through conceptual relationships and mapped to concepts in external sources. Since the core SKOS model only allows for relations between concepts, the SKOS-XL extension has been created to provide support for modelling relations between concept labels. The latter include the relations between abbreviations and their full forms (e.g., between 'EU' and 'European Union').

Some members of the NOVA CLUNL research group are currently working on modelling lexicographic information, specifically focusing on the relationships between abbreviations and their respective complete forms (Costa, Salgado & Almeida, 2021a; 2021b). In the context of the *Digital Edition of the Vocabulário Ortográfico da Língua Portuguesa* (VOLP-1940) project¹³⁴, SKOS allows the modelling of knowledge

¹³¹ <https://www.w3.org/TR/2008/WD-skos-reference-20080125/>

¹³² <https://op.europa.eu/en/web/eu-vocabularies>

¹³³ <https://www.getty.edu/research/tools/vocabularies/aat/>

¹³⁴ <https://clunl.fcsh.unl.pt/en/investigacao/projetos-curso/edicao-digital-do-vocabulario-ortografico-da-lingua-portuguesa-volp-1940/> and <https://www.volp-acl.pt/index.php/vocabulario-1940/projeto>

organisation systems, acting on microstructural information and enabling the connection to other existing systems and resources. The modelling of lexicographic categories and their linguistic realisations (i.e., abbreviations and full forms) in SKOS facilitates the future exploration of VOLP-1940 as linked data. For example, the language category allows a system to extract every entry that has been adopted from another language (e.g., ‘*croché*’ in Portuguese borrowed from the French ‘*crochet*’), which would be an important application for linguistics scholars interested in loanwords and word-formation processes. For interoperability purposes, the lexicographic categories modelled in SKOS should be aligned with external vocabularies and ontologies, such as the widely used LexInfo¹³⁵ ontology of lexical categories. For example, our class for nouns should be mapped to LexInfo’s noun class, which would facilitate the reuse of VOLP-1940’s subset of nouns as linked data (Costa, Salgado & Almeida, 2021a, p. 196).

8.3 OntoLex-Lemon

OntoLex-Lemon was developed by the W3C Ontology-Lexica Community Group (Cimiano, McCrae, & Buitelaar, 2016) based on previous models – in particular the Lexicon Model for ONtologies or *lemon model* (McCrae et al., 2012). This has become widely known for representation of lexical data on the semantic web, including Princeton WordNet¹³⁶ and FrameNet¹³⁷, and has gradually acquired the status of a *de facto* standard according to the principles of linked data.

Concerning the conversion of lexicographic resources into linked data, this model is the preferred choice of many researchers (Klimek & Brümmer, 2015; Declerck et al., 2019; Abromeit et al., 2016; Bosque-Gil et al., 2016a; McCrae et al., 2019).

The *lemon* model was first proposed in 2011 (McCrae, Spohr & Cimiano, 2011). As implied by the name of this model, its aim is not to represent dictionaries but ‘to provide rich linguistic grounding for ontologies’¹³⁸. The emergence of the Linked Open

¹³⁵ <https://www.lexinfo.net/>

¹³⁶ <https://wordnet.princeton.edu/>

¹³⁷ <https://framenet.icsi.berkeley.edu/fndrupal/>

¹³⁸ <https://www.w3.org/community/ontolex/>

Data (LOD) movement, and Linguistic Linked Open Data or LLOD¹³⁹, created the need to represent lexical data as an ontology for the semantic web. After the foundation of the OntoLex community in late 2011, the group took on the improvement of Lexicon Model for Ontologies (Lemon) and its update to create OntoLex-Lemon (McCrae et al., 2017). Since its creation, the OntoLex group has focused more on structuring the model and collecting various usage cases to expand the coverage of Lemon. Simultaneously, the proliferation of tools that make it easier to search, link and visualise such resources has been attractive for many projects. OntoLex-Lemon has been adopted in some lexical resources (e.g., Apertium dictionaries, Forcada et al., 2011; BabelNet, Navigli & Ponzetto, 2012, converted into Lemon-BabelNet, Ehrmann et al., 2014; Global series of K Dictionaries¹⁴⁰, Bosque-Gil et al., 2019).

The OntoLex-Lemon vocabulary is the one used for the publication of lexical data as a knowledge graph. OntoLex-Lemon modelling is based on RDF triplets: subject, predicate and object. The Lexical Markup Framework (ISO 24613-1, 2019) standard also played an important role in defining OntoLex-Lemon – the LMF directly inspired this module in defining lexical entries as the core element of the lexicon.

One of the major issues encountered was that the lexical entry defined in the core module had strict requirements to make it suitable for NLP applications. In particular, it required that each lexical entry had a single lemma, part of speech, morphology, and etymology.

Trying to overcome some OntoLex-Lemon limitations when modelling lexicographic information as LD, the OntoLex community developed the *lexicography module (lexicog)*¹⁴¹, a model to encode existing dictionaries as LLOD. This module operates in combination with the OntoLex core module. This specification was published by the Ontology-Lexica Community Group. Nevertheless, it is not a W3C Standard, nor is it on the W3C Standards Track¹⁴².

¹³⁹ <https://linguistic-lod.org/>

¹⁴⁰ <https://lexicala.com/>

¹⁴¹ <https://www.w3.org/2019/09/lexicog/>

¹⁴² 'Although W3C hosts these conversations, the groups do not necessarily represent the views of the W3C Membership or staff.' See <https://www.w3.org/community/ontolex/>.

The core elements of *lexicog* are `lexicog:LexicographicResource`, `lexicog:Entry` and `lexicog:LexicographicComponent`, along with `lexicog:entry` and `lexicog:describes` properties. The `lexicog:LexicographicResource` class represents a collection of lexicographic entries (`lexicog:Entry`) in accordance with the lexicographic criteria followed in the development of that resource, which are grouped in the dictionary through `lexicog:entry`. Since the correspondence between a dictionary entry and `ontolex:Entry` is not always 1:1 (e.g., a single dictionary entry can describe a lexical unit that assumes different parts of speech and therefore corresponds to more than one `ontolex:LexicalEntry`), this difference also extends to the entire dictionary and distinguishes a `lime:Lexicon` from a `lexicog:LexicographicResource`. An `Entry` is a structural element that represents a lexicographic article or record as it is arranged in a source lexicographic resource.

OntoLex-Lemon was able to open new perspectives for lexical data, offering a structure for their representation on the semantic web and consequently overcoming *ad hoc* serialisation and format problems. However, this initiative has some limitations when it comes to using its scheme as a native format to model lexical structures and their relationships. One such limitation is its triple-based vocabulary's relative complexity compared to standards such as TEI or LMF. Use cases, based mainly on conversion scenarios, support this claim (Klimek & Brümmer, 2015; Bosque-Gil et al., 2016b). Second, since there are no explicit guidelines on how to model certain lexicographic components, such as the relation between a multiword expression and the lemma in a dictionary converted to OntoLex-Lemon, it represents an obstacle to arriving at a unified representation of lexical information that reduces exchange and comparability alternatives. In addition, the standard is under active review, and significant progress has already been made to support the modelling of new, more granular classes of information. However, it remains insufficiently developed to cover the modelling requirements and differences covered in printed dictionaries but has great potential to allow high interoperability and exchange allowed by the semantic web technologies.

8.4 Lexical Markup Framework

The Lexical Markup Framework (LMF) is a *de jure* multi-part standard within ISO 24613 (ISO 24613-1, 2019; ISO 24613-2, 2020; ISO 24613-3, 2021; ISO 24613-4, 2021; ISO 24613-5, 2021) that provides a common standardised structure to have a flexible specification platform for lexical data, mainly for NLP applications with an extension to represent lexical resources and machine-readable dictionaries (MRD). This standard has been prepared and maintained by the technical subcommittee ISO/TC37/SC4/WG4.¹⁴³

The work of the LMF began at the beginning of the 21st century after a series of international projects in the 1990s such as *Acquilex*¹⁴⁴, *Genelex*¹⁴⁵ and *Parole*¹⁴⁶, among others, and was financed by the European Commission. The first group of experts who began developing the LMF aimed to design a general structure based on the general characteristics of the existing lexicons and to develop a consistent terminology that described each component of these lexicons, thereby generating a comprehensive model that better represented all lexicons and their respective components.

The LMF specification was first presented to the scholarly community through an article (Francopoulo et al., 2006) and was officially published as ISO 24613 in 2008. Subsequently, a book entirely dedicated to the LMF was published (Francopoulo, 2013).

The metamodel is structured in two sections – a core package that represents the basic information in a lexical entry and interlinked extension packages, ‘which are expressed in a framework that describes the reuse of the core components in conjunction with the additional components required for a specific lexical resource’ (Francopoulo et al., 2006, p. 234) for the representation of the MRD.

The original ISO 24613 (2008) standard was conceived as an abstract metamodel providing a standardised framework for the construction of computational lexica. At the time this thesis was written, the original LMF standard had been under review since 2016 and was subdivided into several parts (Romary et al., 2019): the Core Model (ISO

¹⁴³ Committee on Language Resource Management and working group Lexical Resources: <https://www.iso.org/committee/297592.html>

¹⁴⁴ <https://www.cl.cam.ac.uk/research/nl/acquilex/>

¹⁴⁵ <http://www.ilc.cnr.it/EAGLES96/lexarch/node15.html>

¹⁴⁶ <https://cordis.europa.eu/project/id/LE24017>

24613-1, 2019); the machine-readable dictionary (MRD) model (ISO 24613-2, 2020); the Etymological extension (ISO 24613-3, 2021), which aims to make both standards interoperable and fully compatible; the TEI serialisation (ISO 24613-4, 2020), describing the serialisation of the LMF standard defined as an XML model compliant with the TEI guidelines; and, finally, Lexical base exchange (LBX) serialisation, a W3C XML serialisation for MRD (ISO/DIS 24613-5, 2020). The main objective of this new version is to create a more modular, flexible and durable model. These standards also include definitions of terms applicable to our research.

The metamodel for lexical entries provides an abstract representation format for lexical information. LMF is a native UML framework. The standard provides guidelines on how to convert the UML model to the XML schema for lexica.

The main classes are `Lexical Resource`, `Global Information`, `Lexicon`, `Lexical Entry`, `Lemma` and `Word Form`. The UML specification can be somewhat abstract for the lexicographer community, who are not used to formalising data structures.

`Lexical Entry` is the key class of any LMF metamodel and is the backbone of the lexical description. Morphological and semantic information are presented through the classes `Form` and `Sense` as well as their subclasses.

The `List Of Components` and their classes, belonging to several overlapping extensions, represent the main modelling mechanism. The TEI serialisation directly maps the class names in the metamodel and XML elements' names or attributes.

Under the general notion of `Word Form`, the LMF gathers information that documents, classifies or structures a lexical unit's written or spoken representation. The `Sense` component is organised as a fully iterative and recursive structure, which, according to the scope of the actual lexical database to be implemented, can be further characterised by any number of restrictions, such as register, usage, grammatical or syntactic variants, collocations or translations.

8.5 Text Encoding Initiative

The Text Encoding Initiative (TEI) (Sperberg-McQueen & Burnard, 1994) is now the recognised international *de facto* standard for the digital representation of textual resources (ranging from books and manuscripts to mathematical formulae, culinary recipes, music notation, among many other types) in the scholarly research community. Despite not having the legal status of a standard (Stührenberg, 2012), it is widely used by the scholarly research community in the humanities, particularly by the lexicographic community in several dictionary projects for digitally-created lexicographic data (e.g., Budin, Majewski & Mörrth, 2012) or retro-digitised projects (e.g., Bohbot et al., 2018).

The TEI is the basis for many current lexicographic projects, such as BASNUM¹⁴⁷, Nénufar¹⁴⁸, ARTFL¹⁴⁹, VICAV¹⁵⁰ or the Berlin-Brandenburg Academy of Sciences project for digitising and transcribing legacy dictionaries¹⁵¹, VOLP-1940 (Salgado & Costa, 2020) and MORDigital (Costa et al., 2021c), to cite a few. Although the original target audience was the academic community, libraries and publishers, among other organisations, have also used the TEI.

The TEI was created in 1987 by a consortium of several institutions, known as the TEI Consortium, to develop a standardised format for the electronic edition of textual content in multiple formats. The TEI Consortium is responsible for the continuous development of *P5: Guidelines for Electronic Text Encoding and Interchange*.¹⁵²

The TEI Guidelines comprise comprehensive documentation and define a markup language to represent structural and conceptual characteristics of text documents. Their first draft (P1) was published in 1990, and the current version, P5, was published in 2007 and has been updated regularly. When writing this thesis, the Guidelines version 4.3.0 were made available on 31 August 2021 and will continue to be subject to constant updates. Many different individuals are responsible for the maintenance and development of the Guidelines. This interaction is mainly carried out

¹⁴⁷ <https://anr.fr/Project-ANR-18-CE38-0003>

¹⁴⁸ <http://nenufar.huma-num.fr/>

¹⁴⁹ <https://artfl-project.uchicago.edu/>

¹⁵⁰ <https://vicav.acdh.oeaw.ac.at/>

¹⁵¹ <https://gitlab.com/xlhrlld/retro-dict>

¹⁵² <https://tei-c.org/Vault/P5/4.2.1/doc/tei-p5-doc/en/html/>

through the TEI-L mailing list,¹⁵³ whose response has always been swift and, in many cases, exhaustive. Also, any bugs or inconsistencies can be reported to the TEI community on their mailing list or via GitHub Issues.¹⁵⁴ Furthermore, the guidelines are available under the open Creative Commons BY licence from the TEI website¹⁵⁵ and also from GitHub.¹⁵⁶

Initially, the Standard Generalised Markup Language (SGML) was used to encode documents. However, after the widespread adoption of XML, the P4 version of the TEI Guidelines, published in 2002, switched to the new encoding language. These guidelines are developed as a modular and extensible XML schema, which make it possible to convert TEI-encoded texts into various other document formats. The specification language in which the TEI is defined and which is used to express a customisation of the TEI scheme is ‘one document does it all’ (ODD), i.e., the schema should contain documentation favouring its flexible nature.

Although it is a text metamodel, the TEI is based on XML and essentially defines several hundred elements and their respective attributes. As a metalanguage, the TEI provides a vocabulary (a set of elements and attributes) and a grammar (a schema) that can be used to describe, structure and validate data. Its specific XML syntax and semantics make it a method of textual analysis for digital processing.

The guidelines provide the formal modelling of text in documents through categories that bring together related XML elements called modules. The P5 version of the standard comprises 21 modules, three to mark almost any text, with the ninth being dedicated to dictionary encoding. In our work, we chose to follow this standardised format for several reasons. First, it is commonly used for digital editing and digital preservation of documents. Second, it has a specific module for dictionaries, which applies to the encoding of various lexical resources provided in Chapter 9 (‘Dictionaries’)¹⁵⁷ of the TEI Guidelines.

¹⁵³ <https://tei-c.org/support/>

¹⁵⁴ <https://github.com/DARIAH-ERIC/lexicalresources/projects/1>

¹⁵⁵ <https://tei-c.org/>

¹⁵⁶ <https://github.com/TEIC/TEI>

¹⁵⁷ <https://tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>

All TEI documents must include a metadata section, named *TEI header*, and share a set of common annotation features defined as the core module in the standard (Chapter 3).¹⁵⁸ This set includes structural elements such as paragraphs, lists or bibliographical references.

TEI is a descriptive recommendation that does not enforce a single form to encode a specific document. This extreme flexibility – this is the characteristic that perhaps justifies its wide adoption – of the TEI Guidelines in structuring data offers the possibility of several types of encoding for the same components. Its ‘one document does it all’ (ODD) specification language also underlines its adaptability to any new requirement. However, for interoperable reasons the flexibility offered needs to be restricted. For instance, to create cross-references, the preferred way is to use the `<xr>` tag. Nevertheless, it is also possible to create links using `<anchor>`, `<ptr>` or `<link>`. On the other hand, ODD attribute lists are of three types – (i) ‘closed’ (only the values defined in `<valList>` are permitted); (ii) ‘semi-open’ (the values defined in `<valList>` are treated as suggested values, but others are allowed); and (iii) ‘open’ (any value is allowed; values in `<valList>` are treated as mere examples). The semi-open and open lists present particular problems for interoperability. Looking at the current statistics at the time of writing this thesis, there is still much work to be done to arrive at a set of values that are as closed as possible.¹⁵⁹ In some cases, TEI makes no binding requirements for the possible values since there are many possibilities across different projects. However, in a given lexicographic project, it is likely that standardising an agreed set of values will be very helpful. In this sense, it is better to customise or change the schema by providing more restrictions. This explains why, for example, there is the need to restrict the scope of usage information (Salgado et al., 2019). Then, it will be necessary to know what they actually cover and signify and ensure that all the documents use only the agreed-upon set of values.

¹⁵⁸ <https://tei-c.org/Vault/P5/1.3.0/doc/tei-p5-doc/es/html/CO.html>

¹⁵⁹ Statistics kindly provided by Laurent Romary (TEI All, 22 August 2021): Number of defined attributes (attDef): 535. Number of value list in defined attributes (attDef/valList): 146. Statistics on the type attribute of attDef/valList: semi: 40; open: 50; closed: 55 The one that is not specified is @break in att.breaking, but ‘open’ is the default for valList/@type.

8.5.1 The TEI Dictionary Module

From the very beginning, the TEI Guidelines have had a module explicitly focused on the encoding of dictionaries¹⁶⁰. For dictionaries, Chapter 9¹⁶¹ of the TEI Guidelines starts by defining the dictionary structure as a book, namely front matter, body or back matter. The elements defined in this module are mainly intended to encode human-oriented dictionaries but can also help encoding computational lexicons.

The TEI Guidelines provide solutions to encoding the original layout of a dictionary page, i.e., how the entries are organised visually (typographic view), the properties of a text modelled as a sequence of tokens (editorial view) and the underlying lexical structures concerned with the conceptual or linguistic content of a dictionary (lexical view), revealing a more abstract and focused perspective for dealing with linguistic content (Ide & Véronis, 1995). In Chapter 9, we explore these different views of modelling (section 9.1, pp. 276–277). The first level of text encoding aims to reflect the physical structure of a document, whereas the second level deals with text structures.

However, the overall flexibility, which raises the possibility of individual lexicographic phenomena being encoded in multiple ways, has been a cause of concern from the point of view of interoperability (Salgado et al., 2019). This freedom, together with its widespread adoption among lexicographers who have their own background and views on the logical structure of a dictionary, has produced an array of encoding solutions. Ironically, a standard that was supposed to unify the encoding formats under the umbrella of a common structure may sometimes appear as an uncontrolled modelling space. Flexibility, therefore, is both a virtue and a shortcoming of TEI. To reduce this freedom and define a specific format for dictionaries, forcing dictionary encoders to follow the same structural rules, the lexicographic and dictionary-encoding communities are currently discussing a new format with a particular focus on retro-digitised dictionaries. This is known as TEI Lex-0 (Tasovac, Romary et al., 2018; Romary

¹⁶⁰ Here, the word ‘dictionaries’ is taken in its most general sense, i.e., encompassing not only dictionaries but also other types of lexical resources.

¹⁶¹ <https://tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>

& Tasovac, 2018; Bański, Bowers & Erjavec, 2017), which is a TEI-compliant but streamlined format to facilitate interoperability.

8.5.2 The TEI Lex-0

TEI Lex-0¹⁶² is a stricter subset of TEI that aims for a stricter TEI representation of heterogeneous TEI-based lexical resources. Its goal is to establish a baseline encoding and a target format to facilitate the interoperability of heterogeneously encoded lexical resources. Some of the experiments on TEI Lex-0 in digital lexical databases can already be referred to, e.g., the studies by Bohbot et al. (2019), Bowers et al. (2019), Khan et al. (2020) and Salgado et al. (2019).

In the context of the ELEXIS project, TEI Lex-0 has been adopted, together with OntoLex, as one of the baseline formats for the ELEXIS infrastructure (McCrae et al., 2019). As the layout of this format is not yet closed, we have been actively contributing to its development by creating issues on GitHub.¹⁶³

TEI Lex-0 was launched in 2016, and it is led by the DARIAH Working Group on Lexical Resources¹⁶⁴, made up of experts in lexical resources. Its goal is to define a clear and versatile annotation structure, but one that is not too permissive, to facilitate the interoperability of heterogeneously encoded lexical resources. The TEI Lex-0 should not be seen as a replacement for the TEI Dictionary Module. It should be first considered as ‘format that existing TEI dictionaries can be unequivocally transformed to in order to be queried, visualised, or mined uniformly’ (Tasovac, Romary et al., 2018, para. 3). While the TEI Lex-0 is being developed, some of its best-practice recommendations are also changing the recommendations of the TEI Guidelines themselves.

The TEI Lex-0 imposes different types of restrictions compared with the TEI, as follows:

- It reduces the number of elements available (e.g., the TEI Lex-0 uses only <entry>, while TEI has several elements for the basic microstructure unit of

¹⁶² <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>

¹⁶³ <https://github.com/DARIAH-ERIC/lexicalresources/projects/1>

¹⁶⁴ <https://www.dariah.eu/activities/working-groups/lexical-resources/>

the dictionary). These are <entry> (a single structured entry in any kind of lexical resource), <entryFree> (a single structured entry), <superEntry> (a single unstructured¹⁶⁵ entry), <re> (an entry related to a lemma within an entry), and <hom> (homograph within an entry). While the document precisely describes when each should be used (entry forces a structure) (Bański, Bowers & Erjavec, 2017), <entryFree> provides a flat representation and allows unstructured entries that should be avoided but may be necessary for some dictionaries, and <superEntry> is a mechanism that can group other entries, such as homonyms). This freedom makes it difficult for different authors to keep their dictionaries coherent in terms of structure.

- It makes certain attribute values required (e.g., `xml:lang` and `xml:id` in <entry>).
- It reduces the number of possible attribute values on specific elements (such as <usg>).
- It applies additional syntactic constraints (e.g., <def> can only appear within a <sense>) or, when necessary, allows for new syntactic constructions.

We decided to use this new and stricter subset of the TEI Guidelines for its interoperability since it has been extensively tested in some Portuguese projects with good results (Costa et al., 2021b; Salgado et al., 2019), and arguing that a simplified array of elements can lead to a more coherent and legible encoding without sacrificing its semantic expressivity, and whose application will be detailed in Chapter 9.

¹⁶⁵ Unstructured means that an element can appear anywhere within any entry level. The elements are provided to support much wider variation.

CHAPTER 9

TEI Lex-0 in action

TEI Lex-0 aims at establishing a target format to facilitate the interoperability of heterogeneously encoded lexical resources.

ROMARY & TASOVAC (2018)

This chapter discusses the encoding of terms in general language dictionaries using TEI Lex-0, a customised version of TEI for lexicographic datasets.

The application of TEI Lex-0 will be demonstrated with samples of some selected terms from the DLP (soon to be made available online) as a case study to present our ongoing work related to the encoding of terms. We try, whenever possible, to select examples from the domains under study. However, whenever we intend to illustrate a particular feature that is important for the encoding of terms and we could not find examples of these domains, we exemplify the observations with terms belonging to other fields.

The goal of this chapter is threefold: (1) to illustrate how existing TEI Lex-0 specifications can be used in an actual dictionary project to consistently mark up different microstructural components, including simple domain labels; (2) to show how the currently recommended TEI Lex-0 practice for representing domain labels as flat values is not robust enough to deal with more complex, hierarchical domain structures; and (3) to propose alternative ways of encoding taxonomies of domain labels in TEI Lex-0 as our contribution to the development of this community standard. In other words, the goal of this chapter is to translate the conceptual work we have done in previous chapters into a practical means of implementation using TEI Lex-0.

Throughout this chapter, one cannot forget that dictionary encoders work on formal representations of the actual lexicographic content of existing dictionaries. The discussion will be from the point of view of lexicographic data modelling, i.e., the process of explicitly marking up the structural hierarchies and the scope of particular textual elements from existing dictionary entries to convert them to an electronic format as part

of a lexicographic digitisation workflow (Tasovac & Petrović, 2015). The encoding of a simple dictionary entry will be presented first to highlight its main aspects: the basic structure of an entry, the domain label and the defining hierarchy as it is relevant to this research and the formal representation of polylexical terms. We will also present examples of entries with related entries and demonstrate the difference between the encoding of usage examples (whether extracted from a corpus or created by the lexicographer) and illustrative quotations (taken from books or periodicals). A comprehensive encoding of the analysed terms in this thesis are available in a GitHub¹⁶⁶ repository.

We adopt some typographic specifications for use throughout this chapter. TEI P5 terms (element names, attribute names, attribute values, etc.) are written in a fixed-width (monospace) font and:

- for individual element names, we surrounded the name of the element with angle brackets (<entry>);
- for the names of nested elements, we used the XPath notation¹⁶⁷, e.g., (cit/quote/bibl);
- for attribute names, we used the @ sign before the name of the attribute, e.g., @type;
- for attribute values, we surrounded the string with quotation marks ("), e.g., "domain".

9.1 Different Views of Modelling

There are two main approaches that we can take in modelling lexicographic resources in general and retrodigitised dictionaries in particular. We can view a dictionary primarily as a textual artefact with its own specific publishing history and its own verbal expression and visual arrangement of the linguistic content contained within it or we can instead prioritise the linguistic content, ignoring how it is presented and the exact sequence of words used in, for example, the definitions of articles. There is also a

¹⁶⁶ https://github.com/anacastrosalgado/DLP/tree/master/PhD_work

¹⁶⁷ <https://www.w3.org/TR/xpath-31/>

third (potentially more verbose) approach, that is, to do both simultaneously and make sure both kinds of information are aligned.

The distinction made in Chapter 9 ('Dictionaries') of the TEI Guidelines between the typographical, editorial and lexical views of a dictionary is handy in this discussion (Figure 122).

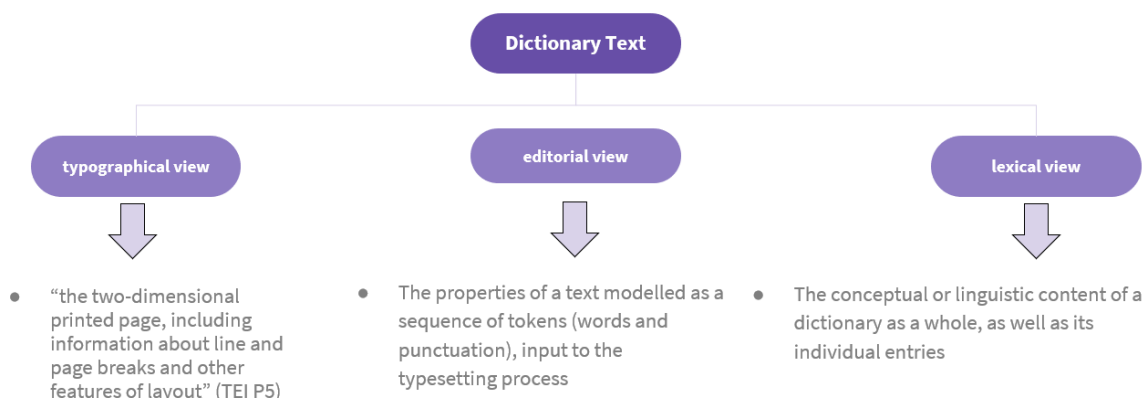


Figure 122: *Different Views on Lexicographic Resources (Khan & Salgado, 2021)*

These views are defined as follows: the typographical view aims to mirror the physical structure of a document using elements from the core module. It concerns the layout of individual pages. These TEI elements can be used to encode the page layout, column and line breaks and highlighted words. Some elements can also be typed to provide more precision on how they are typographically presented in the original printed document. The second level of encoding deals with the semantic and logical function of text structures – the sequential arrangement of individual tokens along with the use of specific font styles, punctuation and special characters. The editorial view is concerned with the properties of a text modelled as a sequence of tokens, and the lexical view is concerned with the conceptual or linguistic content of a dictionary as a whole as well as its individual entries. With the typographical view, we would be interested in, e.g., the position of line breaks in a text or the visual arrangement of entries on any single page, with the editorial view, we are interested in such things as which words are used in the description of the article and in which order. Finally, in the lexical view we are interested in information about the given domain of a term, or that, for instance, a

given headword is a ‘noun’. In addition to these views, TEI also offers extensive provision in the <teiHeader> element for including metadata about an original resource to be modelled (e.g., who the authors were and when it was published) and about the process of its digitisation as well as the creation of that process. For lexical encoding, the dictionary module provides a lexicon designer with an exhaustive set of TEI elements that model different linguistic levels of lexical information. This encoding is also influenced by the linear description of the printed dictionaries. Common practices involve respecting the order of the fields as they appear in the original document.

9.2 The DLPC and DLP as a TEI Dictionary Projects

As previously mentioned, the first complete edition of the DLPC was published in 2001 in a two-volume paper version. The PDF version of the ACL print dictionary was later converted into XML using a customised version of the P5 schema of the Text Encoding Initiative (TEI). A custom-built dictionary writing system – LeXmart¹⁶⁸ – was developed to allow editing and creation of new dictionary entries and validation of their structure and overall dictionary coherence. The DLPC originated the DLP, which is currently being converted to the TEI Lex-0 format for data interoperability purposes (Simões et al., 2019).

The original TEI P5 schema was used as the target format. Some specific standard constructions had to be changed to enable encoding some of the dictionary entries. This process was iterative and interactive, with human intervention needed to fix minor issues in some entries for which the default behaviour could not correctly determine the entry structure. To allow quick editing of the database, the TEI dictionary was split into thousands of small XML documents (one per dictionary entry) that were imported into a native XML database (eXist-DB)¹⁶⁹.

Although we followed TEI in DLPC encoding, the following reasons led us to investigate TEI Lex-0: (1) (1) we had to adapt the standard features because we could not find solutions in the TEI Guidelines that covered all the microstructural components

¹⁶⁸ <http://www.lexmart.eu/>

¹⁶⁹ <http://exist-db.org/exist/apps/homepage/index.html>

of the dictionary, e.g., the entry ‘*a*’, as a preposition, and different types and levels of information: grammatical, semantic and pragmatic (Simões et al., 2019) to indicate the prepositions values; (2) the extreme flexibility of TEI Guidelines (the multiple solutions to encode the same type of information) raised many questions when we were making some decisions in terms of dictionary content reusability.

We found some advantages to the strictness of TEI Lex-0 which can potentially facilitate data exchange and mutual alignment across dictionaries. Discussions with the editors of this format have been fruitful in finding linguistically and structurally valid encoding solutions. The retrodigitised version of DLPC was imported to a database for future archive reference, but it is not being edited. The database was cloned to update the lexicographic articles, thus came the DLP. We are now converting the DLP into TEI Lex-0 encoding, mainly because it allows us to encode the full extent of the dictionary structure without customising the schema ourselves. Therefore, we present some experiments on the encoding of specific parts of the dictionary entries. Our immediate goal is not to have the dictionary only in TEI Lex-0 but to keep the original version in our interpretation of TEI and have another version that can be used for tests and promote the discussion with the TEI Lex-0 community.

Also, as the entries are stored independently in the XML database, our goal is not to produce a complete XML document for the dictionary but a set of small XML files per dictionary entry. Therefore, details about the *TEI header* are deliberately ignored at this stage, and we are not using the complete schema but only the entry portion, considering the <entry> tag as the document root element. In the future, the <teiHeader> can be stored in an independent record in the database, and a simple tool can be used to construct a TEI/TEI Lex-0 file with the complete dictionary, validating the complete schema.

Figure 123 presents a list of some of the essential changes between our DLPC original encoding and DLP conversion into TEI Lex-0:

XML ESSENTIAL CHANGES (some examples)	
Original encoding	Conversion into TEI Lex-0
1. <entry>	<entry xml:id=" id " xml:lang="pt">
2. <gramGrp> part of speech and gender </gramGrp>	<gramGrp><gram type="pos"></gram> <Gram type="gen">f.</gram></gramGrp>
3. <sense>	<sense xml:id=" id ">
4. <quote type="example">	<cit type="example"><quote>
5. <syn> synonym </syn>	<xr type="synonym"><ref type="entry sense"> synonym </ref></xr>
6. <cit><quote> example </quote> <bibl> author ,<title> title </title>,<page> </bibl></cit>	<cit type="example"><quote> example </quote> <bibl><author> author </author> <title> title </title><citedRange> page </citedRange></bibl></cit>
7. <sense><def>V.<xr><ref> cross-reference </ref></xr></def></sense>	<xr><lbl>V.</lbl><ref> cross-reference </ref></xr>

Figure 123: XML Essential Changes – DLCP Original Encoding and DLP Conversion into TEI Lex-0

As shown in Figure 123, we should highlight that in line 1 of TEI Lex-0, the <entry> element requires the attributes @xml:id¹⁷⁰, the unique entry identifier for the element bearing the ID value, and @xml:lang, the appropriate language content code according to IETF BCP 47¹⁷¹ ('pt' for Portuguese), which in turn is based on ISO 639.¹⁷² In terms of the POS (line 2), this grammatical information was initially encoded using only the <gramGrp> element for the part of speech (POS) and gender. In TEI Lex-0, morphosyntactic information is encoded in a typed <gram> element, including the POS of the entry and further specifications, such as the gender and number. The examples (line 4) are now encoded in <cit> (cited quotation) and @type within the "quote" value. We customised our TEI schema to allow <syn> (line 5) in the original encoding, but it was not a TEI element, which at the time we thought was better for the encoders. Instead of using <xr> (line 7), the element used in the TEI Guidelines to refer to entries defined in another entry, we left the cross-reference in <def>. In TEI Lex-0, we switched to the semantically more correct nesting of <xr> (cross-reference container) and <ref> for the actual pointer to a different entry.

9.2.1 Basic Structure of an Entry

A lexicographic article, or an <entry> element, always starts with a lemma (or canonical form). The lemma is encoded using the <form> element with the @type

¹⁷⁰ The XML standard does not allow the use of accented characters in element identifiers.

¹⁷¹ <https://tools.ietf.org/html/bcp47>

¹⁷² <https://www.iso.org/iso-639-language-codes.html>

attribute and the value "lemma". The <orth> element (orthographic form) gives the orthographic form of the lemma, i.e., the written form *per se*. After this information, a basic entry structure can include several elements, such as phonetics (<pron>), grammatical information (<gramGrp>), etymology (<etym>) and meaning (<sense>), as shown below.

```
<entry xml:id="..." xml:lang="pt" type="...">
  <form type="lemma">
    <orth>...</orth>
  </form>
  <gramGrp>
    <gram type="pos"/>
    <gram type="gen"/>
  </gramGrp>
  <sense>
    [...]
  </sense>
</entry>
```

Figure 124 shows an example of a lexicographic article from the DLPC.

cristalografia [kri/ʔelugrefiʔ]. *s. f.* (Do gr. κρύσταλλος 'cristal' + suf. *-grafia*). *Miner.* Ciência que estuda e descreve a forma e a estrutura dos cristais, bem como as leis que regem a sua formação.

Figure 124: Entry 'cristalografia' [crystallography] in the DLPC (ACL)

In the DLPC, the monolexical term "cristalografia" [crystallography], as shown in Figure 124 above (Figure 62 in Chapter 6), has some traditional typographic features, such as the headword (the orthographic form in bold typeface), phonetic transcription (phonetics in square brackets), grammatical information (*s. f.* [feminine noun], the POS followed by the gender, both in italics and abbreviated to save space), etymology (etymological information in round brackets), and finally the definition (meaning). In the DLP, we have the same structure, but as some phonetic transcriptions were lost during

the data conversion task¹⁷³, the new dictionary will probably not include this information in the first moment.

Figure 125 shows the same revised entry in the DLP.

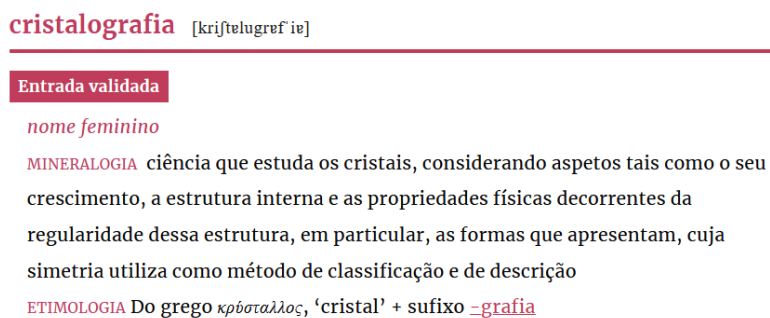


Figure 125: Entry 'cristalografia' [crystallography] in the DLP (ACL)

Comparing Figures 124 and 125 helps reveal some structured changes, namely that we are now using the designation '*nome*' [noun] (and not *s.* or a substantive) when the POS was expanded. The domain label, *Miner.*, appears abbreviated in the DLPC but appears in its full form in the DLP (MINERALOGIA [mineralogy]). The etymology will appear at the end of the lexicographic article introduced by a delimiter, 'ETIMOLOGIA' [etymology], which will be automatically processed. Next, the encoding of this updated dictionary entry is shown below for the sake of context.

The core elements of this dictionary entry will be described in the following subsections.

9.2.2 Macrostructural Level

The outermost structural level of an entry consists of the <entry> element that includes all of the information about the lemma, that is, the <form> element, information on the written and spoken forms related to the description of its spelling and phonetics.

¹⁷³ We lost non-IPA/Greek characters (Simões, Almeida & Salgado, 2016).

The different types of entries are currently marked with the @type attribute in the <entry> element. In Salgado et al. (2019a), we analysed the different types of lexical units that can be headwords. This classification is replicated in this work: monolexical units and polylexical units, affixes and abbreviations (see Chapter 5, p. 131). In the encoding of “cristalografia”, this term is classified as a monolexical term (<entry type="monolexicalUnit"[...]>). This annotation will not be explicit, i.e., the information will not be visible to the end user, but by adopting this classification, we will be able to automatically locate and distinguish the type of lexical units and also extract statistical information. As we have already seen above, the <form> element specifies its attribute value as "lemma" and the orthographic form is given in the <orth> element. Concerning phonetic transcription, this information is given in the <pron> element.

One of the main features of the DLPC, which differentiates it from other contemporary Portuguese dictionaries (e.g., GDLP; HOUAISS), is the treatment of POS homonyms. Homonyms of the same etymological family belonging to different parts of speech are described separately in individual entries and differentiated by numeric superscripts to the right of the lemma (e.g., “paleozóico”¹, *adj.*, “paleozoico”², *s. m.*) as an adjective and as a noun. According to the editors in the Introduction, splitting entries ‘*justifica-se por razões de natureza semântica, morfológica e sintáctica*’ (is justified for semantic, morphological and syntactic reasons) (DLPC, p. xvii). In Figure 126, this point will be illustrated.

paleozóico¹, **a** [paljɔzɔjku, -v]. *adj.* (De *paleo-* + gr. ξῶον ‘animal’ + suf. *-ico*). Que é relativo à era primária ou ao período geológico do Paleozóico. ≈ PRIMÁRIO.
Paleozóico² [paljɔzɔjku]. *s. m.* (De *paleo-* + gr. ξῶον ‘animal’ + suf. *-ico*). *Geol.* Divisão cronológica da história da Terra, anterior ao Mesozóico, que abarca os primeiros 345 milhões de anos do éon fanerozóico, compreendendo os períodos Câmbrico, Ordovícico, Silúrico, Devónico, Carbónico e Pérmico, que se caracteriza por uma grande diversificação da fauna, com o desenvolvimento dos invertebrados e o aparecimento dos primeiros peixes, batráquios, insectos e répteis. ≈ ERA PRIMÁRIA, PRIMÁRIO.

Figure 126: Entry ‘paleozóico’ [palaeozoic] in the DLPC (ACL)

```

<entry type="monolexicalUnit" xml:lang="pt" xml:id="DLPC.paleozoico_1"
n="1">
  <form type="lemma">
    <orth>paleozóico</orth>
    <pron>paljɔ'zɔjku</pron>
  </form>
  <form type="inflected">
    <orth>paleozóico</orth>
    <gramGrp>
      <gram type="gen">m.</gram>
    </gramGrp>
  </form>
  <form type="inflected">
    <orth>paleozóica</orth>
    <gramGrp>
      <gram type="gen">f.</gram>
    </gramGrp>
    <pron>paljɔ'zɔjke</pron>
  </form>
  <gramGrp>
    <gram type="pos" norm="ADJ">adj.</gram>
  </gramGrp>
  <!--etc.-->
  <sense xml:id="DLPC.paleozoico_1_1">
    <def>que é relativo à era primária ou ao período geológico do
Paleozóico</def>
    <xr type="synonymy">
      <ref type="sense">primário</ref>
    </xr>
  </sense>
</entry>

```

There are two structural descriptions for entries: flat and nested entries. We should highlight that the TEI Lex-O schema only uses the <entry> element; once we have constrained the general structure of a lexical entry, in our schema, <entryFree>, <superEntry> and <re> (related entry) from the current TEI Guidelines are not used.

This example shows that TEI Lex-O adopts a constructive approach, making the encoding more structured and verbose, thus facilitating machine processing. The inflected forms (masculine and feminine) are encoded in the <form> element using the @type with "inflected" value. In this case, the grammatical information specific to each inflected form is embedded in the <form> element.

This example (Figure 126) shows yet another detail regarding visual information. As there is more than one entry for the term “paleozóico” (palaeozoic), dictionaries usually include a superscript number next to the lemma to differentiate each headword. We decided to encode the number as the attribute @n (number) in the <entry> element.

In DLP, we changed the criterion for treating POS homonyms so that we now just have one entry with two different POS (Figure 127).

paleozoico, a [paljɔˈzɔjku, -ɐ] ou **paleozóico, a** ^{PRÉ-AO}

Entrada validada

adjetivo

GEOLOGIA relativo ou pertencente ao Paleozoico

SINÓNIMOS *primário*

nome masculino

1. **GEOLOGIA** designação da era inferior (unidade geocronológica) que integra o Fanerozoico, correspondendo ao intervalo de tempo geológico, equivalente a aproximadamente 541 a 250 milhões de anos, durante o qual um conjunto de rochas desse eratema se formou

SINÓNIMOS *primário*

2. **GEOLOGIA** designação do eratema (unidade cronostratigráfica), que integra o Fanerozoico, posicionado acima do Precâmbrico e abaixo do Mesozoico, durante o qual um conjunto de rochas se formou durante a era respetiva

ETIMOLOGIA De *paleo-* + grego *ζῷον*, ‘animal’ + sufixo *-ico*

1) O sistema paleozoico integra o eonotema fanerozoico. 2) O sistema paleozoico integra as séries: Câmbrico, Ordovícico, Silúrico, Devónico, Carbonífero, Pérmico.

Tt Como nome, escreve-se com inicial maiúscula.

Figure 127: Entry ‘paleozoico’ [palaeozoic] in the DLP (ACL)

```
<entry type="monolexicalUnit" xml:lang="pt" xml:id="DLP.paleozoico">
  <form type="lemma">
    <orth>paleozoico</orth>
    <pron>paljɔˈzɔjku</pron>
  </form>
  <form type="inflected">
    <orth>paleozoico</orth>
    <gramGrp>
      <gram type="gen">m.</gram>
    </gramGrp>
  </form>
  <form type="inflected">
    <orth>paleozoica</orth>
    <gramGrp>
      <gram type="gen">f.</gram>
    </gramGrp>
    <pron>paljɔˈzɔjke</pron>
  </form>
  <gramGrp>
    <gram type="pos" norm="ADJ">adj.</gram>
  </gramGrp>
  <sense xml:id="DLPC.paleozoico_1">
    <usg type="domain" corresp="#domain.earth_sciences.geology.stratigraphy"/>
    <def>relativo ou pertencente ao Paleozoico</def>
    <xr type="synonymy">
      <ref type="sense">primário</ref>
    </xr>
  </sense>
  <gramGrp>
    <gram type="pos" norm="NOUN">nome</gram>
  </gramGrp>
</entry>
```

<gram type="gen">masculino</gram>
 </gramGrp>
 <sense n="1" xml:id="DLP-paleozoico_2">
 <usg type="domain" corresp="#domain.earth_sciences.geology.stratigraphy"/>
 <def>designação do eratema inferior (<xr>
 <ref type="entry">cronostratigráfica</ref>
 </xr>) do eonotema Fanerozoico, correspondente ao conjunto de rochas
 formadas durante a
 era respetiva (<xr>
 <ref type="entry">unidade geocronológica</ref>
 </xr>)</def>
 <xr type="synonymy">
 <ref type="entry">primário</ref>
 </xr>
 </sense>
 <sense n="2" xml:id="DLP-paleozoico_3">
 <usg type="domain" corresp="#domain.earth_sciences.geology.stratigraphy"/>
 <def>designação da era inicial (<xr>
 <ref type="entry">unidade geocronológica</ref>
 </xr>) do eón Fanerozoico, correspondente ao intervalo de tempo durante o
 qual se
 formaram as rochas do respetivo eratema (<xr>
 <ref type="entry">cronostratigráfica</ref>
 </xr>), entre 541 e 251 milhões de anos</def>
 <xr type="synonymy">
 <ref type="entry">primário</ref>
 </xr>
 </sense>
 <etym>
 <etym type="grammaticalization">
 <seg type="desc">De</seg>
 <cit type="etymon">
 <form>
 <orth extent="pref">paleo-</orth>
 </form>
 </cit>
 </etym>
 <etym type="inheritance">
 <seg type="desc">grego</seg>
 <cit type="etymon" xml:lang="grc">
 <pc>'</pc>
 <gloss>animal</gloss>
 <pc>'</pc>
 </cit>
 </etym>
 <etym type="grammaticalization">
 <seg type="desc">sufixo</seg>
 <cit type="etymon">
 <form>
 <orth extent="pref">-ico</orth>
 </form>
 </cit>
 </etym>
 </etym>
 <note type="encyclopedic">O sistema/período paleozoico integra as séries/épocas:
 Câmbrico,
 Ordovícico, Silúrico, Devónico, Carbonífero e Pérmico.</note>
 <note type="case">Como nome, escreve-se com inicial maiúscula.</note>
 </entry>

Finally, we want to say a word about the spelling issue. DLP will reflect the new writing rules imposed by the Portuguese Language Orthographic Agreement of 1990¹⁷⁴. Returning to the previous example, “paleozoico” (Figure 127), the spelling of this word according to the new rules¹⁷⁵ no longer has an accent, but we have the old form with accent annotated. If the user looks up old forms, they will find the new lexical forms even if written in the search box without applying the new rules. The result we want to achieve (cf. Bański, Bowers, & Erjavec, 2017) can be seen below.

```
<entry type="monolexicalUnit" xml:lang="pt" xml:id="DLPC.paleozoico_1">
  <form type="lemma">
    <orth>paleozoico</orth>
    <pron>paljɔ'zɔjku</pron>
  </form>
  <form type="variant">
    <orth notAfter="1990" xml:lang="pt-PT">paleozóico</orth>
    <usg type="time">PRÉ-AO</usg>
  </form>
  <!--etc.-->
</entry>
```

Other details that can be observed by looking at Figure 127 will be explored further once we enter the microstructural components.

As we discussed in Chapter 2, the lemma is part of the macrostructure as well as the microstructure. As we have already described this component here, we will proceed with analysing other microstructural components.

9.2.3 Microstructural Level

In the DLP, after the lemma, the internal structure of all its lexicographic articles begins with grammatical information that should be specified as <entry>/<gramGrp>. This element can be used in two different places: as a sibling of the <form> element, when the annotation refers to all the forms present in the <entry>, or as a child of the <form> element when the information is specific to that form. As XML is verbose

¹⁷⁴ <https://volp-acl.pt/index.php/ortografia/texto-integral-do-ao90>

¹⁷⁵ The diphthong ‘oi’ loses its accent in paroxytone words.

enough, annotations in the DLP will mainly appear next to the <form> element, and when used inside it, they will describe only the properties that differ in that form.

Looking again to Figure 125, the next component is the POS (*nome* or noun), followed by the gender (*feminino* or feminine) in italics and abbreviated to save space in the DLPC and expanded in the DLP. We have annotated POS using <gram type="pos"> and also tagged the gender as <gram type="gen">. For interoperability reasons, we also use the @norm attribute for the Universal Dependencies¹⁷⁶ POS values. To guarantee the accuracy of this conversion, a list detailing the possibilities of that tag's content was computed, and the desired annotation was manually added to the POS.

In most cases, the sense has a (lexicographic) definition which is encoded in the <def> element.

The <etym> element contains etymological information. This element only occurs once per entry. The TEI Lex-0 as a TEI guideline recommends tagging separate elements of etymologies using multipurpose TEI tags. In the examples shown in this thesis, we have tried to apply the guidelines, but much work remains to be done in etymology, which will not be addressed here because it is beyond the scope of this work.¹⁷⁷ The examples collected here present the <etym> element with @type attributes containing a recursive <cit> construct for the etymons to be described. The etymons are associated with a language, a form and a bibliographical description. The <metamark> contains the graphic signal '(+)' that indicates the structural composition of the elements of formation of the lexical unit in question.

So far, we illustrate what has just been stated through the examples of Figures 125 and 127. Nevertheless, lexicographers generally assign a domain label preceding the definition when dealing with terms like “cristalografia” or “paleozoico”. This issue will be explored below.

¹⁷⁶ <https://universaldependencies.org/#language-u>

¹⁷⁷ For recent efforts to address etymology in TEI, see: Bowers et al. (2021); Khan et al. (2020); Sagot (2017).

9.3 Encoding Terms

There is no significant difference between encoding a lexical unit and a term in general language dictionaries. The use of the domain label solely characterises the latter.

9.3.1 Encoding Domain Labels

Within usage labels, the domain label is a crucial marker to identify terms in general language dictionaries. As an early step towards harmonising and standardising usage labels across dictionaries, we proposed a set of definitions for usage label categories. Domain label is defined as a ‘marker which identifies the specialised field of knowledge in which a lexical unit is mainly used’ (Salgado, Costa & Tasovac, 2019).

The restrictions that the TEI Lex-0 imposes on the TEI Guidelines are highly advantageous, as they allow a more precise and scientifically accurate encoding. It is considered good practice to restrict the scope of <usg>. The attribute @type must characterise/specify the element, in this case as a domain label. Given this, in TEI Lex-0, the @type is mandatory according to the fixed values set. TEI Lex-0, like the TEI Guidelines, offers a range of sample values to illustrate potential uses of the typed element¹⁷⁸ <usg>. TEI Lex-0 introduces a new naming scheme for the existing TEI original values to specify the observed phenomena.¹⁷⁹ Regarding domain labels, the original TEI Guidelines value, "dom", has been replaced by "domain" in TEI Lex-0 for the sake of clarity and objectivity:

```
<usg type="domain"/>
```

The specialised field of knowledge can be abbreviated (‘label-like descriptors’ in Tasovac, Romary et al., 2018) a very commonly used lexicographic convention for usage information in dictionary systems due to space restrictions in print dictionaries or expanded (‘fuller narrative expressions’ in Tasovac, Romary et al., 2018). The DLE and

¹⁷⁸ ‘Typed element’ means an element that can have a type and that specifies a set of values.

¹⁷⁹ For further details, see the table that shows the differences between suggested values of type in TEI and the required values of type in TEI Lex-0 to restrict the scope of <usg>, Chapter 7 – ‘Usage’: <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html#usage>.

the DLPC use abbreviated forms in the printed editions whose expansion is provided in the initial pages of these dictionaries, the online Spanish dictionary conserves the abbreviations and in the DAF, domain labels already appear expanded on the web, although not all DAF labels are already presented in their full form (cf. Chapter 6). When encoding dictionary data, it is important to normalise the abbreviated and unabbreviated labels to a single value for the sake of consistency and for better information retrieval. Let us demonstrate this with an example from our lexicographic corpus, focusing our analysis only on the domain label.

In Figure 124, the domain label *Miner.* from the DLPC corresponds to the full form MINERALOGIA [mineralogy] from DLP or Figure 125. A good practice is to encode the abbreviated domain label within the element <usg> followed by the attribute type required (@type="domain"). To provide the expanded form of the abbreviation, we may use the @expand attribute, as follows:

```
<usg type="domain" expand="Mineralogia">Min.</usg>
```

Another topic developed in Chapter 6 involved mapping the domain labelling of the three academy dictionaries. Even if a global harmonisation¹⁸⁰ effort is currently beyond the scope of this thesis, the proposal of a multilingual domain map led us to create new metadata to facilitate our analysis. As stated before, we established the equivalent English term as a metalabel assigned to the corresponding domain – see Table 19.

METALABEL	DLPC	DLE	DAF
crystallography	Cristalog. Cristalografía	—	—
geology	Geol. Geología	Geol. geología	géol. Géologie
mineralogy	Miner. Mineralogia	—	minér. Minéralogie
palaeontology	Paleont. Paleontología	—	paléont. Paléontologie
sports	Desp. Desporto	Dep. deportes	—
football	Fut. Futbol	—	—

¹⁸⁰ Although TEI employs terms such as ‘normalized/standardized’, we prefer to talk in terms of harmonisation.

Table 19: *Domains and subdomains under study and their metalabel*

To encode this metadata information, we encourage the use of the @norm attribute. This attribute ‘provides the normalised/standardised form of information present in the source text in a non-normalised form’¹⁸¹. Below, we show how to annotate the GEOLOGY domain:

```
<usg type="domain" expand="Geologia" norm="geology">Geol.</usg>
```

The annotations described above ensure better control of the terminological data and better verify its consistency. Using a metalabel will be beneficial for any work on aligning multiple dictionaries and studying them in parallel. However, an international harmonisation effort across different dictionaries would necessarily require further comparison of more dictionaries and a community-based agreement on the common values for metalabels.

One of the points discussed in the previous chapters refers to the importance of accessing a set of terms of a given domain, both for the lexicographer to control the terminologies included in the dictionary and for the user to search by a specific domain. For such lexical organisation to be possible, we have organised the domains under study in Chapter 7 and propose encoding hierarchical domain labels. In sum, and to illustrate our aim, FOOTBALL is considered a domain of the superdomain SPORTS; the domain GEOLOGY branches out to include terms belonging to the sub-branches of STRATIGRAPHY, MINERALOGY, PETROLOGY, etc., within the superdomain of EARTH SCIENCES.

We selected the geological sciences to illustrate what we propose. In the DLP, the MINERALOGY label could be conserved as a subdomain. However, we could add the domain GEOLOGY and the superdomain EARTH SCIENCES for the reasons presented previously, following the proposed methodology. To indicate a superdomain or a subdomain, they could be encoded using the @subtype attribute.

¹⁸¹ <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html#TEI.att.lexicographic.normalized>

```

<usg type="superdomain" expand="Ciências da Terra">C. Terra</usg>
<usg type="domain" expand="Geologia">Geol.</usg>
<usg type="subdomain" expand="Mineralogia">Min.</usg>

```

There are problems with this approach: first of all, the attribute values "superdomain" and "subdomain" are not valid according to the TEI Lex-0 schema. But even if they were, the <usg> element with @type and @subtype attributes would not present a sufficiently robust mechanism for encoding hierarchical domain labels. The above encoding shows three flat labels: the @type is used to indicate the position of the label in a hierarchy, but there is nothing in this encoding that explicitly indicates that these three labels belong to the same hierarchical chain. It may be implicitly clear to a human reader that CIÊNCIAS DA TERRA is the superdomain of the domain GEOLOGIA. However, from a machine-processing point of view, the link between the two is missing. The problems would be compounded if, in the future, or in a different dictionary, we resorted to the use of a more deeply nested hierarchy, i.e., beyond the tripartite structure of superdomain, domain and subdomain: it would be highly impractical to multiply the prefix 'sub' to indicate levels below subdomain (sub-subdomain, etc.)

To overcome the deficiency of flat representation of labels in general language dictionaries, we would ideally aim at a kind of encoding in which we can separate canonical, possibly multilingual, labels that are defined in one place and then simply pointed to from the dictionary entry. For this reason, we propose to employ the mechanism for the definition of taxonomies already available in the <teiHeader>. This is possible in both plain TEI and TEI Lex-0 but has not been documented until now as a solution for representing usage labels. With this approach, domain labels are documented in <encodingDesc> (encoding description)¹⁸². The domains established in the taxonomy are declared in <classDecl> (classification declarations)¹⁸³. This element is used to group the source of the domain's taxonomy used by the header or elsewhere in the document. First, the <taxonomy> (taxonomy)¹⁸⁴ element identifies the structured

¹⁸² 'Encoding description documents the relationship between an electronic text and the source or sources from which it was derived'; see <http://web.uvic.ca/lancenrd/martin/guidelines/ref-encodingDesc.html>.

¹⁸³ 'Classification declarations contains one or more taxonomies defining any classificatory codes used elsewhere in the text'; see <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-classDecl.html>.

¹⁸⁴ 'Taxonomy defines a typology either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy'; see <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-taxonomy.html>.

taxonomy. The categories will be documented in the <category> element¹⁸⁵. The category elements are described, each defining a single category within the given taxonomy. Then, child categories are defined by the contents of a nested <catDesc> (category description)¹⁸⁶ element, which contains the designation of the domain in question in the identified language. A single category may contain more than one <catDesc> child, and if you proceed with our work, the categories can be described in different languages (xml:lang). As a result of this thought process, we can establish a multilingual hierarchy for EARTH SCIENCES superdomain.

```
<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>
```

The hierarchical domain label for SPORTS domain labels is presented below:

```
<encodingDesc>
  <classDecl>
    <taxonomy xml:id="domain">
      <category xml:id="domain.sports">
        <catDesc xml:lang="en">Sport</catDesc>
        <catDesc xml:lang="pt">Desporto</catDesc>
        <catDesc xml:lang="es">Deporte</catDesc>
        <catDesc xml:lang="fr">Sport</catDesc>
      </category>
    </taxonomy>
  </classDecl>
</encodingDesc>
```

¹⁸⁵ 'Category contains an individual descriptive category, possibly nested within a superordinate category, within a user-defined taxonomy'; see <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-category.html>.

¹⁸⁶ 'Category description describes some category within a taxonomy or text typology, either in the form of a brief prose description or in terms of the situational parameters used by the TEI formal textDesc'; see <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-catDesc.html>.

```

        <category xml:id="domain.sports.teamsports">
          <catDesc xml:lang="en">Team Sports</catDesc>
          <catDesc xml:lang="pt">Desportos de Equipa</catDesc>
          <catDesc xml:lang="es">Deportes de equipo</catDesc>
          <catDesc xml:lang="fr">Sports d'équipe</catDesc>
          <category xml:id="domain.sports.teamsports.football">
            <catDesc xml:lang="en">Football</catDesc>
            <catDesc xml:lang="pt">Futebol</catDesc>
            <catDesc xml:lang="es">Fútbol</catDesc>
            <catDesc xml:lang="fr">Football</catDesc>
          </category>
        </category>
      </taxonomy>
    </classDecl>
  </encodingDesc>

```

The notions of correspondence and alignment are essential to the work that we have been doing concerning hierarchical domain labels. To encode such correspondence, we use the @corresp¹⁸⁷ attribute in the <usg> element. Since the reference points to a local element, its value takes the form of an abbreviated local pointer by simply preceding the destination value with a hash sign '#'. In this case, as the taxonomy is already structured in the <teiHeader>, we use an <usg> empty element that indicates the presence of an empty node within a content model that corresponds to the content inserted in the hierarchical tree in the <teiHeader>.

```
<usg type="domain" corresp="#domain.earth_sciences.geology"/>
```

Flat usage labels are, as we have seen, usually encoded as text values of the <usg> element. For the sake of human readability, one could deploy the same strategy and explicitly add the domain label as the content of the <usg> element even when the full label taxonomy is maintained in the <teiHeader>. This would be especially useful if labels used in a given dictionary are not consistent. For instance, in older dictionaries, one can encounter abbreviated and non-abbreviated labels used for the same domain. The text content of <usg> would then reflect the value of the label as it appears in the print dictionary regardless of the label as it is expressed in the taxonomy. In our case, using the @corresp attribute is sufficient because: (1) we consider this work a revision

¹⁸⁷ 'Corresponds points to elements that correspond to the current element in some way'; see <https://tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SACS>.

of the existing dictionary, not just a structural representation of the existing content; and (2) the way we construct @xml:id attribute on <category> is both machine and human-readable: each @xml:id contains the full hierarchical path for the given label within our taxonomy. For instance, the MINERALOGY subdomain has the @xml:id "earthsciences.geology.mineralogy". When processing the TEI file, it can be decided which labels will be displayed to the end user – e.g., we can choose whether we want all subdomains to be invisible, or just some of them, etc.

The @corresp attribute is one of the global linking attributes whose value, in our case, formalises the correspondence relationship with another identified element. Although the @corresp attribute works, we argue that it would be better for a well-recognised encoding of usage labels if <usg> was a member of att.canonical¹⁸⁸. This is a more precise mechanism, which is currently not allowed by TEI Guidelines for <usg>, but we would recommend implementing it in TEI. This way, we could use the @ref¹⁸⁹ attribute whose value is a tag URI – as defined in RFC 4151¹⁹⁰ – on <usg>.¹⁹¹

Moreover, domain labels can occur at different levels of the entry’s hierarchy. In the example in Figure 125, the position of the domain label can be encoded at the lemma level or even at the sense level since the “cristalografia” dictionary entry has only one meaning. In these cases, the lexicographic work should be uniform throughout the dictionary, so we recommend using the label at the <sense> element level. In addition, we have to suppose that a given lexical unit can generate new meanings in the future.

Recommend	Entry-level label
<pre><entry type="monolexicalUnit" xml:lang="pt" xml:id="DLP.cristalografia"> <form type="lemma"> <orth>cristalografia</orth> <pron>kriʃtelugre'fie</pron> </form> <gramGrp></pre>	<pre><entry type="monolexicalUnit" xml:lang="pt" xml:id="DLP.cristalografia"> <form type="lemma"> <orth>cristalografia</orth> <pron>kriʃtelugre'fie</pron> </form> <gramGrp></pre>

¹⁸⁸ ‘att.canonical provides attributes that can be used to associate a representation such as a name or title with canonical information about the object being named or referenced’; see <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.canonical.html>.

¹⁸⁹ ‘Reference provides an explicit means of locating a full definition or identity for the entity being named by means of one or more URIs’; see <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.canonical.html>.

¹⁹⁰ <https://www.ietf.org/rfc/rfc4151.txt>

¹⁹¹ Concerning this topic, we will open a ticket on GitHub to the TEI Council to make the change in TEI itself.

<pre> <gram type="pos" norm="NOUN">n.</gram> <gram type="gen">f.</gram> </gramGrp> <sense xml:id="DLP.cristalografia_1"> <usg type="domain" corresp="#domain.earth_sciences.geology.mineral ogy"/> <def>...</def> </sense> <etym> <!--etc.--> </entry> </pre>	<pre> <gram type="pos" norm="NOUN">n.</gram> <gram type="gen">f.</gram> </gramGrp> <usg type="domain" corresp="#domain.earth_sciences. geology.mineralogy"/> <sense xml:id="DLP.cristalografia_1"> <def>...</def> </sense> <etym> <!--etc.--> </entry> </pre>
---	---

Table 20: Domain label occurring at different levels of the entry's hierarchy

Furthermore, at the level of sense, and as seen in Chapter 4, the domain label, in addition to serving as an identifying device for a term, works very well as a distinctive element of meaning. Let us go back to an example given earlier, the entry “cratera” [crater] (Figure 34 and now 128) with several senses.

cratera [kretére]. *s. f.* (Do lat. *cratēra* < gr. κρατήρ).
1. Recipiente ou vaso antigo, geralmente de grandes dimensões, com duas asas, onde os gregos misturavam o vinho com água, para ser servido durante as refeições.
2. *Geol.* Abertura larga, geralmente arredondada e em forma de funil, por onde o vulcão deixa sair a lava, as matérias incandescentes ou os gases. *A cratera do vulcão.*
3. *Ind.* Abertura na parte superior de um forno de uma fábrica de vidros. 4. Buraco grande. *Caminho cheio de crateras.* 5. *Mil.* Buraco no solo provocado pela explosão de uma granada, projectil, bomba ou outra carga explosiva.
6. *Astr.* Depressão de origem vulcânica ou meteorítica, de dimensões variáveis, na superfície de um astro. *Cratera lunar. cratera meteorítica*, a que é produzida pelo impacto de um meteorito com a superfície de um planeta. 7. O que pode dar origem a infelicidade, calamidades, grandes acidentes... 8. Zona ou local onde alguma coisa ferve, fervilha, ou mostra grande actividade. ≈ AZÁFAMA.

Figure 128: Entry ‘cratera’ [crater] in the DLPC (ACL)

Senses 2, 3, 5 and 6 have domain labels: in sense 2, *Geol.* indicates that this sense belongs to the domain of GEOLOGY; sense 3 points to INDUSTRY (*Ind.*); sense 5 refers to the MILITARY domain (*Mil.*); and sense 6 is related to the field of ASTRONOMY (*Astr.*). These domain labels must be encoded according to the recommendation given in Table 20, i.e., after the <sense> element.

If we have an entry marked with meanings from the same domain – which, in Portuguese lexicography, often happens in botanical terms (plant and then flower) –, it may make sense that the domain label does not appear as repeated for the end user. This was a criterion adopted in the 2001 version (DLPC). We illustrate with the example “estrelícia” [strelitzia] from the DLPC (Figure 129).

estrelícia [i/ˈtɾilísjɐ]. *s. f.* (Do fr. *strélitzia*, do russo *streletz* ‘soldado do czar’). *Bot.* **1.** Planta caracterizada por folhas dísticas e flores hermafrodita (*Strelitzia*, Ait.), oriunda da África do Sul, cultivada na Europa como ornamental. **2.** Flor dessa planta. *Trouxe da Madeira um ramo de estrelícias.*

Figure 129: Entry ‘estrelícia’ [strelitzia] in the DLPC (ACL)

The domain label, *Bot.*, in this case, appears before the numbers that signal the different senses. In any case, we maintain our recommendation to mark the domain label at the sense level. Later, the domain label can be automatically moved by programming. This practice allows all senses to be kept correctly classified without losing terminological information.

It is also possible to make some domain labels invisible to the end user. In some cases the lexicographic definition may provide sufficient clarification, so the information pertaining to the domain label can be hidden for the user. On the other hand (and as explained in Chapter 7), some assigned subdomains may be hidden in the final version of the dictionary. Anyway, the markers are still helpful to retrieve information for lexicographic purposes. The reasoner and the search engine can use the hidden information to allow the user to find a specific term that belongs to a domain. To signal the visibility/invisibility of a particular label, we are currently using the attribute @rend¹⁹² with the value "hidden".

In brief, we list the steps that we consider relevant as part of best encoding practices regarding the encoding of the domain label in general language dictionaries:

¹⁹² ‘att.global.rendition provides rendering attributes common to all elements in the TEI encoding scheme’; see <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.global.rendition.html>.

- (i) We found advantages in using hierarchical domain labels (superdomain, domain, subdomain). For this, we first need to include a <taxonomy> in the <teiHeader> and then use correspondences designations.
- (ii) Use the element <usg> to annotate data about domain labels.
- (iii) Assign the "domain" value to the attribute @type.
- (iv) If the data uses abbreviated forms, we recommend providing the full form using the @expand attribute if using flat <usg> labels. Encoding the abbreviation and its respective full form at the same time is very useful. Later, we can decide how this information will be viewed when publishing the digital or printed data. If using a taxonomy in the <teiHeader>, full forms should be provided as values in <catDesc> elements.
- (v) The domain label can be associated at various points in the entry hierarchy. Its position must be analysed and evaluated, on a case-by-case basis, by the lexicographer.

Finally, we hope in the future to have the Portuguese Academy dictionary new edition linked to ontologies. Costa et al. (2020) proposed two possible markup approaches to associate the <usg> element to an ontology class: one that only uses the TEI Lex-0 format and another one that allows the expansion of the TEI Lex-0, namely the W3C XML Linking Language (XLink 1.1)¹⁹³ standard.

9.3.2 Encoding Polylexical Terms

As Tasovac, Salgado and Costa (2020) have pointed out, the modelling and encoding of polylexical units is a topic that has not been covered in sufficient depth by the TEI Guidelines. To overcome some issues, the authors (Tasovac, Salgado & Costa, 2020) introduce the notions of ‘macro- and microstructural relevance’ to differentiate between polylexical units that serve as headwords for their independent dictionary entries and those that appear inside entries for different headwords. The lack of consensus within the lexicographic community poses a challenge to the task of encoding

¹⁹³ <https://www.w3.org/TR/xlink11/>

dictionaries. The main question concerning polylexical units is how to describe these units using TEI recommendations formally.

As structural lexicographic components, polylexical terms can appear as entries ('macrostructurally relevant polylexical units' in Tasovac, Salgado & Costa, 2020) or in nested entry-like structures inside entries ('microstructurally relevant polylexical units' in Tasovac, Salgado & Costa, 2020) of a given lexicographic article.

The authors (Tasovac, Salgado & Costa, 2020, p. 34) introduce the notion of 'lexicographic transparency' to distinguish between those units which are not accompanied by an explicit definition and those that are accompanied by an explicit definition. The former are encoded as <form>-like constructs, whereas the latter become <entry>-like constructs, which can have further constraints imposed on them (sense numbers, domain labels, grammatical labels, etc.).

In the context of the DLPC and, more macrostructurally speaking, in the Portuguese orthographic tradition, hyphenation is treated as a mark of lexicalisation and non-compositional meaning, which leads to entry-level lexicographic treatment. For instance, "defesa-direito" [right back] is a lemma. As such, it is considered, from the point of view of the lexicographer, headword material.¹⁹⁴ Nevertheless, there is a strong likelihood that this form is found unhyphenated in corpora, so other dictionaries do not hyphenate it and therefore record it as a subheadword.

Concerning the lexicographic treatment of polylexical terms and their respective encoding, we are interested in analysing the so-called lexicographically non-transparent polylexical units in the microstructure. Such units follow a minimal <entry>-like structure (note that in the print edition, the expression is set in boldface, like a lemma) and are accompanied by a definition (or a pointer to a definition under a different entry). These units can themselves be divided into two further categories, based on the position they take up in the entry microstructure: i) those that are attached to particular senses; and ii) those that appear at the end of the entry, following the description of individual senses.

¹⁹⁴ The hyphen as a marker of semantic opaqueness, however, is to a certain extent a projection of lexicographic idealism. Many polylexicals that are traditionally hyphenated in Portuguese dictionaries are written without the hyphen in common usage.

Take, for instance, the example of Figure 130.

defesa¹ [difeze]. *s. f.* (Do lat. *defensa*). 1. Acção de proteger ou proteger-se; acto ou efeito de defender ou de defender-se. ≠ ATAQUE. *A defesa de um país. Instinto de +; reflexo de +. legítima+ defesa.* 2. O que serve para proteger, defender. *O guarda-chuva era a sua única defesa contra o aguaceiro. A trincheira era a única defesa dos soldados. Este forte já foi a principal defesa da ilha. Arma de +; linha de +; meio de +. defesa nacional*, conjunto de meios destinados a assegurar a integridade territorial de um país contra possíveis ataques estrangeiros. **mecanismo+ de defesa.** 3. Acto de patrocinar, de sustentar, de defender uma ideia, uma causa. *A defesa da liberdade. A defesa de uma tese. A defesa de uma ideologia. A defesa de um ideal. defesa de tese*, acto solene em que um candidato a determinado grau académico apresenta um trabalho inédito e argumenta em sua defesa perante um júri. 4. *Dir.* Advogado que defende um réu em juízo. *O juiz deu a palavra à defesa. A defesa apresentou as alegações finais. Advogado de +.* 5. *Dir.* Acto de justificar, defender ou defender-se um réu em tribunal contra uma acusação, expondo factos e produzindo provas em seu favor. = ALEGAÇÃO. *A defesa e a acusação.* 6. *Dir.* Conjunto de procedimentos adoptados pelo réu ou pelo seu advogado de defesa para refutar as acusações. *O advogado preparou cuidadosamente a defesa do arguido.* 7. Acto de impedir, de proibir. = IMPEDIMENTO, INTERDIÇÃO, PROIBIÇÃO. 8. *Us. pl. Zool.* Chifres de animais. = CHIFRE, CORNO, HASTE. 9. *Us. pl. Zool.* Dente muito saliente em alguns animais. *As defesas do elefante. As defesas do javali. As morsas defrontam-se, utilizando as defesas.* 10. Preservativo. 11. *Desp.* Intercepção da bola pelo guarda-redes, impedindo-a de entrar na baliza. *O guarda-redes fez uma defesa incrível.* 12. *Desp.* Acção de contrariar o ataque, a investida do adversário; acto ou efeito de defender. = ATAQUE. *«Errado é jogar à defesa.» (DN, 26.10.1988).* 13. *Desp.* Conjunto de jogadores que têm como função contrariar o ataque do adversário, actuando na parte recuada do meio campo da sua equipa. *A equipa adversária deparou com uma defesa intransponível.* 14. *Equir.* Resistência oposta por um cavalo à transposição de um obstáculo ou para se subtrair à acção do cavaleiro. 15. *Region. (Alent.).* Grande herdade ou propriedade rústica. **em defesa de**, *loc. prep.*, em apoio de; em favor de. *O governo tomou medidas em defesa do meio ambiente. O sindicato agiu em defesa dos direitos dos trabalhadores.* **em legítima defesa**, *loc. adv.*, em reacção a uma agressão ilícita, actual ou iminente. *Disparou em legítima defesa. Matou em legítima defesa.* **jogar à defesa**. 1. *Fut.* Procurar defender a sua baliza, sem atacar, sem procurar marcar golos. 2. Não se expor. *Não consegue ser espontânea por estar permanentemente a jogar à defesa.* **não ter defesa (possível)**, *Fam.*, não ter qualquer justificação para a prática de um acto considerado errado e reprovável.

defesa² [difeze]. *s. m.* (Do lat. *defensa*). *Desp.* Jogador de futebol ou de outros desportos, que actua na parte recuada do meio campo da sua equipa. *«ficou na frente, não apenas para dar continuidade aos lances mas também para 'prender' os defesas contrários.» (DN, 27.10.1988).*

Figure 130: Entry ‘defesa’ [defence] in the DLPC (ACL)

The lexicographic article “defesa” [defense] from DLPC illustrates a case of a polylexical unit. The monolexical item “defesa” [defense] is the lemma for a lexicographic article with fifteen different numbered senses. Senses 11, 12, and 13 are labelled with *Desp.* (the abbreviation of DESPORTO [sport]). We found a polylexical item (a collocation) related to FOOTBALL (domain label = *Fut.*), ‘jogar à defesa’ [play defence], which appears in boldface, just like the lemma, and has two numbered meanings: 1) ‘procurar defender a sua baliza, sem atacar, sem procurar marcar golos’ [trying to

defend their goal, without attacking, without trying to score goals] and 2) ‘*não se expor*’ [abstaining from exposing yourself]. These senses are not explicitly labelled – they are not accompanied by a label that identifies the given unit as a ‘collocation’. Highlighting is only given by the use of boldface. As indicated in Chapter 6, the first sense of this polylexical unit could be associated with the senses related to SPORT, but here appear at the end of the lexicographic article.

The sense-related non-transparent polylexical unit (*‘jogar à defesa’*) can be encoded in TEI Lex-0 within an <entry> construct.¹⁹⁵ The type of the polylexical unit is indicated by the <gram> element.

Because lexicographically transparent polylexical units are not structured as mini-entries but are instead presented to the reader as a sequence of forms, we recommend encoding them as <form> elements (<form type="collocation">). Finally, because sense-related polylexical units are modelled as nested entries, they can include domain labels as well.¹⁹⁶

```
<!--etc.-->
<sense n="12" xml:id="DLP-defesa_1">
  <usg type="domain" corresp="#domain.sports.football"/>
  <!--etc.-->
  <entry xml:id="jogar_a_defesa" xml:lang="pt" type="relatedEntry">
    <form type="collocation">
      <orth>jogar à defesa</orth>
    </form>
    <sense xml:id="jogar_a_defesa">
      <usg type="domain" corresp="#domain.sports.football"/>
      <def>procurar defender a sua baliza, sem atacar ou sem
procurar marcar golos</def>
    </sense>
  </entry>
  <cit type="example">
    <quote>Errado é jogar à defesa.</quote>
    <bibl>
      <title>DN</title>
      <date>26.10.1988</date>
    </bibl>
  </cit>
```

¹⁹⁵ TEI and TEI Lex-0 diverge somewhat on how they allow this, but the end result is the same: in TEI Lex-0, the content model of <sense> allows elements from the class model1.sensePart as its children, and <entry> is a member of this class; whereas in TEI <sense> has a broader content model which allows members of the class model entryPart as its children.

¹⁹⁶ For a comprehensive encoding of this lexicographic article, see the repository on GitHub: https://github.com/anacastrosalgado/DLP/tree/master/PhD_work.

```
</sense>
<!--etc.-->
```

In Chapter 7, we saw how the modelling of concept systems facilitates the definition of semantic relationships. We move on to see how to encode these relationships.

9.3.3 Encoding Semantic Relations

Semantic relations are encoded within specific senses. The recommended way to encode semantic relations in TEI Lex-0 is the external relation element provided by `<xr>`. The different types of semantic relations are identified in `@type` (e.g., `<xr type="synonymy"></xr>`).

To illustrate the encoding of synonyms, we chose the “guarda-redes” [goalkeeper] entry since this term has a synonym with a usage label (geographical label, *Bras.* or Brazil). The Brazilian units “arqueiro” and “goleiro” are thus equivalent to the Portuguese variant “guarda-redes”.

guarda-redes [gwardɐʁɐˈɐdʁɨ]

Entrada validada

nome masculino e feminino singular e plural

DESPORTO jogador de uma equipa que atua na baliza, cuja função é impedir a entrada da bola na sua baliza com o objetivo de evitar que a equipa adversária marque golos ou pontos

SINÓNIMOS *arqueiro (BRAS.); goleiro (BRAS.)*

🔥 *O guarda-redes, com uma exibição de luxo, foi a figura do jogo.*

🗣️ Termo recorrente em desportos coletivos, designadamente no futebol, andebol, hóquei, etc.

ETIMOLOGIA De forma do verbo *guardar* + *rede*

Figure 131: Entry ‘guarda-redes’ [goalkeeper] in the DLP (ACL)

```
<entry type="monolexicalUnit" xml:lang="pt" xml:id="DLP.guarda_redes">
  <form type="lemma">
    <orth>guarda_redes</orth>
    <pron>gwardɐˈɐdɐʁɨ</pron>
  </form>
  <gramGrp>
    <gram type="pos" norm="NOUN">nome</gram>
    <gram type="gen">masculino</gram>
```

```

        <pc>e</pc>
        <gram type="gen">feminino</gram>
        <gram type="num">singular</gram>
        <pc>e</pc>
        <gram type="num">plural</gram>
    </gramGrp>
    <sense xml:id="DLP.guarda_redes_1">
        <usg type="domain" corresp="#domain.sports"/>
        <def>jogador de uma equipa que atua na baliza, cuja função é
impedir a entrada da bola na sua baliza com o objetivo de evitar que a
equipa adversária marque golos ou pontos</def>
        <xr type="synonymy">
            <ref type="entry">arqueiro</ref>
        </xr>
        <usg type="geographic" corresp="#geographic.brasil">Brasil</usg>
        <xr type="synonymy">
            <ref type="entry">goleiro</ref>
        </xr>
        <usg type="geographic"
corresp="#geographic.brasil">Brasil</usg>
        <cit type="example">
            <quote type="example">O guarda-redes, com uma exibição de
luxo, foi a figura do jogo.</quote>
        </cit>
        <note type="use">Termo recorrente em desportos coletivos,
designadamente no futebol, andebol, hóquei, etc.</note>
    </sense>
    <etym>De forma do verbo guardar + rede</etym>
    <etym type="grammaticalization">
        <seg type="desc">Da forma do verbo</seg>
        <cit type="etymon" xml:lang="pt">
            <form>
                <orth>guardar</orth>
            </form>
        </cit>
    </etym>
    <etym type="grammaticalization">
        <metamark>+</metamark>
        <cit type="etymon" xml:lang="pt">
            <form>
                <orth>rede</orth>
            </form>
        </cit>
    </etym>
</entry>

```

This example thus illustrates that usage labels, in this case, a geographic label (*Bras.* or *Brazil*), can also be associated with synonyms.

In the geology context, we have established (see Chapter 7) that “idade”, “época”, “período”, “era” and “éon” (specific terms) are the hyponym of the hypernym “unidade geocronológica” (generic term). In TEI Lex-0, hyperonyms are encoded inside <xr type="hypernymy"></xr>. The hyponyms are encoded inside <xr type="hyponymy"></xr>. We illustrate a case of a hyperonymy using the lexicographic article “éon”.

```

<entry type="monolexicalUnit" xml:lang="pt" xml:id="DLP.eon">
  <form type="lemma">
    <orth>éon</orth>
    <pron>'ɛn</pron>
  </form>
  <gramGrp>
    <gram type="pos" norm="NOUN">nome</gram>
    <gram type="gen">masculino</gram>
  </gramGrp>
  <sense n="1" xml:id="DLP-eon_1">
    <def>divisão de tempo infinitamente longa</def>
  </sense>
  <sense xml:id="DLP-eon_2" n="2">
    <usg type="domain">Filos.</usg>
    <def>espírito que emana da inteligência eterna</def>
  </sense>
  <sense xml:id="DLP-eon_3" n="3">
    <usg type="domain"
corresp="#domain.earth_sciences.geology.stratigraphy"/>
    <def>intervalo de tempo geológico (<xr>
      <ref type="entry">unidade geocronológica</ref>
</xr>) durante o qual se formou um eonotema (<xr type="hypernymy">
      <ref type="entry">unidade cronostatigráfica</ref>
</xr>)</def>
    <form type="collocations">
      <form type="collocation">
        <orth>
          <ref type="oRef">
            <lbl>+</lbl>
          </ref>
          <seg>éon fanerozoico</seg>
        </orth>
        <gramGrp>
          <gram type="mwe" value="co-ocorrente_privilegiado"/>
        </gramGrp>
      </form>
    </form>
    <note type="encyclopedic">1) Na escala do tempo geológico, o éon é
a categoria hierárquica
    mais elevada. 2) O éon integra várias eras.</note>
  </sense>
  <etym>
    <etym type="inheritance">
      <seg type="desc">Do latim médio</seg>
      <cit type="etymon" xml:lang="la">
        <form><orth>aeon</orth></form>
      </cit>
      <seg type="desc">pelo grego</seg>
      <cit type="etymon" xml:lang="grc">
        <form><orth>ἀἰών</orth></form>
        <pc>'</pc>
        <gloss>eternidade</gloss>
        <pc>'</pc>
      </cit>
    </etym>
  </etym>
  <note type="plural">Plural: éones</note>
</entry>

```

9.3.4 Encoding Other Components

The <cit> (cited quotation) element contains a text fragment with at least one occurrence of the word form, used in the sense described. In the DLP, we can have usage examples, fragments extracted from corpora or even made up by the lexicographer (<cit type="example">), and illustrative quotations from books, newspapers or periodicals (cit/quote/bibl) that contain a loosely structured bibliographic citation whose sub-components may be explicitly tagged. The last element always contains a bibliographic reference to its source. We provide an example (Figure 132) that illustrates the (cit/quote) element.

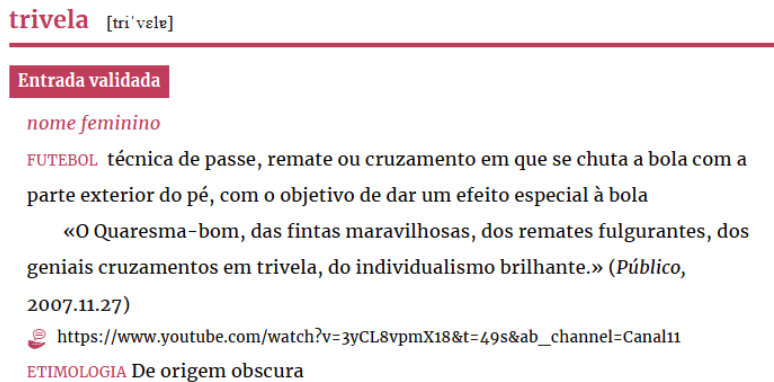


Figure 132: Entry 'trivela' in the DLP (ACL)

As we stated in Chapter 6, a link to a YouTube video could be provided to illustrate what a 'trivela' is in the football context. We decided to include the link in <note> with the @type attribute with the value "media" followed by the URL.¹⁹⁷

```
<entry type="monolexicalUnit" xml:lang="pt" xml:id="DLP.trivela">
  <form><orth>trivela</orth>
    <pron>tri'vele</pron>
  </form>
  <gramGrp>
    <gram type="pos" norm="NOUN">n.</gram>
    <gram type="gen">f.</gram>
  </gramGrp>
  <sense xml:id="DLP-trivela_1-dbf7a-1">
    <usg type="domain" corresp="#domain.sports.football"/>
  </sense>
</entry>
```

¹⁹⁷ We follow this approach because there are several notes in the DLP, namely, encyclopaedic, usage and spelling (case), and all of them are appropriately marked.

```

    <def>técnica de passe, remate ou cruzamento em que se chuta a bola com
a parte exterior do pé, com o objetivo de dar um efeito especial à bola</def>
    <cit type="example">
      <quote>O Quaresma-bom, das fintas maravilhosas, dos remates
fulgurantes, dos geniais
      cruzamentos em trivela, do individualismo brilhante.</quote>
      <bibl><title>Público</title>
      <date>2007.11.27</date>
    </bibl>
    </cit>
    <note
type="media">https://www.youtube.com/watch?v=3yCL8vpmX18& t=49s& ab_chan
nel=Canal11>
    </note>
    </sense>
    <etym>De origem obscura</etym>
</entry>

```

The encodings presented here attest that the TEI Lex-0 specifications respond positively to our current needs. Instead of having only a flat label system, we propose that a hierarchical treatment of usage labels be explicitly included in the TEI Lex-0 Guidelines. This could be an important basis for the eventual harmonisation of usage labels across TEI-based dictionaries and different languages. Finally, for a comprehensive encoding of all the lexicographic articles mentioned throughout this chapter and others encoded terms that illustrate our purposes, see the repository on GitHub¹⁹⁸. We will now move on to the next chapter of this study, where the conclusion and some future directions will be presented.

¹⁹⁸ https://github.com/anacastrosalgado/DLP/tree/master/PhD_work

CONCLUDING REMARKS

The primary motivation for this study was to improve the lexicographic work carried out on the ACL. Nevertheless, we invested in a broader multilingual scale within the European lexicographic arena. Thus, so as not to restrict our research to the national level, we selected other academy dictionaries as our objects of study. The main reason for creating a contrasting corpus is that although languages and dictionaries are different, they have similar problems. By observing and comparing various lexicographic resources, we believe we are taking an essential step towards a possible homogenisation of the representation of lexicographic data striving to solve the problems detected.

The paradigm change from paper to digital underlines the need to rethink the theoretical and methodological assumptions of the Portuguese lexicographic tradition. Furthermore, this emphasises the importance of distinguishing between the units that belong to the common language and the terms that occur in different specialised texts and discourses. We took this opportunity to invest in the quality of the specialised meanings that will soon be available when the DLP becomes publicly available.

The practical lexicographic work involves multiple tasks; thus, we have restricted our research to the treatment of terms in general language dictionaries. The increasingly frequent inclusion of terms in those dictionaries is related to the democratisation of knowledge and technological advances. We have seen that it is not the original degree of specialisation of a given term that justifies its inclusion in a general language dictionary, but rather how much users and speakers of a given language need that term.

Throughout this thesis, the methodology we proposed answered positively the questions raised in the Introduction of this work. In the following lines, we briefly summarise the discussion undertaken in this research by returning to the questions.

(i) Might principles and methods of terminology work contribute to lexicographic work?

This research project aimed to discuss certain decisions traditionally taken by lexicographers. In our view, the customary methodology needs to be reformulated

regarding the treatment of terms. In the first theoretical chapters, we saw that establishing boundaries between words and terms is very difficult. Terms also appear in general language dictionaries, and it is difficult to identify them when the domain label is absent.

We assume that a terminologically-based methodology could be advantageous and improve the quality of the lexicographic product both in terms of representation and organisation of knowledge and the description of terms themselves – the conceptual and linguistic dimensions.

From the very beginning, the conclusions offered by this work were intended to be logically dependent on the assumptions (theoretical foundation) from which it departed. To achieve this goal, we use real examples of how lexicographers should treat terms based on proper terminological analysis. During the course of this research, interaction with specialists proved to be essential. Specialists provided all the relevant information for acquiring fundamental knowledge, indicated the essential literature that should be read and aided in the subsequent constitution of the corpus, answering our questions and validating the concept system.

(ii) How are terms treated in general language dictionaries, namely in academy dictionaries?

All three academy dictionaries lack explicit explanatory information regarding the treatment of terms. In addition to finding marked meanings, the lexicographic methodology is the same whether we deal with lexical units (words in general) or terminological units (terms). We propose following a terminological-based approach to the treatment of terms in general language dictionaries. We favoured the so-called terminological definitions rather than lexicographical ones to guarantee the quality of the final product and provide greater clarity to the lexicographer, who often feels insecure (or uncomfortable) when they have to define terms. We argue that the definition, even the (lexicographic) definition, is needed to place the term in its appropriate position in the knowledge structure. Since it is a purely terminological activity, we can call it a terminological definition even in general language dictionaries.

Concerning the inclusion of terms in general language dictionaries, we consider their presence unquestionable. However, highly specialised terms used only by a very limited number of specialists should not be included. When they are required to write definitions, their inclusion is mandatory, but when in doubt while modelling concept systems, the use of corpora can help the lexicographer decide whether a given term should be included.

We argue that domain labels should not only be shown as a flat list in the outside matter. A well-organised hierarchy will help lexicographers and end-users better understand the relations between concepts.

(iii) What domains are currently represented in these works? Are those domains conceptually organised?

The importance of diatechnical information in lexicography is indisputable. We examined the front matter of the print editions of the DLPC and the DLE, as well as the introductory texts available on the DAF webpage, to ascertain whether explicit references were made to the adopted labelling system and/or to any criterion or justification for the presence of diatechnical information. Some inconsistencies were observed in the dictionaries analysed in this thesis, which can be attributed to the absence of an explicit methodology when they were originally compiled. The three academy dictionaries include only brief references to usage labelling and do not explain the use of domain labels. Additionally, the number of labels selected by the lexicographers in charge of these dictionaries is unbalanced. There is also an imbalance in the scope of the labels, where the DLPC and the DAF have many examples of the so-called subdomains that the DLE ignores. A proposal for international harmonisation, therefore, is still a mirage. The dictionaries under study seem to be supported only by a flat list of abbreviations that contains different types of information. For more structured and founded domain lists, we questioned the presence of general domains accompanied by unstructured subdomains. To ameliorate this situation, we believe that the criteria followed by lexicographers to make decisions on the inclusion of terminological data should be included in future editions of those works, even digital ones.

Structuring a domain is a terminological task. This organisation is fundamental to improving the labelling systems in dictionaries. We analysed the domain labelling, suggesting the elimination of any unnecessary or repetitive markings as well as those distinctions that can sometimes seem arbitrary because they are too narrow, both from the point of view of a lexicographer and that of a regular user of the dictionary. This was the starting point to move from a non-hierarchical organisation to a hierarchical system, which consequently increases the consistency of annotation and information retrieval.

After collecting all the domain labels used in the academy dictionaries that constitute our corpus and analysing them, we decided to structure two domains: GEOLOGY and related geological sciences and FOOTBALL. After a practical exercise, we were able to show how much the quality of the definitions improved following the application of a terminological methodology.

(iv) What is the role or function of the domain label in academy dictionaries?

The role of a domain label is to identify the specialised field of knowledge in which a lexical unit is mainly used. Domain labelling can be seen as a lexicographic device for knowledge organisation in a given lexical resource. Our analysis confirms that domain labels point to terms. In addition, in the three dictionaries under observation, other ways of labelling domains, such as linguistic formulae found in the definitions, have the same functions as domain labels. From our point of view, these labelling systems are in need of an urgent revision, eliminating unnecessary or repetitive labels, as well as those distinctions that are too fine. Sometimes these excessively fine distinctions seem arbitrary from the perspective of both lexicographers and dictionary users. Some inconsistencies were also observed in the usage of abbreviated forms, which are used only occasionally. Our findings, however, are relevant not only for lexicographic practice but also for dictionary encoders. The tacit knowledge and implicit rules of lexicographic procedures make not only the encoders' jobs more difficult but the dictionary itself less transparent to users.

(v) Is it possible to map the domain labels between the different academy lexicographic resources?

By proposing hierarchical domain labels, we organise knowledge and establish higher and lower categories. The fact that we define a domain hierarchy does not mean that all proposed labels will be visible in the final product. This means that the lexicographer must structure the domains thoroughly and identify the terms according to the classification adopted. However, later on, the decision to make domain categories visible to the public must be weighed and considered taking into account the number of terms classified with that label and also looking at the set of tags and their statistics in the set of an established superdomain. The decision to make domain labels visible or invisible must be made by teams of editors and lexicographers. To implement good practices, lexicographers should join forces to collaborate in the proposal to harmonise domain labels and thus improve the diatechnical marking process in academy dictionaries. We have to recognise that there is no ideal or unique model to follow. Still, we argue for the necessity of following good practices. This harmonisation is all the more valuable as it further advances structured lexical databases based on standards that allow access to the construction of lexicographic resources adapted to the necessary interoperability.

(vi) If we organise the domains, identify the concepts and the relations drawn between them, model concept systems and then search for the terms linked to the identified concepts, will all this improve the definitions of the concepts pointed at by the terms?

We should emphasise that we endorse the definition of the concept. The onomasiological perspective makes us look at the concept, identify it, isolate it, specify its characteristics and differentiate that concept from others that belong to the same concept system. Only after these relationships are established, the lexicographer will be able to propose a definition that can be validated by the domain specialist. The analysis of the definitions according to the conceptual aspect is relevant in dictionaries even if the audience is not made up of experts. As recommended by ISO 704 (2009), we conclude that intensional definitions are beneficial. In addition to domain labels, we found other mechanisms to mark specialised information, such as the use of formulae present in the definition. In this case, as we demonstrated in our examples, we think that the best place to provide additional information is in a note field.

We also showed that conceptual identifiers and linguistic markers may help lexicographers draft definitions. Focusing on the characteristics of a given concept is a fundamental step when defining it. In the DLP, we tested the creation of natural language definitions using concept systems. The results obtained are immensely satisfactory, ensuring greater definition accuracy and quality. Instead of working a dictionary by classical alphabetical ordering (from A to Z), i.e., letter by letter, we found advantages in treating entries by sets of terms, first identifying the generic concept and describing its characteristics, and thus distinguishing it from other concepts.

(vii) Do the TEI Lex-0's specifications meet the identified requirements to represent terms?

By examining the encoding of the terms analysed here, we confirmed that TEI Lex-0 meets our research needs. After encoding the microstructural components needed when terms are at the core of lexicographic work, we can ensure the interoperability and reusability of the specialised data. The advantage of applying TEI Lex-0 lies in the fact that lexicographers and terminologists are currently trying to apply TEI to the ongoing review of the ISO LMF. Given TEI Lex-0 (still) has a non-standard nature, it can be changed to accommodate relevant dictionary structures. We intend to demonstrate that the results obtained are helpful for computational lexical encoding and can serve the purpose of natural language processing. One of the main contributions of this research was to analyse, confront and discuss the different domain labels used in academy dictionaries. We have shown that the currently recommended TEI Lex-0 practice for representing domain labels as flat values is not robust enough to deal with more complex, hierarchical domain structures. The proposal that we present here for encoding hierarchical domain labels has the advantage of being usable in any dictionary, including multilingual ones. We recognise, however, that it is only a starting point for what we consider to be a joint effort to standardise domain labels and that only two domains were worked in with a sampling of examples in each. In the future, we are also interested in exploring the results in the field of ontology, as we did for OntoDomLab-Med (Costa et al., 2020; Costa et al., 2021d).

The need to apply standardised models within the lexicographic universe reveals that these cannot be closed models. As long as there is no harmonisation between the

various European and world lexicographical resources, there is always a need to change the scheme of these formal representations to respond to the requirements of these resources. On the other hand, the desire to link data across the web calls for the alignment of these resources.

We conclude with five final considerations:

1. Our research has strictly lexicographic purposes, using terminological methods to contribute to the guidelines for a methodology for processing terms in general language dictionaries and for definitions, namely the dictionaries of the national academies analysed here, proposing a new dictionary model that combines lexicographic methods and terminological practices in a harmonised and balanced way.

2. Combining conceptual and linguistic dimensions involves an iterative procedure. Knowing the domain and then organising it are necessary tasks for a quick and systematic identification of basic concepts, which will result in a better description of the lexicon. This facilitates encoding by fostering a more orderly data classification depending on each element, such as the entry or sense. Because these units are marked with domain labels, specialists must intervene and assist in organising knowledge and validating the lexicographic content, resulting in more accurate encoding. Lexicographers should select a limited number of concepts to avoid inconsistencies, structure them into concept systems and locate them in the system. The use of diagrams proved to be helpful for the organisation work.

3. Finding out that the dictionaries that make up our lexicographic corpus share common problems concerning the examples analysed led us to suggest that it would be interesting to present identical solutions for all of them. We believe that our methodology is helpful for lexicographers to organise the domain labelling system, improving and bringing accuracy to the process of writing terminological definitions adapted to general language dictionaries. The solutions presented for the Portuguese language dictionary (DLP) can be replicated in other language dictionaries.

4. Although we have restricted our analysis to two specific domains, we believe this methodology can be replicated in other domains. The next step we have in mind is to test this methodology on terms from mathematics (since we found a strong

presence of subdomains related to the domain of mathematics), chemistry (chemical elements from the periodical table), and metrology (units of measurement from the International System of Units). Expert validation is a must.

5. The continuous expansion of the multilingual information society has led to a pressing demand for multilingual linguistic resources suitable for different applications. In this regard, specific important works include WordNet domains (e.g., Magnini & Cavaglià, 2000; Bentivogli et al., 2004; Gella et al., 2014). Concepts such as interoperability, reusability, linking data and data alignment are increasingly necessary for a lexicographer. For this reason, we argue that lexicographic metadata should be harmonised between different lexicographic resources where possible. This so mainly because we deal with a large amount of data, consequently increasing the difficulty of maximising the reusability of these resources. The retrodigitisation of printed lexicographic works highlights the inconsistencies of the labelling system. The harmonisation of existing language resources requires international standards and guidelines (Ide & Romary, 2007) to develop language technologies and conceptual modelling based on ISO standards (704, 2009; 1087, 2019) that yield terminologies that benefit the development of this multilingual information society. In terms of interoperability, the use of hierarchical domain labels is advantageous; it allows labels to be brought closer to different dictionaries and, in turn, makes their reusability profitable. An agreement between academies and other institutions would be desirable to systematise and optimise a new type of lexicography that can better represent the entire European lexicographic heritage.

We will continue to invest in an effective trans-disciplinary approach that combines theories and methods of terminology and lexicography, and even other disciplines, placing best practice standards at the core of our research. Unquestionably, terminology, with its interdisciplinary nature, is at the core of knowledge conceptualisation and organisation, which justifies our approach.

BIBLIOGRAPHY

Dictionaries

DA (1770) = Real Academia Española. (1770). *Diccionario de Autoridades*.

DAF = Académie Française. (2021). *Dictionnaire de l'Académie Française* (9th ed.). (2021). Retrieved from <http://www.dictionnaire-academie.fr/>.

DAF (1694) = *Le Dictionnaire de l'Académie Française, Dédié au Roy*. (1694). 1st edition. Paris: Chez Vve J. B. Coignard et J. B. Coignard. Retrieved from <https://gallica.bnf.fr/ark:/12148/bpt6k503971>.

DAF (1718) = *Nouveau Dictionnaire de l'Académie Française Dédié au Roy*. (1718). 2nd edition. Paris: Chez Jean-Baptiste Coignard. Retrieved from <https://gallica.bnf.fr/ark:/12148/bpt6k12803909>.

DLE = Real Academia Española. (2021). *Diccionario de la lengua española* (24th ed.). Retrieved from www.rae.es/rae.

DLE (2014) = Real Academia Española. (2014). *Diccionario de la lengua española*. 23th edition.

DLP = Academia das Ciências de Lisboa (2021). *Dicionário da Língua Portuguesa*. Salgado, A. (Coord.). Lisboa: Academia das Ciências de Lisboa. [New digital edition under revision.]

DLPC = Academia das Ciências de Lisboa (2001). *Dicionário da Língua Portuguesa Contemporânea*, 2 vols. Castelleiro, J. M. (Coord.). Lisboa: Academia das Ciências de Lisboa and Editorial Verbo.

GDLP (2010) = *Grande Dicionário da Língua Portuguesa*. (2010). Porto: Porto Editora.

HOUAISS = *Grande Dicionário Houaiss da Língua Portuguesa*. (2015). Lisboa: Círculo de Leitores.

INFOPÉDIA = *Dicionário Infopédia da Língua Portuguesa*. (2021). Porto Editora. Retrieved from <https://www.infopedia.pt/>.

MACMILLAN (2007) = *Macmillan dictionary for children*. (2007). Ed. by Cristopher G. Morris. Australia: Simon & Schuster.

MACMILLAN (2021) = *Macmillan Dictionary*. (2021). Retrieved from <https://www.macmillandictionary.com/>.

OED = *Oxford English Dictionary*. (2021). Oxford University Press. Retrieved from <https://www.oed.com/>.

PE (1956) = Costa, J. A., & Melo, A. S. (1956). *Dicionário da Língua Portuguesa*. 3.^a ed. muito corrigida e aumentada. Porto Editora.

PRIBERAM (2021). *Dicionário Priberam da Língua Portuguesa*. Retrieved from <https://dicionario.priberam.org/>.

Literature

- Abel, A. (2012). Dictionary writing systems and beyond. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 83–106). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199654864.003.0005.
- Abromeit, F., Chiarcos, C., Fäth, C., & Ionov, M. (2016). Linking the tower of Babel: modelling a massive set of etymological dictionaries as RDF. In J. McCrae et al. (Eds.), *Proceedings of the 5th Workshop on Linked Data in Linguistics (LDL-2016): Managing, Building and Using Linked Language Resources, Portoroz, Slovenia* (pp. 11–19). Retrieved from http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-LDL2016_Proceedings.pdf.
- ACL. (1780). *Plano de Estatutos, em que convierão os primeiros socios da Acaemia das Sciencias de Lisboa, com beneplacito de sua Magestade*. Lisboa, Regia Officina Typografica.
- ACL. (1793). Planta para se formar o Diccionario da lingoa portugueza. In *Diccionario da lingoa portugueza*, t. 1, A (pp. i-xx). Academia Real das Ciências de Lisboa. Lisboa: Na Officina da mesma Academia.
- ACL. (1799). *Catalogo dos livros, que se hão de ler para a continuação do diccionario da língua portugueza: Mandado publicar pela Academia Real das Sciencias de Lisboa*. Lisboa: Na Typographia da mesma Academia. Retrieved from <https://bibdig.biblioteca.unesp.br/handle/10/28356>.
- ACL. (1870). *Relatório da Comissão encarregada de propor à Academia Real das Sciencias de Lisboa o modo de levar a efeito a publicação do Diccionario da Lingua Portuguesa*. Lisboa: Typographia da Academia.
- ACL. (1987). *Instituto de Lexicologia e Lexicografia da Língua Portuguesa*. Lisboa: Academia das Ciências de Lisboa.
- Adamska-Sałaciak, A. (2019). Lexicography and theory: clearing the ground. *International Journal of Lexicography*, 32(1), 1–19. doi:10.1093/ijl/ecy017.
- AF. (1635/1995). *Statuts et règlements*. Retrieved from https://www.academie-francaise.fr/sites/academie-francaise.fr/files/statuts_af_0.pdf.
- AF. (1694). Préface de la première édition. In *Dictionnaire de l'Académie Française*, s. p. Retrieved from <https://www.academie-francaise.fr/le-dictionnaire-les-neuf-prefaces/preface-de-la-premiere-edition-1694>.
- AF. (1798). Préface de la cinquième édition. In *Dictionnaire de l'Académie Française*, s. p. Retrieved from <https://www.academie-francaise.fr/le-dictionnaire-les-neufs-prefaces/preface-de-la-cinquieme-edition-1798>.
- AF. (2021). *La nouvelle édition numérique du Dictionnaire de l'Académie française, dans ses différentes éditions*. Retrieved from <https://www.dictionnaire-academie.fr/presentation.html>.
- Ahmadi, S., McCrae, J., Nimb, S., Khan, F., Monachini, M., Pedersen, B., Declerck, T., Wissik, T., Bellandi, A., Pisani, I., Troelsgård, T., Olsen, S., Krek, S., Lipp, V., Váradi T., Simon, L., Gyorffy, A., Tiberius, C., Schoonheim, T., Ben Moshe, Y., Rudich, M., Abu Ahmad, R., Lonke, D., Kovalenko, K., Langemets, M., Kallas, J., Dereza, O.,

- Fransen, T., Cillessen, D., Lindemann, D., Alonso, M., Salgado, A., Luis Sancho, J., Ureña-Ruiz, R.J., Porta Zamorano, J., Simov, K., Osenova, P., Kancheva, Z., Radev, I., Stanković, R., Perdihi, A., & Gabrovsek, D. (2020). A Multilingual Evaluation Dataset for Monolingual Word Sense Alignment. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC2020)*, 11–16 May (pp. 3232–3242). France: Marseille.
- Ahumada, I. (Ed.) (2002). *Diccionarios y lenguas de especialidad*. Jaén: Universidad de Jaén.
- Al-Kasimi, A. M. (2019). The history of Arabic lexicography and terminology. *Handbook of Terminology*, vol. 2, pp. 7–30.
- Alves, D. (2016). As humanidades digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português. *Ler História*, 69. doi:10.4000/lerhistoria.2496.
- Alves, I. M. (1997). Contribuição ao estudo do vocabulário da habitação: a palavra casa nos dicionários da Língua Portuguesa. *Anais do Museu Paulista: História E Cultura Material*, 5(1), 163–172. doi:10.1590/S0101-47141997000100005.
- Amaral, I. (2012). *Notas históricas sobre os primeiros tempos da Academia das Ciências de Lisboa*. Lisboa: Colibri.
- Amsler, R. A. (1980). *The Structure of the Merriam-Webster Pocket Dictionary*. Austin: University of Texas.
- Arnold, I. V. (1986). *Lexicology of modern English: A textbook for students of institutes and faculties of foreign languages*. Moscow: Graduate School.
- Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Ayres, C. (1927). Para a história da Academia das Ciências de Lisboa. *Boletim da Segunda Classe* 13, pp. 1–544.
- Baalbaki, R. (2014). *The arabic lexicographical tradition: From the 2nd/8th to the 12th/18th Century*. Leiden: Brill.
- Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., & Summers, E. (2013). Key Choices in the Design of Simple Knowledge Organization System (SKOS). *Journal of Web Semantics*, 20, 35–49. doi:10.1016/j.websem.2013.05.001.
- Bakhtin, M. (1992). *Estética da criação verbal*. São Paulo: Martins Fontes.
- Baldinger, K. (1960). Alphabetisches oder begrifflich gegliedertes Wörterbuch? *Zeitschrift für romanische Philologie*, 76, 521–536.
- Baldwin, T., & Kim, S. N. (2010). Multiword expressions. In N. Indurkha & F. J. Damerau (Eds.), *Handbook of natural language processing* (2nd ed.) (pp. 267–292). Boca Raton: CRC Press.
- Bański, P., Bowers, J., & Erjavec, T. (2017). TEI Lex-0 guidelines for the encoding of dictionary information on written and spoken forms. In Kosem, I., Tiberius, C., Jakubíček, M., Kallas, J., Krek, S., & Baisa, V. (Eds.), *Electronic Lexicography in the*

21st Century: Proceedings of eLex 2017 Conference (pp. 485–494). Brno: Lexical Computing CZ s.r.o.

- Beaujot, J.-P. (1989). Dictionnaire et idéologie. In F. J. Hausmann et al. (Eds.), *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie Internationale de Lexicographie*, vol. 1 (pp. 79–88). Berlin: Walter de Gruyter.
- Béjoint, H. (1988). Scientific and technical words in general dictionaries. *International Journal of Lexicography*, 1(4), 354–368. doi:10.1093/ijl/1.4.354.
- Béjoint, H. (2000). *Modern lexicography: An introduction*. Oxford: Oxford University Press Inc.
- Bentivogli, L., Forner, P., Magnini, B., & Pianta, E. (2004). Revising the wordnet domains hierarchy: semantics, coverage and balancing. In *Proceedings Workshop on Multilingual Linguistic Resources, MLR'04* (pp. 101–108), Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bergenholtz, H., & Gouws, R. H. (2012). What is lexicography?. *Lexicos*, 22, 31–42. doi:10.5788/22-1-996 .Bergenholtz, H., & Tarp, S. (1995). *Manual of specialised lexicography: The preparation of specialised dictionaries*. Amsterdam: John Benjamins Publishing. doi:10.1075/btl.12.
- Bergenholtz, H., & Tarp, S. (1995). *Manual of Specialised Lexikography*. Preparation of LSP dictionaries-problems and suggested solutions. Amsterdam–Philadelphia: John Benjamins.
- Bergenholtz, H., & Tarp, S. (2003). Two opposing theories: On H. E. Wiegand's recent discovery of lexicographic functions. *Hermes*, 31, 171–196. doi:10.7146/hjlc.v16i31.25743.
- Bergenholtz H., Nielsen, S., & Tarp, S. (2009). *Lexicography at a crossroads. dictionaries and encyclopedias today. Lexicographical tools tomorrow*. Bern: Peter Lang.
- Berry, D. M., & Fagerjord, A. (2017). *Digital humanities: Knowledge and critique in a digital age*. Cambridge: Polity Press.
- Biderman, M. T. C. (1984). A ciência da lexicografia. *Alfa*, 28, 1–26.
- Blecua, J. M. (2006). *Principios del diccionario de autoridades*. Madrid: Real Academia Española. Retrieved from https://www.rae.es/sites/default/files/Discurso_Ingreso_Jose_Manuel_Blecua.pdf.
- Bogaards, P. (2010). Lexicography: science without theory? In Schryver G.-M. (Ed.), *A way with words (festschrift for Patrick Hanks)* (pp. 313–322). Kampala, Uganda: Menha Publishers.
- Bohbot, H., Frontini, F., Luxardo, G., Khemakhem, M., & Romary, L. (2018). Presenting the Nénufar Project: A diachronic digital edition of the Petit Larousse Illustré. In GLOBALEX 2018 – *Globalex workshop at LREC2018*, May 2018, Miyazaki, Japan (pp. 1–6). Retrieved from <https://hal.archives-ouvertes.fr/hal-01728328>.

- Bosque-Gil, J., Gracia, J., & Gómez-Pérez, A. (2016a). Linked data in lexicography. *Kernerman Dictionary News*, 24:19–24. Retrieved from https://lexicala.com/wp-content/uploads/2021/03/kdn24_2016.pdf.
- Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E., & Aguado-de Cea, G. (2016b). Modelling multilingual lexicographic resources for the web of data: The K dictionaries case. In Kernerman I., Kosem I., Krek S., & Trap-Jensen L. (Eds.), *GLOBALEX 2016 – Lexicographic Resources for Human Language Technology Workshop Programme* (pp. 65–72). [s.n.]: [s.l.]. Retrieved from http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-GLOBALEX_Proceedings-v2.pdf.
- Bosque-Gil, J., Lonke, D., Gracia, J., & Kernerman, I. (2019). Validating the ontolex-lemon lexicography module with k dictionaries' multilingual data. In Kosem, I. et al (Eds.), *Electronic lexicography in the 21st century. Proceedings of eLex 2019 conference. 1–3 October 2019, Sintra, Portugal* (pp. 726–746). Brno: Lexical Computing CZ, s.r.o. Retrieved from https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_41.pdf.
- Bothma, T. J. D. (2017). Lexicography and information science. In Fuertes-Olivera, P. A. (Ed.), *Routledge handbook of lexicography*. London: Routledge.
- Boulanger, J.-C. (2001). L'aménagement des marques d'usage technolèctales dans les dictionnaires généraux bilingues, dans Les dictionnaires de langue française. In *Dictionnaire d'apprentissage, dictionnaires spécialisés de la langue, dictionnaires de spécialité* (pp. 247–271). Paris: Honoré Champion éditeur.
- Boulanger, J.-C., & L'Homme, M.-C. (1991). Les technolèctes dans la pratique dictionnaire générale. Quelques fragments d'une culture. *Meta*, 36(1), 23–40. doi:10.7202/002113ar.
- Bourdieu, P., Dauncey, H., & Hare, G. (1998). The state, economics and sport. *Culture Sport Society*, 1(2), 15–21. doi:10.1080/14610989808721813.
- Bowers, J., Herold, A., Romary, L., Tasovac, T. (2021). TEI Lex-0 Etym – Towards terse recommendations for the encoding of etymological information. Preprint. Retrieved from <https://halinria.fr/hal-03108781>.
- Bowker, L. (2017). Lexicography and terminology. In Fuertes-Olivera, P. A. (Ed.), *The Routledge Handbook of Lexicography*. London: Routledge. Doi:10.4324/9781315104942.ch9.
- Bray, L. (1990). La lexicographie française des origines à Littré. In Hausmann, F. J. et al. (Eds.), *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie Internationale de Lexicographie*, vol. 2 (pp. 1789–1818). Berlin: Walter de Gruyter.
- Budin, G., Majewski, S., & Mörth, K. (2012). Creating lexical resources in TEI P5. A schema for multi-purpose digital dictionaries. *Journal of the Text Encoding Initiative [Online]*, 3. doi:10.4000/jtei.522.

- Burada, M., & Sinu, R. (Eds.). (2020). *A local perspective on lexicography: Dictionary research, practice, and use in Romania*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Burke, P. (2010). *Languages and communities in early modern Europe*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511617362.
- Cabré, M. T. (1994). Terminologie et dictionnaires. *Meta*, 39(4), 589–597. doi:10.7202/002182ar.
- Cabré, M. T. (1995). La terminología hoy: Concepciones, tendencias y aplicaciones. *Ciência da Informação*, 24(3). Retrieved from <http://revista.ibict.br/ciinf/article/view/567>.
- Cabré, M. T. (1998). El discurs especialitzat o la variació funcional determinada per la temàtica: Noves perspectives. *Caplletra: Revista Internacional de Filologia*, 25, 173–193.
- Cabré, M. T. (1999). *Terminology: Theory, methods and applications*. Amsterdam: John Benjamins. doi:10.1075/tlrp.1.
- Cabré, M. T. (2003). Theories of terminology: their description, prescription and explanation. *Terminology*, 9(2), 163–199. doi:10.1075/term.9.2.03cab.
- Calzolari, N., Zampolli, A., & Lenci, A. (2002). Towards a Standard for a Multilingual Lexical Entry: the EAGLES/ISLE Initiative. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing. Third International Conference, CICLing 2002, Mexico City, Mexico, February 17–23, 2002 Proceedings* (pp. 264–279). Berlin / New York: Springer-Verlag. doi:10.1007/3-540-45715-1.
- Candel, D. (1979). La présentation par domaines des emplois scientifiques et techniques dans quelques dictionnaires de langue. *Langue française*, 43, 100–118. doi:10.3406/lfr.1979.6165.
- Carras, C. (2002). *Le vocabulaire économique et commercial dans la presse brésilienne (années 1991–1992): étude comparative et proposition de dictionnaire bilingue portugais / français* (Doctoral thesis, Université Lyon II). Retrieved from http://theses.univ-lyon2.fr/documents/lyon2/2002/carras_c.
- Carrère d'Encausse, H., Broglie, G., Dotoli, G., & Selvaggio, M. (Eds.) (2017). *Le dictionnaire de l'Académie française. Langue, littérature, société*. Paris: Hermann Éditeurs.
- Carrington da Costa, J. C. S. (1931). *O Paleozóico português. (Síntese e crítica)*. (Doctoral dissertation. Universidade do Porto.
- Casares, J. (1982). *Introducción a la lexicografía moderna*. Madrid: Editorial CSIC.
- Casteleiro, J. M. (1981). Estudo linguístico do 1.º dicionário da Academia. *Memórias da Academia das Ciências de Lisboa*, 22, 47–67.
- Casteleiro, J. M. (2008). Actividades lexicográficas da Academia das Ciências de Lisboa. In González Seoane, E., Santamarina, A., & Varela Barreiro, X. (Ed.), *A lexicografía galega moderna. Recursos e perspectivas* (pp. 315–322). Santiago de Compostela: Consello da Cultura Galega; Instituto da Lingua Galega.

- Cimiano, P., McCrae, J. P., & Buitelaar, P. (2016). Lexicon Model for Ontologies: Community Report. W3C Community Group Final Report. Retrieved from <https://www.w3.org/2016/05/ontolex/>.
- Coelho, J. P. (1974). Plano a que obedece o dicionário Académico. *Boletim da Academia das Ciências de Lisboa*, 31, 247–259.
- Cohen, K. M., Finney, S. C., Gibbard, P. L. & Fan, J.-X. (2017). The ICS International chronostratigraphic Chart. Episodes 36: 199-204. Retrieved from: <https://stratigraphy.org/ICSchart/ChronostratChart2017-02PTPortuguese.pdf>.
- Cohen, K. M., Finney, S. C., Gibbard, P. L. & Fan, J.-X. (2021). The ICS International Chronostratigraphic Chart, v 2021/07. Episodes 36: 199–204. Retrieved from: <https://stratigraphy.org/ICSchart/ChronostratChart2021-07.pdf>.
- Collinot, A., & Mazière, F. (1997). *Un prêt à parler: le dictionnaire*. Paris: Presses Universitaires de France.
- Considine, J. (2014). *Academy dictionaries 1600–1800*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107741997.
- Correia, M. (2008). Lexicografia no início do século xxi – Novas perspectivas, novos recursos e suas consequências. In Júnior, M. A. (Ed.), *Lexicon – Dicionário de Grego-Português, Actas de colóquio* (pp. 73–85). Lisboa: Centro de Estudos Clássicos.
- Correia, M. (2009). *Os dicionários portugueses*. Lisboa: Editorial Caminho.
- Costa, R. (2006a). Texte, terme et contexte. In Blampain, D., Thoiron, P., & Van Campenhoudt, M. (Eds.), *Mots, termes et contextes. Actes des VII Journées Scientifiques du Réseau Lexicologie, Terminologie et Traduction* (pp. 79–88). Paris: Éditions des Archives Contemporaines.
- Costa, R. (2006b). Plurality of theoretical approaches to terminology. In Picht, H. (Ed.), *Modern approaches to terminological theories and applications* (pp. 77–89). Bern: Peter Lang.
- Costa, R. (2013). Terminology and Specialised Lexicography: two complementary domains. *Lexicographica*, 29(1), 29–42. doi:10.1515/lexi-2013-0004.
- Costa, R. (2021). Terminology in the Digital Age: the Ontological Turn: Part 2. TOTh Training School 2021, 1–2 June 2021, France, Université Savoie Mont Blanc. Bourget du Lac.
- Costa, R., Carvalho, S., Salgado, A., Simões, A., & Tasovac, T. (2020). Ontologie des marques de domaines appliquée aux dictionnaires de langue générale. In Blanco, X. (Ed.), *La lexicographie en tant que méthodologie de recherche en linguistique. Langue(s) et parole [Special issue]*. *Revue de Philologie Française et Romane*, 5, 201–230. Retrieved from <https://raco.cat/index.php/Langue/article/view/379305>.
- Costa, R., Ramos, M., Salgado, A., Carvalho, S., & Almeida, B., & Silva, R. (2021b, forthcoming). Neoterm or neologism? A closer look at the determinologisation process. In *Proceedings of 3rd Globalex Workshop on Lexicography and Neology*. Adelaide, Australia.

- Costa, R., Salgado, A., & Almeida, B. (2021a). SKOS as a Key Element for Linking Lexicography to Digital Humanities. In K. Golub, & Liu, Y. (Eds.), *Information and knowledge organisation*. Routledge. ISBN 9780367675516.
- Costa, R., Salgado, A., & Almeida, B. (2021b). Going digital: the case of a historical Portuguese lexicographical resource. In EADH2021 'Interdisciplinary Perspectives on Data', *2nd International Conference of the European Association for Digital Humanities* (EADH) – Krasnoyarsk (Russia), 21–25 September 2021.
- Costa, R., Salgado, A., Khan, A., Carvalho, S., Romary, L., Almeida, B., Ramos, M., Khemakhem, M., Silva, R., & Tasovac, T. (2021c). MORDigital: the advent of a new lexicographical Portuguese project. In I. Kosem et al. (Eds.), *Electronic lexicography in the 21st century: post-editing lexicography. Proceedings of the eLex 2021 conference* (pp. 312–324). Brno: Lexical Computing CZ. ISSN 2533-5626.
- Cowie, A. P. (1994). Phraseology. In Asher, R. E. (Ed.), *The encyclopedia of language and linguistics* (pp. 3168–3171). Oxford, UK: Pergamon.
- Cunha, P. P., Lemos de Sousa, M. J., Pinto de Jesus, A., Rodrigues, C. F., Telles Antunes, M., & Tomás, C. A. (2012). O carvão em Portugal: Geologia, petrologia e geoquímica. In M. J. Lemos de Sousa, C. F. Rodrigues & M. A. P. Dinis (Eds.), *O carvão na actualidade*, vol. 1, *Petrologia, métodos analíticos, classificação e avaliação de recursos e reservas, papel no contexto energético, carvão em Portugal* (pp. 309–381), Porto, Lisboa: Universidade Fernando Pessoa, Academia das Ciências de Lisboa.
- D'Alembert, J. R. (1751). Discours préliminaire des éditeurs. In Diderot, D., & D'Alembert, J. R. (Eds.), *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers, etc.*, University of Chicago: ARTFL Encyclopédie Project (Spring 2021 Edition), Robert Morrissey and Glenn Roe (Eds). Retrieved from <https://encyclopedia.uchicago.edu/node/88>.
- Dantas, J., (1936). As nomenclaturas científicas no Dicionário da Academia. In *Memórias, Classe de Letras* (pp. 301–303), tomo 2. Lisboa: Academia das Ciências de Lisboa.
- De Bessé, B. (1990). La définition terminologique. In Chaurand, J. & Mazière, F. (Ed.), *Actes du Colloque la Définition, organisé par le CELEX (Centre d'études du Lexique) de l'Université Paris-Nord (1988)* (pp. 252–261). Paris: Larousse.
- De Bessé, B. (2000). Le domaine. In H. Béjoint & P. Thoiron (Eds.), *Le sens en terminologie* (pp. 182–197). Lyon: Presses Universitaires de Lyon.
- Declerck, T., McCrae, J., Navigli, R., Zaytseva, K., & Wissik, T. (2019). ELEXIS – European lexicographic infrastructure: Contributions to and from the linguistic linked open data. In Kernerman, I., & Simon, K (Eds.), *Proceedings of the 2nd GLOBALEX Workshop. GLOBALEX (GLOBALEX-2018) Lexicography & WordNet located at 11th Language Resources and Evaluation Conference (LREC 2018)*, Miyazaki Japan (pp. 17–22). Paris: ELRA. Retrieved from https://www.dfki.de/fileadmin/user_upload/import/9709_elexis-european-lexicographic.pdf.
- Decreto-Lei n. 157/2015, de 10 de Agosto de 2015. Estatutos da Academia de Ciências de Lisboa.

- Decreto-Lei n. 390/87, de 31 Dezembro de 1987. Estatutos da Academia das Ciências de Lisboa.
- Delavigne, V. (2002). Le domaine aujourd'hui. Une notion à repenser. In Candel, D. (Ed.), *Le traitement des marques de domaine en terminologie*. Retrieved from <https://hal.archives-ouvertes.fr/hal-00924228/>.
- Depecker, L. (2003). Entre signe et concept. Eléments de terminologie générale. In Candel, D., *Le traitement des marques de domaine en terminologie*. Paris: Presses de la Sorbonne Nouvelle.
- Derouin, M.-J., & Le Meur, A. (2002). Ongoing changes in lexicographical international standards: Report on the revision of ISO 151 lexicographical symbols and typographical conventions for use in terminography and proposals for the first draft: Presentation/representation of entries in dictionaries. In Braasch, A., & Povlsen, C. (Eds.), *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002. Copenhagen, Denmark, August 13–17, 2002*, vol. 2 (pp. 689–696). [S.l.]: Center for Sprogteknologi. Retrieved from https://www.euralex.org/elix_proceedings/Euralex2002/.
- Derouin, M.-J., & Le Meur, A. (2006). ISO 151: A revised standard for lexicography. *Kernerman Dictionary News*, no. 14, July 2006. Retrieved from <https://www.kdictionaries.com/kdn//2006/ISO%20151%20%20A%20revised%20standard%20for%20lexicography%20-%20Andr%C3%A9%20Le%20Meur%20and%20Marie-Jeanne%20Derouin.pdf>.
- Derouin, M.-J., & Le Meur, A. (2008). Presentation of the new ISO-Standard for the representation of entries in dictionaries: ISO 151. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May–1 June 2008, Marrakech, Morocco* (pp. 754–757). [S.l.]: European Language Resources Association. Retrieved from <http://www.lrec-conf.org/proceedings/lrec2008/summaries/190.html>.
- Devapala, S. (2004). Typological Classification of Dictionaries. *The Asia Lexicography Conference, 24–26 May*. Chiangmai, Thailand.
- Dias, J. A. (2018). *A Academia Real das Ciências de Lisboa (1779–1834) – Ciências e hibridismo numa periferia europeia*. Lisboa: Colibri.
- Dubois, C. (1990). Considérations generales sur l'organisation du travail lexicographique. *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie Internationale de Lexicographie*, vol. 2. (pp. 1574–158). Berlin: Walter de Gruyter.
- Dubois, C., & Dubois, J. (1971). *Introduction à la lexicographie: le dictionnaire*. Paris: Librairie Larousse.
- Dubois, J. (1962). Recherches lexicographiques: esquisse d'un dictionnaire structural. *Études de linguistique appliquée* 1, 43–48.
- Dubois, J. (1970). Dictionnaire et discours didactique, *Langages*, 5(19), 35–47. Doi:10.3406/lgge.1970.2590.

- Eco, U. (2001). *Semiótica e Filosofia da Linguagem*. Lisboa: Instituto Piaget.
- Ehrmann, M., Ceconi, F., Vannella, D., McCrae, J. P., Cimiano, P., & Navigli, R. (2014). A Multilingual Semantic Network as Linked Data: lemon-BabelNet.
- Engelberg, S., & Lemnitzer, L. (2009). *Lexikographie und Wörterbuchbenutzung* (5th ed.). Tübingen: Stauffenburg.
- Englund, R., & Nissen, H. (1993). *Die lexikalischen Listen der archaischen Texte aus Uruk* (ATU 3).
- Estopà, R. B. (1998). El léxico especializado en los diccionarios de lengua general: las marcas temáticas. *Revista de la Sociedad Española de Lingüística*, 28(2), 359–387.
- Faber, P. (2009). The cognitive shift in terminology and specialized translation. *MonTi: Monografías de Traducción e Interpretación*, 1, 107–134. doi:10.6035/MonTI.2009.1.5.
- Faber, P. (Ed.). (2012). *A cognitive linguistics view of terminology and specialized language*. Berlin, Boston: De Gruyter. doi:10.1515/9783110277203.
- Faber, P. (2015). Frames as a framework for terminology. In Kockaert, H. J. & Steurs F. (Eds.), *Handbook of terminology*, vol. 1 (pp. 14–33). Amsterdam: John Benjamins Publishing Company. doi:10.1075/hot.1.fra1.
- Fajardo, A. (1994). La marcación técnica en la lexicografía española. *Revista de Filología de la Universidad de La Laguna*, 13, 131–143.
- Fajardo, A. (1996/1997). Las marcas lexicográficas: concepto y aplicación práctica en la lexicografía española. *Revista de Lexicografía*, 3, 31–57. A Coruña: Universidade da Coruña.
- Fedorova, I. V. (2004). Style and usage labels in learner's dictionaries: Ways of optimization. In Williams, G., & Vessier, S. (Eds.), *Proceedings of the 11th EURALEX International Congress* (pp. 265–272). Lorient: Université de Bretagne-Sud, Faculté des Lettres et des Sciences Humaines.
- Felber, H. (1987). *Manuel de Terminologie*. Paris: UNESCO, Infoterm.
- Fellbaum, C. (2016). Treatment of multi-word units. In Durkin, P. (Ed.), *The oxford handbook of lexicography* (pp. 411–424). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199691630.001.0001.
- Fish, S. (2018). Stop trying to sell the humanities. *The Chronicle of Higher Education*, 64(36). Retrieved from <https://www.chronicle.com/article/stop-trying-to-sell-the-humanities>.
- Fleury, E. (1922). *O que pode ler-se na Carta Geológica de Portugal*. Separata do *Jornal de Ciências Naturais*, Volume I, 1921. Lisboa: Biblioteca Nacional.
- Fontenelle, T. (1997). *Turning a bilingual dictionary into a lexical-semantic database*. Tübingen: Niemeyer.
- Forcada, M., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J., Tyers, F. (2011). Apertium: A free/open-source platform for rule-based machine

- translation. *Machine Translation*, 25(2), 127–144. Retrieved August 25, 2021, from <http://www.jstor.org/stable/41487458>.
- France, A. (1921). *Léxique*. In *La Vie littéraire*, vol. 2 (pp. 275–283). Paris: Calmann-Lévi.
- Francopoulo, G. (Ed.). (2013). *LMF – Lexical Markup Framework*. London: ISTE/Wiley.
- Francopoulo, G., Bel, N., George, M., Calzolari, N., Monachini, M., Pet, M., Soria, C. (2006). Lexical markup framework (LMF) for NLP multilingual resources. In Witt, A., Sérasset, G., Armstrong, S., Breen, J., Heid, U., Sasaki, F., (Eds.), *Proceedings of the Workshop on Multilingual Language Resources and Interoperability; 2006 Jul 23; Sydney, Australia* (pp. 1–8). Stroudsburg, PA: Association for Computational.
- Frawley, W. (1989). The dictionary as text. *International Journal of Lexicography*, 2(3), 231–248. doi:10.1093/ijl/2.3.231.
- Fuertes-Olivera, P. A., & Bergenholtz, H. (2011). Introduction: The construction of internet dictionaries. In Fuertes-Olivera, P. A., & Bergenholtz H. (Eds.), *E-Lexicography: The internet, digital initiatives and lexicography* (pp. 1–16). London and New York: Continuum. doi:10.5040/9781474211833.0005.
- Fuertes-Olivera, P. A., & Tarp, S. (2008). La teoría funcional de la lexicografía y sus consecuencias para los diccionarios de economía del español. *Revista de Lexicografía* 14, 89–109.
- Furetière, A. (1685). *Factum pour Messire Antoine Furetière, abbé de Chalivoy, contre quelques uns de l'Académie Française*. Amsterdam: H. Desbordes.
- Furetière, A. (1690). *Dictionnaire Universel, contenant généralement tous les mots françois tant vieux que modernes, et les termes de toutes les sciences et des arts*. La Haye/Rotterdam: Arnout & Reinier Leers.
- Galisson, R. (1978). *Recherches de lexicologie descriptive: la banalisation lexicale. Le Vocabulaire du football dans la presse sportive. Contribution aux recherches sur les langues techniques*. Paris: Nathan.
- Gantar, P., Colman, L., Parra Escartín, C., & Martínez Alonso, H. (2018). Multiword expressions: Between lexicography and NLP. *International Journal of Lexicography*, 32(2), 138–162. doi:10.1093/ijl/ecy012.
- Gapporov, B., Vositov, V., & Ibragimova, G. (2020). Typological classification of dictionaries. *ISJ Theoretical and Applied Science*, 1(81), 581–584.
- García de la Concha, V. (2014). *La Real Academia Española. Vida e historia*, Madrid: Real Academia Española.
- Gaudin, F. (1990). Socioterminology and expert discourses. In *TKE'90: Terminology and knowledge engineering*, vol. 2 (pp. 631–641). Retrieved from: hal-01090697.
- Gaudin, F. (2007). *Socioterminologie: une approche sociolinguistique de la terminologie*. Bruxelles: Duculot.
- Geeraerts, D. (1984). Dictionary classification and the foundations of lexicography. *I.T.L. Review*, 63(1), 37–63. doi:10.1075/itl.63.03gee.

- Geeraerts, D., & Janssens, G. (1982). *Wegwijs in woordenboeken. Een kritisch overzicht van de lexicografie van het Nederlands*. Assen: Van Gorcum.
- Gella, S., Strapparava, C., Nastase, V. (2014). Mapping WordNet domains, WordNet topics and Wikipedia categories to generate multilingual domain specific resources. In Calzolari, N., Choukri, K., Declerck, T., et al. (Eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)* (pp. 1117–1121). Reykjavik: European Language Resources Association (ELRA). Retrieved from http://www.lrec-conf.org/proceedings/lrec2014/pdf/122_Paper.pdf.
- Gershuny, H. (1974). Sexist semantics in the dictionary. *ETC: A Review of General Semantics*, 31(2), 159–169. Retrieved from <http://www.jstor.org/stable/42576397>.
- Godfrey-Smith, P. (2009). Models and fictions in science. *Philosophical Studies*, 143, 101–116. doi:10.1007/s11098-008-9313-2.
- Gold, M. K., & Klein, L. F. (Eds.) (2016). *Debates in the Digital Humanities*. Mineápolis: University of Minnesota Press.
- Gonçalves, M. F. (2002). As ‘autoridades’ no Vocabulário Português e Latino (1712-1728) de D. Rafael Bluteau. Retrieved from <https://dspace.uevora.pt/rdpc/bitstream/10174/8802/1/As%20%E2%80%9CAutoridades%E2%80%9D%20no%20Vocabulário%20Português%20e%20Latino%20%281712-1728%29.htm>.
- Gonçalves, M. F., & Banza, A. P. (Eds.) (2013). *Património Textual e Humanidades Digitais: da antiga à Nova Filologia* (pp. 73–111). Évora: CIDEHUS. doi:10.4000/books.cidehus.1088.
- Gouws, R. H. (2005). Meilensteine auf dem historischen Weg der Metalexikographie. *Lexicographica* 21, 158–178. doi:10.1515/9783484604742.158.
- Gouws, R. H. (2011). Learning, unlearning and innovation in the planning of electronic dictionaries. In Fuertes-Olivera, P. A., & Bergenholtz H. (Eds.), *E-Lexicography: The internet, digital initiatives and lexicography* (pp. 17–29). London and New York: Continuum. doi:10.5040/9781474211833.ch-001.
- Gouws, R. H. (2020). Special field and subject field lexicography contributing to lexicography. *Lexikos*, 30, 1–28. doi:10.5788/30-1-1568.
- Gouws, R. H., & Prinsloo, D. J. (2005). *Principles and practices of South African lexicography*. Stellenbosch, South Africa: African Sun Media.
- Gouws, R. H., Heid, U., Schweickard, W., & Wiegand, H. E. (Eds.). (2014). *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent developments with focus on electronic and computational Lexicography*. Berlin, Boston: De Gruyter Mouton. doi:10.1515/9783110238136.
- Granger, H. (1983). Aristotle on genus and differentia in the topics and categories. *The Society for Ancient Greek Philosophy Newsletter*, 106, 1-23. Retrieved from <https://orb.binghamton.edu/sagp/106/>.

- Granger, S. (2012). Introduction: Electronic lexicography – from challenge to opportunity. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography* (pp. 1–11). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199654864.003.0005.
- Grazzini, G. (1991). *L'Accademia della Crusca, Firenze* (4th ed.). Firenze: Nencioni.
- Guerra Salas, L., & Gómez Sánchez, M. (2005). El léxico especializado en los diccionarios monolingües de ELE. In Castillo Carballo, M. A., Cruz Moya, O., García Platero, J. M., & Mora Gutiérrez, J. P. (Eds.), *Actas del XV Congreso de Asele. Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: Deseo y realidad* (pp. 427–434). Sevilla: Universidad de Sevilla. Retrieved from https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/15/15_0425.pdf.
- Guerrero Ramos, G., & Pérez Lagos, M. F. (2017). La definición en el diccionario desde la teoría lingüística. *Pragmalingüística*, 25, (286–310). <https://doi.org/10.25267/Pragmalinguistica.2017.i25.15>.
- Guilbert, L. (1973). La spécificité du terme scientifique et technique. In Guilbert, L., and Peytard & J. Les vocabulaires techniques et scientifiques [Numéro thématique]. *Langue française* 17, 5–17. doi:10.3406/lfr.1973.5617.
- Guilbert, L. (1975). *La créativité lexicale*. Paris: Larousse.
- Haensch, G. (1997). *Los diccionarios del español en el umbral del siglo XXI*. Salamanca: Ediciones Universidad de Salamanca.
- Haensch, G., Wolf, L., Ettinger, S., & Werner, R. (1982). *La lexicografía (De la lingüística teórica a la lexicografía práctica)*. Gredos: Madrid.
- Harris, R., & Hutton, C. (2007). *Definition in theory and practice: Language, lexicography and the Law*. London and New York: Continuum.
- Hartmann, R. R. K. (2005). Pure or hybrid? The development of mixed dictionary genres. *Facta Universitatis. Linguistics and literature*, 3(2), 193–208. Retrieved from <http://facta.junis.ni.ac.rs/lal/lal2005/lal2005-06.pdf>.
- Hartmann, R. R. K. (Ed.) (2003). *Lexicography: critical concepts*, vol. 1, *Dictionaries, compilers, critics and users*. London: Taylor & Francis.
- Hartmann, R. R. K., & James, G. (1998/2002). *Dictionary of Lexicography*. London and New York: Routledge/Taylor and Francis.
- Hausmann et al. (Eds.). (1989). *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie Internationale de Lexicographie*, vol. 1. Berlin: Walter de Gruyter.
- Hausmann et al. (Eds.). (1990). *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie Internationale de Lexicographie*, vol. 2. Berlin: Walter de Gruyter.
- Hausmann et al. (Eds.). (1991). *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of*

Lexicography/Encyclopédie Internationale de Lexicographie, vol. 3. Berlin: Walter de Gruyter.

- Hausmann, F. J. (1989). Die Markierung in einem allgemeinen einsprachigen Wörterbuch: eine Übersicht. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, L. Zgusta (Eds.), *Wörterbücher. Ein internationales Handbuch zur Lexikographie* (pp. 649–657). Berlin: Walter de Gruyter.
- Hausmann, F. J., & Wiegand, H. E. (1989). Component parts and structures of general monolingual dictionaries: A survey. In F. J. Hausmann, O. Reichmann, H. E. Wiegand and E. Zgusta (Eds.), *Wörterbücher. Ein internationales Handbuch zur Lexikographie* (pp. 328–360). Berlin: Walter de Gruyter.
- Holm, P., Jarrick, A., & Scott, D. (2015). *Humanities world report 2015*. Springer. doi:10.1057/9781137500281.
- Hulbert, J. R. (1955). *Dictionaries British and American*. London: Deutsch.
- Humbley, J. (2002). Nouveaux dictionnaires, nouveaux rapports avec les utilisateurs. *Meta*, 47(1), 95–104. doi:10.7202/007994ar.
- Humbley, J., & Candell, D. (1997). Explorations terminologiques dans un dictionnaire de langue, domaine: géologie. In Lapierre, L., Oore, I.; Runte, H. R. (Eds.), *Mélanges de linguistique offerts à Rostislav Kocourek* (pp. 35–48). Halifax: Les Presses de l'Alpha.
- Iamartino, G. (2014). Lexicographers as censors: Checking verbal abuse in early english dictionaries. In Iannaccaro, G., & Iamartino, G. (eds), *Enforcing and eluding censorship: British and Anglo-Italian perspectives* (pp. 168–196). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Iamartino, G. (2020). Lexicography as a mirror of society: Women in John Kersey's dictionaries of the English language, in *Textus. English Studies in Italy*, 1, 35–67, doi:10.7370/97351.
- Ide, N. M., & Véronis, J. (1995). *Text Encoding Initiative: Background and Contexts*. Cambridge, MA: The MIT Press.
- Ide, N., & Romary, L. (2007). *A formal model of dictionary structure and content*. Brighton: University of Brighton.
- Illson, R. (2012). IJL: The first ten years – And beyond. *International Journal of Lexicography* 25(4), 381–385.
- Iriarte Sanromán, A. (2001). *A unidade lexicográfica. Palavras, colocações, frasesmas, pragmatemas*. Braga: Centro de Estudos Humanísticos – Universidade do Minho.
- Iriarte Sanromán, A. (2015). Reverse search in electronic dictionaries. In J. P. Silvestre & A. Villalva (Eds.), *Planning Non-Existent Dictionaries* (pp. 153–162). Lisboa/Aveiro: Centro de Linguística da Universidade de Lisboa/Universidade de Aveiro.
- ISO 1087. (2019). *Terminology Work – Vocabulary – Part 1: Theory and Application*. Geneva: International Organization for Standardization.

- ISO 1951. (1973). Lexicographical symbols particularly for use in classified defining vocabularies.
- ISO 1951. (1997). Lexicographical symbols and typographical conventions for use in terminography. Geneva: International Organization for Standardization.
- ISO 1951. (2007). *Presentation/representation of entries in dictionaries – Requirements, recommendations and information*. Geneva: International Organization for Standardization.
- ISO 24613. (2008). *Language resource management - Lexical markup framework (LMF)*. Geneva: International Organization for Standardization.
- ISO 24613-1. (2019). *Language resource management – Lexical markup framework (LMF) – Part 1: Core model*. Geneva: International Organization for Standardization.
- ISO 24613-2. (2020). *Language resource management – Lexical markup framework (LMF) – Part 2: Machine Readable Dictionary (MRD) model*. Geneva: International Organization for Standardization.
- ISO 24613-3. (2021). *Language resource management – Lexical Markup Framework (LMF) – Part 3: Etymological Extension*. Geneva: International Organization for Standardization.
- ISO 24613-4. (2021). *Language resource management – Lexical Markup Framework (LMF) – Part 4: TEI serialisation*. Geneva: International Organization for Standardization.
- ISO 24613-5. (2018). *Language resource management – Lexical markup framework (LMF) – Part 5: Lexical base exchange (LBX) serialization*. Geneva: International Organization for Standardization.
- ISO 25964-1. (2011). *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*. Geneva: International Organization for Standardization.
- ISO 639-1. (2002). *Codes for the representation of names of languages – Part 1: Alpha-2 code*. Geneva: International Organization for Standardization.
- ISO 639-2. (1998). *Codes for the representation of names of languages – Part 2: Alpha-3 code*. Geneva: International Organization for Standardization.
- ISO 639-3. (2007). *Codes for the representation of names of languages – Part 3: Alpha-3 code*. Geneva: International Organization for Standardization.
- ISO 704. (2009). *Terminology work – Principles and methods*. Geneva: International Organization for Standardization.
- ISO/IEC 2382. (2015). *Information technology – Vocabulary*. Geneva: International Organization for Standardization.
- Jackson, H. (2002). *Lexicography: An introduction*. London and New York: Routledge.
- Jessen, A. (1996). *The presence and treatment of terms in general dictionaries*. M. A. Thesis. Ottawa: University of Ottawa.

- Johnson, S. (1747). *The plan of a dictionary of the English language*. London: Printed for J. and P. Knapton.
- Johnson, S. (1755). *A dictionary of the English language*. London: J. F., & C. Rivington.
- Jónsson, J. H. (2009). Lemmatisation of multiword lexical units: Motivation and benefits. In H. Bergenholtz, S. Nielsen & S. Tarp (Eds.), *Lexicography at a crossroads. Dictionaries and encyclopedias today, lexicographical tools tomorrow* (pp. 165–194). Bern: Peter Lang AG.
- Josselin-Leray, A., & Roberts, R. (2010). De la sélection des termes pour inclusion dans le dictionnaire général. Etat des lieux général et analyse critique de la terminologie informatique dans le New Oxford Dictionary of English (2000). In Hassan Hamzé (Ed.), *Le terme scientifique et technique dans le dictionnaire général. Actes de la 7è édition des RIL (Rencontres Internationales de Lexicographie)* (pp. 85–120). Beirut: Dar Wa Maktabat al-Hilal. Retrieved from <https://hal-univ-tlse2.archives-ouvertes.fr/hal-00983047>.
- Kallas, J., Koeva, S., Langemets, M., Tiberius, C., & Kosem, I. (2019). Lexicographic practices in Europe: Results of the ELEXIS survey on user needs. In Kosem, T., Kuhn, Z., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubíček, M., Krek, S., & Tiberius, C., (Eds.), *Electronic Lexicography in the 21st Century, Proceedings of the eLex 2019 Conference, Sintra, Portugal, 1–3 October 2019* pp. 519–536). Brno: Lexical Computing CZ, s.r.o. Retrieved from https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_30.pdf.
- Khan, A., Romary, L., Salgado, A., Bowers, J., Khemakhem, M., & Tasovac, T. (2020). Modelling etymology in LMF/TEI: The Grande Dicionário Houaiss da Língua Portuguesa Dictionary as a use case. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020), 11–16 May* (pp. 3172–3180). France: Marseille.
- Khan, F., & Salgado, A. (2021). Modelling lexicographic resources using CIDOC CRM, FRBROO and Ontolex Lemon. In A. Bikakis et al. (Eds.), *SWODCH 2021 – Semantic Web and Ontology Design for Cultural Heritage 2021. Proceedings of the International Joint Workshop on Semantic Web and Ontology Design for Cultural Heritage co-located with the Bolzano Summer of Knowledge 2021 (BOSK 2021)* (pp. 1–12). Bozen-Bolzano: CEUR-WS.
- Kilgarrieff, A. (1997). 'I don't believe in word senses'. In *Computers and the Humanities*, 31, 91–113. doi:10.1023/A:1000583911091.
- Kinable, D. (2015). Reflections on the concept of a scholarly dictionary. *Kernerman Dictionary News*, 23, 11–12. Retrieved from https://www.elexicography.eu/wp-content/uploads/2015/10/kdn23_21_20150507_kinable.pdf.
- Klein, K. (2015). Lexicology and lexicography. In Wright, J. D., *International Encyclopedia of the Social & Behavioral Sciences* (2nd Edition) (pp. 938–942). Elsevier. doi:10.1016/B978-0-08-097086-8.53059-1.
- Klimek, B. & Brümmer, M. (2015). Enhancing lexicography with semantic language databases. *Kernerman Dictionary News*, 23, 5–10. Retrieved from https://www.kdictionaries.com/kdn/kdn23_2015.pdf.

- Klosa, A., & Gouws, R. (2015). Outer features in e-dictionaries / Außentexte in Online-Wörterbüchern / Caractéristiques extérieures dans les dictionnaires en ligne. *Lexicographica*, 31(1), 142–172. <https://doi.org/10.1515/lexi-2015-0008>.
- Krumbein, W., & Sloss, L. (1963). *Stratigraphy and Sedimentation*. San Francisco: W. H. Freeman and Co.
- L’Affiche du Manifeste des Digital Humanities (2010). THATCamp Paris. Retrieved from <https://tcp.hypotheses.org/443>.
- L’Homme, M. C. (2004). *La terminologie: principes et techniques*. Montréal: Presses de l’Université de Montréal. Doi:10.4000/books.pum.10693.
- L’Homme, M-C., & Cormier, M. (2014). Dictionaries and the digital revolution: A Focus on users and lexical databases. *International Journal of Lexicography* 27(4), 331–340.
- Landau, S. I. (1974). Scientific and technical entries in American dictionaries. *American Speech* 49, 241–244. doi:10.2307/3087804.
- Landau, S. I. (2001). *Dictionaries. The art and craft of lexicography*. Cambridge: Cambridge University Press.
- Lara, L. F. (1997). *Teoría del diccionario monolingüe*. México: Colegio de México.
- Legoinha, P. (2008). Carbónico ou carbonífero, eis a questão! In Callapez, P., Rocha, R. B., Marques, J. F., Cunha, L. S., & Dinis, P. M. (Coords.). *A Terra – Conflitos e ordem: Homenagem ao Prof. António Ferreira Soares* (pp. 439–443). Coimbra: Museu Mineralógico e Geológico da Universidade de Coimbra.
- Lemnitzer, L., Romary, L., & Witt, A. (2013). Representing human and machine dictionaries in markup languages (SGML, XML). In Gouws, R., Heid, U., Schweickard, W., & Wiegand H. (Eds.), *Supplementary volume dictionaries. An International Encyclopedia of Lexicography* (pp. 1195–1209). Berlin: De Gruyter. doi:10.1515/9783110238136.1195.
- Lemos de Sousa, M. J. (1961). *A respeito de nomenclatura geológica*. Porto.
- Lemos de Sousa, M. J., Telles Antunes, M., & Salgado, A. (2015). Apresentação Geral. *Thesaurus de Ciências da Terra*. Academia das Ciências de Lisboa.
- Lépinette, B. (1990). Lexicographie bilingue et traduction. *Meta* 35(3), 571–581. doi:10.7202/003468ar.
- Leroyer, P. (2011). Change of paradigm in lexicography. From linguistics to information science and from dictionaries to lexicographic information tools. In Fuertes-Olivera, P. A., & Bergenholtz, H. (Eds.), *E-Lexicography: internet, digital initiatives and lexicography* (pp. 121–140). London and New York: Continuum. doi:10.5040/9781474211833.ch-006.
- Leroyer, P., & Simonsen, H. K. (2020). Reconceptualizing lexicography: the broad understanding. In Gavrilidou, Z., Mitsiaki, M., & Flaitouras, A. (Eds.), *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion* (vol. 1, pp. 183–192). Komotini: SynMorPhose Lab, Democritus University of Thrace. Retrieved from https://euralex2020.gr/wp-content/uploads/2020/11/EURALEX2020_ProceedingsBook-p183-192.pdf.

- Lew, R. (2007). Linguistic semantics and lexicography: A troubled relationship. In Fabiszak, M. (Ed.), *Language and meaning: cognitive and functional perspectives* (pp. 217–224). Frankfurt: Peter Lang.
- Lew, R. (2011). Space restrictions in paper and electronic dictionaries and their implications for the design of production dictionaries. In Banski, P., & Wojtowicz, B. (Eds.), *Issues in Modern Lexicography*. Retrieved from <https://www.semanticscholar.org/paper/Space-restrictions-in-paper-and-electronic-and-for-Lew/56446b2107374f86cce44ce6b23df9e6d530ec7c>.
- Lino, M. T. (1992). Lexicografia e terminologia. *Seminário, Português, Língua de Comunicação Internacional* (Conference presentation). Lisbon.
- Lino, T. (2018). Portuguese lexicography in the internet era. In Fuertes-Oliveira, P. A. (Ed.), *The Routledge handbook of lexicography*. Abingdon: Routledge.
- Lipoński, W. (2009). 'Hey, ref! Go, milk the canaries!' On the distinctiveness of the language of sport. *Studies in Physical Culture & Tourism*, 16, 19–36.
- Livet, Ch.-L. (1858). Article XXV des statuts. In Pellisson-Fontanier, P., & Olivet, P.-J., *Histoire de l'Académie Française*, édition augmentée et commentée, vol. 1. Paris: Chez J. B. Coignard.
- Löckinger, G., Kockaert, H. J., & Budin, G. (2015). Intensional definitions. In Hendrik J. Kockaert & Frida Steurs (Eds.), *Handbook of Terminology*, vol. 1 (pp. 60–81). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/hot.1.int1>.
- Lorentzen, H. (1996). Lemmatization of multi-word lexical units: In which entry? In M. Gellerstram et al. (Eds.), *Proceedings of the 7th EURALEX International Congress on Lexicography* (pp. 415–421). Goteborg, Sweden: Goteborg University Department of Swedish. Retrieved from <https://euralex.org/publications/lemmatization-of-multi-word-lexical-units-in-which-entry/>.
- Luhmann, J., & Burghardt, M. (2021). Digital humanities – A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology*, 1–24. doi:10.1002/asi.24533.
- Lynch, J. (2016). *You could look it up: The reference shelf from Ancient Babylon to Wikipedia*. New York: Bloomsbury Press.
- Magnini, B., & Cavaglià, G. (2000). Integrating subject field codes into WordNet. In Gavrilidou, M., Crayannis, G., Markantonatu, S., Piperidis, S., Stainhaouer, G. (Eds.), *Proceedings of LREC-2000, Second International Conference on Language Resources and Evaluation, Athens, Greece, 31 May–2 June 2000* (pp. 1413–1418). Retrieved from <http://www.lrec-conf.org/proceedings/lrec2000/pdf/219.pdf>.
- Malkiel, Y. (1962). A typological classification of dictionaries on the basis of distinctive features. In Householder, F. W., & Saporta, S. (Eds.), *Problems in lexicography* (Supplement to the *International Journal of American Linguistics*, 28, pp. 217–227). Bloomington: Indiana University.

- Malkiel, Y. (1976). *Etymological dictionaries. A tentative typology*. Chicago: University of Chicago Press.
- Margalitadze, T. (2018). Once again why lexicography is science. *Lexikos*, 28, 245–261. doi:10.5788/28-1-1464.
- Markoff, J. (2006). Entrepreneurs see a web guided by common sense. *The New York Times*. Retrieved from http://www.nytimes.com/2006/11/12/business/12web.html?_r=3&adxnnl=1&andoref=sl.
- Martelli, F., Navigli, R., Krek, S., Tiberius, C., Kallas, J., Gantar, P., Koeva, S., Nimb, S., Pedersen, B. S., Olsen, S., Langements, M., Koppel, K., Üksik, T., Dobrovolić, K., Ureña-Ruiz, R.-J., Sancho-Sánchez, J.-L., Lipp, V., Varadi, T., Györfy, A., László, S., Quochi, V., Monachini, M., Frontini, F., Tempelaars, R., Costa, R., Salgado, A., Čibej, J., & Munda, T. (2021). Designing the ELEXIS Parallel Sense-Annotated Dataset in 10 European Languages. In I. Kosem et al. (Eds.), *Proceedings of the eLex 2021 conference* (pp. 377–395). Brno: Lexical Computing CZ. ISSN 2533-5626.
- Martínez de Sousa, J. (1995). *Diccionario de lexicografía práctica*. Barcelona: Vox-Bibliograf.
- McCarty, W. (2015). Becoming interdisciplinary. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 69–83). West Sussex, UK: Wiley. doi:10.1002/9781118680605.ch5.
- McCracken, J. (2016). The exploitation of dictionary data and metadata. In Durkin, P. (Ed.), *The Oxford handbook of lexicography* (pp. 501–514). Oxford: Oxford University Press.
- McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P. & Cimiano, P. (2017). TheOntoLex-Lemon Model: development and applications. In *Proceedings of eLex 2017*, pages 587–597.
- McCrae, J. P., de Cea, G. A., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., & Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, 46(6), 701–709. doi:10.1007/s10579-012-9182-3.
- McCrae, J. P., Tiberius, C., Khan, A. F., Kernerman, I., Declerck, T., Krek, S., Monachini, M., & Ahmadi, S. (2019). The ELEXIS interface for interoperable lexical resources. In *Proceedings of the eLex 2019 conference. Biennial Conference on Electronic Lexicography (eLex-2019) Electronic lexicography in the 21st century. October 1–3 Sintra Portugal* (pp. 642–659). Brno: Lexical Computing CZ, s.r.o. Retrieved from https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_37.pdf.
- McCrae, J., Spohr, D., & Cimiano, P. (2011). Linking lexical resources and ontologies on the semantic web with Lemon. In Antoniou, G. (Ed.), *Proceedings of the 8th Extended Semantic Web Conference (ESWC)* (pp. 245–259). Berlin: Springer. doi:10.1007/978-3-642-21034-1_17.
- Meier, H. H. (1969). Lexicography as applied linguistics. *English Studies*, 50(1–6), 141–151. doi:10.1080/00138386908597328.

- Mel'čuk, I., & Polguère, A. (2018). Theory and practice of lexicographic definition. *Journal of Cognitive Science* 19(4), 417–470. doi:10.17791/jcs.2018.19.4.417.
- Mel'čuk, I., Arbatchewsky-Jumarie, N., Iordanskaja, L., Mantha, S., & Polguère, A. (1984/1999). *Dictionnaire Explicatif et Combinatoire du Français Contemporain.*, vol. IV, *Recherches lexico-sémantiques*. Montréal: Les Presses of l'Université de Montréal.
- Meyer, I., & Mackintosh, K. (1996). The corpus from a terminographer's viewpoint. *International Journal of Corpus Linguistics*, 1(2), 257–285. doi:10.1075/ijcl.1.2.05mey.
- Meyer, I., & Mackintosh, K. (2000). When terms move into our everyday lives: An overview of de-terminologization. *Terminology* 6, 111–138. doi: 10.1075/term.6.1.07mey.
- Miles, A., & Bechhofer, S. (2009). SKOS. Simple knowledge organization system namespace document. Retrieved from <http://www.w3.org/2009/08/skos-reference/skos.html>.
- Milroy, J., & Milroy, L. (1990). *Authority in Language: Investigating Standard English*. London: Routledge.
- Monson, S. C. (1973). Restrictive labels – Descriptive or prescriptive? In McDavid, R. I., & Duckert, A. R. (Eds.), *Lexicography in English* (pp. 208–212). New York: New York Academy of Sciences.
- Moon, R. (1989). Objective or Objectionable? Ideological Aspects of Dictionaries, *ELR Journal* 3, pp. 59–91.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- Morris, D. (1985). *A Tribo do Futebol*. Lisboa: Publicações Europa-América.
- Mugglestone, L. (2011). *Dictionaries. A very short introduction*. Oxford: Oxford University Press.
- Müller-Spitzer, C. (2008). The lexicographic portal of the IDS: Connecting heterogeneous lexicographic resources by a consistent concept of data modelling. In Bernal, E., & DeCesaris, J. (Eds.), *Proceedings of the Thirteenth EURALEX International Congress, Barcelona, Spain, July 15th–19th, 2008* (pp. 457–461). Barcelona: Universitat Pompeu Fabra and Institut Universitari de Lingüística Aplicada. Retrieved from <https://euralex.org/publications/the-lexicographic-portal-of-the-ids-connecting-heterogeneous-lexicographic-resources-by-a-consistent-concept-of-data-modelling/>.
- Müller-Spitzer, C. (2013). Textual structures in electronic dictionaries. In Gouws, Rufus H., et al. (Eds.), *Wörterbücher/Dictionaries/Dictionnaires: Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie Internationale de Lexicographie* (pp. 367–381). Berlin: De Gruyter Mouton. doi:10.1515/9783110238136.367.

- Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250. doi:10.1016/j.artint.2012.07.001.
- Neuendorf, K. E., Mehl Jr., J. P., & Jackson, J. A. (2011). *Glossary of Geology*. 5th ed. Alexandria, Virginia: American Geosciences Institute. Springer Science & Business Media.
- Nielsen, S. (2013). The future of dictionaries, dictionaries of the future. In Jackson, H. (Ed.), *The Bloomsbury Companion to Lexicography* (pp. 355–372). London: Bloomsbury Academic.
- Nielsen, S. (2018). Lexicography and interdisciplinarity. In Fuertes-Olivera, P. A. (Ed.), *The Routledge Handbook of Lexicography* (pp. 93–104). London: Routledge.
- Nielsen, S., & Tarp, S. (2009). *Lexicography in the 21st century*. In Honour of Henning Bergenholtz. Amsterdam: John Benjamins Publishing Company. doi:10.1075/tlrp.12.
- Nomdedeu Rull, A. (2008). Hacia una reestructuración de la marca de ‘deportes’ en lexicografía. In Azorín Fernández, D., et al. (Eds.), *El diccionario como puente entre las lenguas y culturas del mundo. Actas del II Congreso Internacional de Lexicografía Hispánica* (pp. 764–770). Alicante: Biblioteca Virtual Miguel de Cervantes. Retrieved from <https://dialnet.unirioja.es/servlet/articulo?codigo=5511595>.
- Nová, J. (2018). Terms embraced by the general public: How to cope with determinologization in the dictionary? In Čibej, J., Gorjanc, V., Kosem, I., Krek, S. (Eds.), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana, Slovenia, 17–21 July 2018 (pp. 387–398). Ljubljana: Ljubljana University Press. Retrieved from <https://euralex.org/publications/terms-embraced-by-the-general-public-how-to-cope-with-determinologization-in-the-dictionary/>.
- O’Reilly, T. (2005). What is Web 2.0: Design patterns and business models for the next generation of software. Retrieved from <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>.
- Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning: A study of the Influence of language upon thought and of the science of dymbolism*. New York: Harcourt, Brace & World.
- Pais, J., & Rocha, R. (2010). *Quadro de divisões estratigráficas*. Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa.
- Pavel, S., & Nolet, D. (2001). *Handbook of terminology / Précis de terminologie*. Ottawa: Terminology and Standardization, Translation Bureau.
- Paz Battaner, M. (1996). Terminología y diccionarios. In *Actes de la Jornada Panllatina de Terminologia* (pp. 93–117). Barcelona: Institut Universitari de Lingüística Aplicada.
- Peixoto, J. P. (1997). A ciência em Portugal e a Academia das Ciências de Lisboa. *Colóquio/Ciências*, 19, 71–84.

- Pereira, R. R., & Nadin, O. L. (2019). Dicionário enquanto gênero textual: Por uma proposta de categorização. *Acta Scientiarum Language and Culture*, 41(1), 1–8. doi:10.4025/actascilangcult.v41i1.43835.
- Pérez Pascual, J. I. (2012). El léxico de especialidad. In Luque Toro, L., Medina Monteiro, J., & Luque, R. (Eds.), *Léxico español actual III* (pp. 189–219). Venecia: Libreria Editrice Cafoscarina.
- Pilehvar, M. T., & Navigli, R. (2014). A Robust approach to aligning heterogeneous lexical resources. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, vol. 1, *Long Papers* (pp. 468–478). Baltimore, Maryland: Association for Computational Linguistics.
- Pinto de Jesus, A., Lemos de Sousa, M. J., Chaminé, H. I., Dias, R., Fonseca, P. E., & Gomes, A. (2010). O carbonífero em Portugal. In J. M. Coteló Neiva, A. Ribeiro, L. Mendes Victor, F. Noronha & Magalhães Ramalho, M. (Eds.), *Ciências geológicas: Ensino, investigação e sua história, vol. I, Geologia clássica* (pp. 341–355). Lisboa: Associação Portuguesa de Geólogos (APG), Sociedade Geológica de Portugal.
- Piotrowski, T. (2013). A Theory of lexicography – Is there one? In Jackson, H. (Ed.), *The Bloomsbury Companion to Lexicography* (pp. 303–320). London and New York: Bloomsbury Academic.
- Porto Dapena, A. (2002). *Manual de Técnica Lexicográfica*. Madrid: Arco/Libros.
- Pruvost, J. (2006). *Les dictionnaires français: Outils d'une langue et d'une culture*. Paris: Ophrys.
- Ptaszyński, M. O. (2010). Theoretical considerations for the improvement of usage labelling in dictionaries: A combined formal-functional approach. *International Journal of Lexicography*, 23(4), 411–442. doi:10.1093/ijl/ecq029.
- Quemada, B. (1968). *Les dictionnaires du français moderne, 1539–1863: Étude sur leur histoire, leurs types et leurs méthodes*. Paris: Didier.
- Quemada, B. (1987). Notes sur lexicographie et dictionnaire. *Cahiers de lexicologie*, 51(2), 229–242. Paris.
- Quemada, B. (Ed.). (1997). *Les préfaces du dictionnaire de l'Académie française (1694–1992): Textes, introductions et notes*. Paris: Champion.
- RAE. (1715). Fundación y estatutos de la Real Academia Española. Madrid: Imprenta Real. Retrieved from https://www.rae.es/sites/default/files/Estatutos_1715.pdf.
- Rey, A. (1970). Typologie génétique des dictionnaires. *Langages*, 19, 48–68.
- Rey, A. (1979). *La terminologie: noms et notions*. Paris: Presses Universitaires de France.
- Rey, A. (1983). Norme et dictionnaire (domaine du français). In Bédard, E., & Maurais, J. (Eds.), *La norme linguistique*. Québec: Le Robert.
- Rey, A. (1984/2001). Préface du *Grand Robert de la langue française*. In *Grand Robert de la langue française*. Retrieved from <https://grandrobert.lerobert.com/AideGR/Pages/Preface6.HTML>.

- Rey, A. (1985). La terminologie dans un dictionnaire général de la langue française: Le Grand Robert. *TermNet News*, 14, 5–7.
- Rey, A. (1989). Linguistic absolutism. In Hollier, D. (Ed.), *A new history of French literature* (pp. 373–379). Harvard: Harvard University Press.
- Rey, A. (1990). Les marques d'usage et leur mise en place dans les dictionnaires du XVIII^e siècle: le cas Furetière. In Glatigny, M. (Coord.), *Les marques d'usage dans les dictionnaires (XVII^e–XVIII^e siècles)* (pp. 17–29). Lille: Presses Universitaires de Lille.
- Rey, A. (1995). *Essays on Terminology*. Amsterdam: John Benjamins Publishing.
- Rey, A. (2003). La renaissance du dictionnaire de langue française au milieu du XX^e siècle: une révolution tranquille. In Cormier, M. C., Francoeur, A., & Boulanger J.-C. (Eds.), *Les dictionnaires Le Robert. Genèse et évolution* (pp. 88–99). Montréal: Presses de l'Université de Montréal.
- Rey, A. (2008). *De l'artisanat des dictionnaires à une science du mot. Images et modèles*. Paris: Armand Colin.
- Rey, A., & Delesalle, S. (1979). Problèmes et conflits lexicographiques. *Langue Française*, 43, 4–26.
- Rey-Debove, J. (1966). La définition lexicographique: recherches sur l'équation sémique. *Cahiers de lexicologie*, 8, 71–94. doi:10.15122/isbn.978-2-8124-4261-2.p.0077.
- Rey-Debove, J. (1971). *Étude linguistique et sémiotique des dictionnaires français contemporains*. Paris: The Hague.
- Richelet, P. (1680). *Dictionnaire françois, contenant les mots et les choses, plusieurs nouvelles remarques sur la langue françoise*. Genève: Chez Jean Herman Widerhold.
- Roberts, R. P. (2004). Terms in general dictionaries. In Bravo Gozalo, J. M. (Ed.), *A new spectrum of translation studies* (pp. 121–140). Valladolid: Universidad de Valladolid.
- Roche, C. (2012). Ontoterminology: How to unify terminology and ontology into a single paradigm. In Calzolari, N., Choukri, K., Declerck, T., et al. (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC- 2012)*. Istanbul, Turkey, May 23-25 (pp. 2626–2630). Istanbul: European Language Resources Association (ELRA).
- Roche, C. (2015). Ontological definition. In Kockaert H. J., & Steurs, F. (Eds.), *Handbook of terminology* (vol. 1, pp. 128–152). Amsterdam: John Benjamins Publishing Company.
- Roche, C., Calberg-Challot, M., Damas, L., & Rouard, P. (2009). Ontoterminology: A new paradigm for terminology. In *International Conference on Knowledge Engineering and Ontology Development, Oct 2009* (pp. 321–326). Funchal: [s.n.].
- Rodríguez Barcia, S. (2016). El Diccionario de la Lengua Española (2014): Análisis del nuevo discurso lexicográfico de la RAE. *Lexis*, 40(2), 331–374. Retrieved from http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S0254-92392016000200004&lng=es&tlng=es.

- Romary, L. (2013). Standardization of the formal representation of lexical information for NLP. In Gouws, R. H., Heid, U., Schweickard, W., & Wiegand, H. E. (Eds.), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary volume: Recent developments with special focus on electronic and computational lexicography* (pp. 1266–1274). Berlin, Boston: De Gruyter Mouton. doi:10.1515/9783110238136.1266.
- Romary, L., & Tasovac, T. (2018). TEI Lex-O: A target format for TEI-Encoded dictionaries and lexical resources. In *Proceedings of the 8th Conference of Japanese Association for Digital Humanities* (pp. 274–275). Retrieved from https://tei2018.dhii.asia/AbstractsBook_TEI_0907.pdf.
- Romary, L., & Wegstein, W. (2012). Consistent modelling of heterogeneous lexical structures. *Journal of the Text Encoding Initiative*, 3. doi:10.4000/jtei.540.
- Romary, L., Khemakhem, M., Khan, F., Bowers, J., Calzolari, N., George, M., Pet, M., & Bański, P. (2019). LMF reloaded. In Ahmet, M. G., Çiçekler, & N., Taşdemir, Y. (Eds.), *Proceedings of the 13th International Conference of the Asian Association for Lexicography* (pp. 533–539). Istanbul: Istanbul University Department of Linguistics. Retrieved from https://cdn.istanbul.edu.tr/FileHandler2.ashx?f=asialex_proceedings.pdf.
- Rondeau, G. (1984). *Introduction à la Terminologie*. Montréal: Gaëtan Morin.
- Rundell, M. (2010). What future for the learner's dictionary? In Kernerman I. J., & Bogaards, P. (Eds.), *English Learners' Dictionaries at the DSNA 2009* (pp. 169–175). Jerusalem: Kdictionaries.
- Rundell, M. (2012). The road to automated lexicography: An editor's viewpoint. In Granger, S., & Paquot, M. (Eds.), *Electronic Lexicography* (pp. 15–30). Oxford: Oxford University Press.
- Rundell, M. (2015). From print to digital: Implications for dictionary policy and lexicographic conventions. *Lexikos*, 25(1). doi:10.5788/25-1-1301.
- Rundell, M. (2019). Computer Corpora and Their Impact on Lexicography and Language Teaching. In Mullings, C., Stephanie, K., Deegan, M., & Ross, S. (Eds.), *New Technologies for the Humanities* (pp. 198–216). Berlin: K. G. Saur, 2019. doi:10.1515/9783110978278-012.
- Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam: John Benjamins Publishing Company.
- Sager, J. C. (2000). *Essays on definition*. Amsterdam: John Benjamins Publishing Company.
- Sager, J. C. (2004). The structure of the linguistic world of concepts and its representation in dictionaries: Eugen Wüster (1898–1977). *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 10(2), 281–306. doi:10.1075/term.10.2.08sag.
- Sagot, B. (2017). Extracting an etymological database from wiktionary. In *Electronic Lexicography in the 21st century (eLex 2017), Sep 2017, Leiden, Netherlands* (pp. 716–728). Retrieved from <https://hal.inria.fr/hal-01592061>.

- Sakwa, L. N. (2011). Problems of usage labelling in English lexicography. *Lexicos* 21, 305–315. doi:10.5788/21-1-47.
- Salgado, A., & Costa, R. (2019a). Marcas temáticas en los diccionarios académicos ibéricos: estudio comparativo. *RILEX. Revista sobre investigaciones léxicas* 2(2), 37–63. doi:10.17561/rilex.v2.n2.2.
- Salgado, A., & Costa, R. (2019b). A good TACTIC for lexicographical work: football terms encoded in TEI Lex-O. In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development: TOTh Conference 2019 – Terminology & Ontology: Theories and applications*, pp. 381–398. Chambéry, França: SciTePress – Science and Technology.
- Salgado, A., & Costa, R. (2020). O projeto Edição Digital dos Vocabulários da Academia das Ciências: o VOLP-1940. *Revista da Associação Portuguesa de Linguística*, 7, 275–294. doi:10.26334/2183-9077/rapln7ano2020a17.
- Salgado, A., Costa, R. & Tasovac, T. (2019). Improving the consistency of usage labelling in dictionaries with TEI Lex-O. *Lexicography: Journal of ASIALEX*, 6(2), 133–156. doi:10.1007/s40607-019-00061-x.
- Salgado, A., Costa, R., & Tasovac, T. (2021a). Comprender el mundo para mejorar un diccionario: las marcas temáticas en el Diccionario de la Lengua Española de la Real Academia Española. In *IX Congreso Internacional de Lexicografía Hispánica: Lexicografía del Español. Internacionalización e Intercomunicación*, May 25–27, Universidad de La Laguna, Spain.
- Salgado, A., Costa, R., & Tasovac, T. (2021b). Is there a place for orthographic dictionaries in the 21st Century? In *The International Conferences for Historical Lexicography and Lexicology (ICHLL)*, University of La Rioja, Logroño, Spain.
- Salgado, A., Costa, R., & Tasovac, T. (2021c). Mapping domain labels of dictionaries. In *Proceedings of XIX EURALEX International Congress: Lexicography for Inclusion*. Greece: Alexandroupolis, Greece.
- Salgado, A., Costa, R., Tasovac, T., & Simões, A. (2019). TEI Lex-O In Action: Improving the encoding of the Dictionary of the Academia das Ciências de Lisboa. In I. Kosem et al. (Eds.), *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference, 1-3 October 2019* (pp. 417–433). Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o.
- Salgado, A., Sina, A., Simões, A., Costa, R., & McCrae, J. (2020). Challenges of Word Sense Alignment: Portuguese Language Resources. In M. Ionov et al. (Eds.), *Proceedings of 7th Workshop on Linked Data in Linguistics (LDL-2020) Building Tools and Infrastructure*, 45–51. France: Marseille. ISBN 979-10-95546-46-7.
- Santos, C. (2010). *Terminologia e ontologias: metodologias para representação do conhecimento*. (Doctoral dissertation). Retrieved from <http://hdl.handle.net/10773/2876>.
- Santos, C., & Costa, R. (2015). Domain specificity: Semasiological and onomasiological knowledge representation. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of Terminology*, vol. 1 (pp. 153–179). Amsterdam: John Benjamins Publishing Company.

- Schreibman, S., Siemens, R., & Unsworth, J. (Eds.) (2004). *A Companion to Digital Humanities*. Oxford: Blackwell Retrieved from <http://www.digitalhumanities.org/companion/>. ISBN 9781405103213.
- Sebeok, T. (1962). Materials for a typology of dictionaries. In *Lingua*, 11, 363–374.
- Shcherba, L. (1940/1995). Towards a general theory of lexicography (Trans. D. M. T. Cr. Farina). *International Journal of Lexicography* 8(4): 305–349. (Translated from *Opyt obshchei teorii leksikografii. Izvestiia Akademii Nauk SSSR, Otdelenie literatury i iazyka*, 3, 1940, 89–117). doi:10.1093/ijl/8.4.314.
- Silva, R. (2014). *Gestão de terminologia pela qualidade. Faculdade de Ciências Sociais e Humanas*. (Doctoral dissertation). Retrieved from <http://hdl.handle.net/10362/13664>.
- Silvestre, J. P. (2008). *Bluteau e as origens da lexicografia moderna*. Lisboa: Imprensa Nacional-Casa da Moeda.
- Silvestre, J. P. (2016). Lexicografia. In A. M. Martins & E. Carrilho (Eds.), *Manual de linguística portuguesa* (pp. 200–223). Berlin: De Gruyter Mouton. doi:10.1515/9783110368840-010.
- Silvestre, J. P., Villalva, A., & Pacheco, P. (2014). The spectrum of red colour names in Portuguese. In *Proceedings of the 50th Anniversary Convention of the AISB*. Retrieved from <http://doc.gold.ac.uk/aisb50/AISB50-S20/aisb50-S20-silvestre-paper.pdf>.
- Simões, A. (2014). Informáticos, linguistas e linguagens. In Macedo, A. G., Sousa, C. M., & Moura, V. (Eds.), *XV Colóquio de Outono: As Humanidades e as Ciências. Disjunções e Confluências*. V. N. Famalicão: Edições Húmus. Retrieved from <http://repositorium.sdum.uminho.pt/handle/1822/42238>.
- Simões, A., Almeida, J. J., & Salgado, A. (2016). Building a dictionary using XML technology. In Mernik, M., Leal, J. P., Oliveira, H. G. (Eds.), *5th Symposium on Languages, Applications and Technologies (SLATE'16)* (14:1–14:8). Germany: Dagstuhl. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. doi:0.4230/OASlcs.SLATE.2016.0.
- Simões, A., Salgado, A., Costa, R., & Almeida, J. J. (2019). LeXmart: A smart tool for lexicographers. In Kosem, I., Zingano Kuhn, T., Correia, M. Ferreira, J. P., Janson, M., Pereira, I., Kallas, J., Jakubicek, M., Krek, S. & Tiberius, C. (Eds.), *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference* (pp. 453–466). Sintra, Portugal, Brno: Lexical Computing CZ, s.r.o. ISSN 2533-5626.
- Simões, A., Salgado, A., & Costa, R. (2021). LeXmart: A platform designed with lexicographical data in mind. In I. Kosem et al. (Eds.), *Electronic lexicography in the 21st century: Post-editing lexicography. Proceedings of the eLex 2021 conference* (pp. 529–541). Brno: Lexical Computing CZ. ISSN 2533-5626.
- Smit, M. (1996). Wiegand's metalexicography as a framework for a multilingual, multicultural, explanatory music education dictionary for South Africa. Unpublished D. Litt. Thesis. Stellenbosch: University of Stellenbosch.

- Souffi, S. (2009). Le dictionnaire de l'Académie française: between good use and culture. *Ela. Études de linguistique appliquée*, 2(2), 155–176. doi:10.3917/ela.154.0155.
- Sperberg-McQueen, C. M., Burnard, L., et al. (1994). *Guidelines for Electronic Text Encoding and Interchange*, vol. 1. Text Encoding Initiative Chicago and Oxford.
- Stührenberg, M. (2012). The TEI and current standards for structuring linguistic data. *Journal of the Text Encoding Initiative*, 3. doi:10.4000/jtei.523.
- Svensén, B. (1993). *Practical lexicography: Principles and methods of dictionary-making*. Oxford: Oxford University Press.
- Svensén, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary Making*. Cambridge: Cambridge University Press.
- Svensson, P. (2009). Humanities computing as digital humanities. *Digital Humanities Quarterly* 3(3). Retrieved from <http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html>.
- Swanepoel, P. (2010). Improving the functionality of dictionary definitions for lexical sets: The role of definitional templates, definitional consistency, definitional coherence and the incorporation of lexical conceptual models. *Lexikos*, 20, 425–449. doi:10.5788/20-0-151.
- Taborek, J. (2012). The language of sport: some remarks on the language of football. In Lankiwewicz, H., & Waiskiewicz-Firlej, E. (Eds.), *Informe teaching – premises of modern foreign language pedagogy* (pp. 229–255). Pila: Stanisława Staszica.
- Tarp, S. (2008). *Lexicography in the borderland between knowledge and non-knowledge: general lexicographical theory with particular focus on learner's lexicography*. Berlin, New York: Max Niemeyer. doi:10.1515/9783484970434.
- Tasovac, T. (2010). Reimagining the dictionary, or why lexicography needs digital humanities. *Digital Humanities 2010*. Abstract retrieved from <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-883.html>.
- Tasovac, T. (2020). The historical dictionary as an exploratory tool: A digital edition of vuk stefanović karadžić's lexicon serbicogermanico-latinum. (Doctoral dissertation). Trinity College, Dublin. Retrieved from <http://hdl.handle.net/2262/92750>.
- Tasovac, T., & Petrović, S. (2015). Multiple access paths for digital collections of lexicographic paper slips. In Kosem, I., Jakubiček, M., Kallas, J., & Krek, S. (Eds.), *Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital Age. Proceedings of the eLex 2015 Conference* (pp. 384–396). Ljubljana/Brighton: Institute for Applied Slovene Studies and Lexical Computing. Retrieved from https://elex.link/elex2015/proceedings/eLex_2015_25_Tasovac+Petrovic.pdf.
- Tasovac, T., Romary, L., Bański, P., Bowers, J., Does, J. de, Depuydt, K., Erjavec, T., Geyken, A., Herold, A., Hildenbrandt, V., Khemakhem, M., Petrović, S., Salgado, A., e Witt, A. (2018). TEI Lex-0: A baseline encoding for lexicographic data. Version 0.8.5. *DARIAH Working Group on Lexical Resources*. Retrieved from <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>.

- Tasovac, T., Salgado, A., & Costa, R. (2020). Encoding polylexical units with TEI Lex-0: A case study. *Slovenščina 2.0: Empirical, Applied and Interdisciplinary Research*, 8(2), 28–57. doi:10.4312/slo2.0.2020.2.28-57. e-ISSN 2335-2736.
- TEI Consortium, (Eds.). TEI P5: *Guidelines for Electronic Text Encoding and Interchange*. [Version 4.3.0]. [Last updated on 2021-08-31]. TEI Consortium. Retrieved from <http://www.tei-c.org/Guidelines/P5/>.
- Teixeira, C. (1944). *O Antrocolítico continental português. (Estratigrafia e tectónica)*. (Doctoral dissertation). Universidade do Porto.
- Tekorienė, D., & Maskaliūnienė, N. (2004). *Lexicography: British and American dictionaries*. Vilnius: Vilnius University Press.
- Temmerman, R. (2000). *Towards new ways of terminology description. The sociocognitive-approach*. Amsterdam: John Benjamins Publishing Company.
- Ten Hacken, P. (2018). Terms between standardization and the mental lexicon. *Roczniki Humanistyczne*, 66(11), 59–77. doi:10.18290/rh.2018.66.11-4.
- Terras, M., Nyhan, J., & Vahouette, E. (Eds.). (2013). *Defining Digital Humanities: A Reader*. London: Ashgate.
- The Digital Humanities Manifesto 2.0*. (2009). Retrieved from http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf.
- III Jubileu da Academia das Ciências de Lisboa*. (1931). Coimbra: [s.e.].
- Tiberius, C., Costa, R., Erjavec, T., Krek, S., McCrae, J., Roche, C., & Tasovac, T. (2020). Best practices for lexicography – intermediate report. In *ELEXIS – European Lexicographic Infrastructure*. Retrieved from https://elex.is/wp-content/uploads/2020/02/ELEXIS_D1_2_Best_practices_for_Lexicography_Intermediate_Report.pdf.
- Tournier, J. (1992). Problèmes de terminologie en lexicologie anglaise et générale. *Recherches en linguistique étrangère*, 16, 215–226.
- Trap-Jensen, L. (2018). Lexicography between NLP and Linguistics: Aspects of Theory and Practice. In Čibej, J., Gorjanc, V., Kosem, I., & Krek, S. (Eds.), *Proceedings of the 18th EURALEX International Congress: Lexicography in Global Contexts* (pp. 17–21). Ljubljana: Ljubljana University Press, Faculty of Arts. Retrieved from <https://euralex.org/wp-content/themes/euralex/proceedings/Euralex%202018/118-4-2949-1-10-20180820.pdf>.
- Van Sterkenburg, P. (Ed.). (2003). *A practical guide to lexicography*. Amsterdam: John Benjamins.
- Verdelho, T. (1994). Tecnolectos. In Holtus, G., Metzeltin, M., & Schmitt, C. (Eds.), *Lexikon der Romanistischen Linguistik*, vol. 6(2) (pp. 339–355). Max Niemeyer: Tübingen.
- Verdelho, T. (1998). Terminologias na língua portuguesa. Perspectiva diacrónica. In J. Brumme, (Ed.), *La història dels llenguatges iberoromànics d'especialitat (segles XVII-XIX): solucions per al present* (pp. 98–131). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

- Verdelho, T. (2002). O dicionário de Morais Silva e o início da lexicografia moderna. In Head, B. F., Teixeira, J., Lemos, A. S., Barros, A. L., & Pereira, A. (Eds.), *História da Língua e História da Gramática – Actas do encontro* (pp. 473–490). Braga: ILCH, Universidade do Minho.
- Verdelho, T. (2007). Dicionários portugueses: Breve história. Verdelho, T., & Silvestre, J. P. (Orgs.), *Dicionarística portuguesa, inventariação e estudo do património lexicográfico* (pp. 11–60). Aveiro, Universidade de Aveiro.
- Verkuyl, H. J., Janssen, M., & Jansen, F. (2003). The codification of usage by labels. In Sterkenburg, P. (Ed.), *A practical guide to lexicography* (pp. 297–311). Amsterdam: John Benjamins. doi:10.1075/tlrp.6.33ver.
- Villalva, A., & Williams, G. (2019). *The landscape of lexicography*. Lisboa and Aveiro: Centro de Linguística da Universidade de Lisboa and Universidade de Aveiro.
- Villers, M.-É. (2006). *Profession lexicographe*. New edition (online). Montréal: Presses de l'Université de Montreal. doi: <https://doi.org/10.4000/books.pum.135>.
- Vogel, C. (1979). *A history of Indian literature*, vol. 4, *Indian lexicography*. Wiesbaden: Otto Harrassowitz.
- Walczak, B. (1991). *La terminologie dans les dictionnaires généraux*. *Neoterm*, 13(16), 126–130.
- Wang, S. (2016). Lexicultura na língua chinesa e na lexicografia bilingue de chinês-português. (Doctoral dissertation). Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa. Retrieved from <https://run.unl.pt/handle/10362/17164>.
- Wiegand, H. E. (1984). On the structure and contents of a general theory of lexicography. Hartmann, R. R. K. (Ed.). In *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983* (pp. 13–30). Tübingen: Max Niemeyer Verlag.
- Wiegand, H. E. (1985). Eine neue Auffassung der sog. Lexikographischen Definition. In Hyldgaard-Jensen, K., & Zettersten, A. (Eds.), *Symposium on Lexicography II. Proceedings of the Second International Symposium on Lexicography, May 16–17, 1984 at the University of Copenhagen, Tübingen, Niemeyer* (pp. 15–100). Tübingen: Max Niemeyer Verlag. doi:10.1515/9783111341132-002.
- Wiegand, H. E. (1989a). Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In Hausmann, F. J. et al. (Eds.), *Wörterbücher, dictionaries, dictionnaires. Ein internationales Handbuch zur Lexikographie*, vol. 1 (pp. 409–462). Berlin and New York: De Gruyter.
- Wiegand, H. E. (1989b). Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. In Hausmann, F. J. et al. (Eds.), *Wörterbücher, dictionaries, dictionnaires. Ein internationales Handbuch zur Lexikographie*, vol. 1 (pp. 462–501). Berlin and New York: De Gruyter.
- Wiegand, H. E. (1996/2011). Über die Mediostrukturen bei gedruckten Wörterbüchern. In Kammerer, M., & Wolski, W. (Eds.), *Kleine Schriften. Eine Auswahl aus den*

- Jahren 1970-1999 in zwei Bänden. Bd 1: 1970-1988. Bd 2: 1988-1999.* (pp. 1163-1192). Berlin, Boston: De Gruyter. doi:10.1515/9783110808117.1163.
- Wiegand, H. E. (1998). *Wörterbuchforschung*. Berlin: De Gruyter.
- Wiegand, H. E., Gouws, R. H., Kammerer, M., Mann, M., & Wolski, W. (2020). *Dictionary of Lexicography and Dictionary Research*, vol. 3, I-U. Berlin and Boston: De Gruyter.
- Williams, G. (2019). The problem of interlanguage diachronic and synchronic markup. In Villalva A., & Williams, G. (Eds.), *The landscape of lexicography*. Lisboa and Aveiro: Centro de Linguística da Universidade de Lisboa and Universidade de Aveiro.
- Wooldridge, R. (1977). *Les débuts de la lexicographie française: Estienne, Nicot et le Thresor de la langue françoise (1606)*. Toronto: University of Toronto Press.
- Wooldridge, R. (2004). Lexicography. In Schreibman, S., Siemens, R., & Unsworth, J. (Eds.), *A Companion to Digital Humanities* (pp. 69–78). Oxford: Blackwell. Retrieved from <http://www.digitalhumanities.org/companion/>.
- Wüster, E. (1968). *The machine tool: An interlingual dictionary of basic concepts; comprising an alphabetical dictionary and a classified vocabulary with definitions and illustrations*. London: Technical Press.
- Wüster, E. (1979/1998). *Introducción a la teoría general de la terminología y a la lexicografía terminológica*. Barcelona: Institut Universitari De Lingüística Aplicada, Universitat Pompeu Fabra. (Einführung in die Allgemeine Terminologielehre und terminologische Lexikographie. Bonn: Romanistischer Verlag, 1979).
- Xue, S. (1982). Chinese lexicography past and present. *Dictionaries: Journal of the Dictionary Society of North America* 4, 151–169. doi:10.1353/dic.1982.0009.
- Yong, H., & Peng, J. (2007). *Bilingual lexicography from a communicative perspective*. Amsterdam: John Benjamins Publishing Company. doi:10.1075/tlrp.9.
- Yong, H., & Peng, J. (2008). Chinese lexicography: A history from 1046 BC to AD 1911. *Cahiers de linguistique – Asie orientale*, 39(1), pp. 81–94.
- Zgusta, L. (1971). *Manual of lexicography*. Prague and The Hague: Academia and Mouton.

LIST OF FIGURES

- Figure 1:** *The Digital Humanities Stack* (Berry & Fagerjord, 2017)
- Figure 2:** Definition 1 – Entry ‘lexicographie’ [lexicography] in the DAF (AF)
- Figure 3:** Definition 2 – Entry ‘lexicografía’ [lexicography] in the DLE (RAE)
- Figure 4:** Entry ‘lexicografia’ [lexicography] in the DLPC (ACL)
- Figure 5:** *The Theoretical and Practical Components of Lexicography*
- Figure 6:** Definition 1 – Entry ‘terminologie’ [terminology] in the DAF (AF)
- Figure 7:** Definition 2 – Entry ‘terminología’ [terminology] in the DLE (RAE)
- Figure 8:** Definition 3 – Entry ‘terminologia’ [terminology] in the DLPC (ACL)
- Figure 9:** Definition 4 – Entry ‘terminology’ in the OED, Oxford University Press
- Figure 10:** *Lexicography vs Terminology*
- Figure 11:** *Dictionary seen as a diamond with multiple facets*
- Figure 12:** *Categories of a Dictionary’s Taxonomic Classification*
- Figure 13:** *Classification of the Academy Dictionaries under study*
- Figure 14:** *Model of a Dictionary Structure*
- Figure 15:** *Emblem of the Académie Française (AF)*
- Figure 16:** *Charter of the Académie Française (1635)*
- Figure 17:** *Title page of the Dictionnaire de l’Académie Française, engraved by Pierre-Jean Mariette in 1694*
- Figure 18:** *Le Dictionnaire de l’Académie Française, Dédié au Roy, 1st edition (DAF, 1694, p. 289)*
- Figure 19:** *Nouveau Dictionnaire de l’Académie Française Dédié au Roy, 2nd edition*
- Figure 20:** *Front page of Dictionnaire de l’Académie Française (2021), AF*
- Figure 21:** *Paduan academic’s emblem and the emblem of the RAE*
- Figure 22:** *Charter of the Real Academia Española (RAE, 1715), 1st edition*
- Figure 23:** *Title page of the Diccionario de la Lengua Castellana, RAE (1780)*
- Figure 24:** *Front page of the Diccionario de Lengua Española en línea (2021), RAE*
- Figure 25:** *Emblem of the Academia das Ciências de Lisboa (ACL)*
- Figure 26:** *Diccionario da Lingoa Portuguesa (1793), ACL*
- Figure 27:** *Dicionário da Língua Portuguesa Contemporânea (2001), ACL*
- Figure 28:** Entry ‘femelle’ [female], *Dictionnaire François* (1680), AF
- Figure 29:** Entry ‘demi-ton’ [semitone], *Dictionnaire François* (1680), AF
- Figure 30:** Entry ‘eluviação’ [eluvium] in the DLPC (ACL)
- Figure 31:** Entry ‘musivario’ [mosaic, mosaicist, mosaicking] in the DLE (RAE)

Figure 32: Entry ‘*abcesso*’ [abscess] in the DLPC (ACL)

Figure 33: Entry ‘*escanteio*’ [corner] in the DLPC (ACL)

Figure 34: Entry ‘*cratera*’ [crater] in the DLPC (ACL)

Figure 35: Entry ‘*pança*’ [paunch, belly] in the DLPC (ACL)

Figure 36: Entry ‘*haut-de-chausses*’ [breeches] in the DAF (AF)

Figure 37: Entry ‘*banana*’ [banana] in the DAF (AF)

Figure 38: Entries ‘*iceberg*’ and ‘*icebergue*’ [iceberg] in the DLPC (ACL)

Figure 39: Entry ‘*friolero*’ [sensitive to the cold] in the DLE (RAE)

Figure 40: Entry ‘*printemps*’ [spring] in the DAF (AF)

Figure 41: List of abbreviations of the *Diccionario de Autoridades* (1770), RAE

Figure 42: The Relationship of Concept and Term mirroring the double dimension of terminology (adapted from Costa, 2021)

Figure 43: Formal representation of lexical entries in the DPLC (Salgado et al., 2019)

Figure 44: The Meaning Triangle (adapted from Ogden and Richards, 1923)

Figure 45: The entry ‘rock’ in different English dictionaries

Figure 46: Fragment of the DLPC list

Figure 47: Fragment of the DLE list

Figure 48: Fragment of the DAF list

Figure 49: Domain labels in the DLPC (184)

Figure 50: Domain labels in the DLE (74)

Figure 51: Domain labels in the DAF (132)

Figure 52: Areas of knowledge with the highest representation in the DLCP and the DLE

Figure 53: Less frequent domains in the DLPC and the DLE

Figure 54: DLPC vs DLE – Correspondence between domain labels in both dictionaries (65)

Figure 55: DLPC vs DAF – Correspondence between domain labels in both dictionaries (136)

Figure 56: DLE vs DAF – Consensus between domain labels in both dictionaries (53)

Figure 57: Entry ‘*geologia*’ [geology] in the DLPC (ACL)

Figure 58: Entry ‘*geología*’ [geology] in the DLE (RAE)

Figure 59: Entry ‘*géologie*’ [geology] in the DAF (AF)

Figure 60: Entry ‘*cristalografia*’ [crystallography] in the DLPC (ACL)

Figure 61: Entry ‘*cristalografía*’ [crystallography] in the DLE (RAE)

Figure 62: Entry ‘*cristalographie*’ [crystallography] in the DAF (AF)

Figure 63: Entry ‘*mineralogia*’ [mineralogy] in the DLPC (ACL)

Figure 64: Entry ‘*mineralogía*’ [mineralogy] in the DLE (RAE)

Figure 65: Entry ‘*minéralogie*’ [mineralogy] in the DAF (AF)

Figure 66: Entry ‘paleontología’ [paleontology] in the DLPC (ACL)

Figure 67: Entry ‘paleontología’ [paleontology] in the DLE (RAE)

Figure 68: Entry ‘paléontologie’ [paleontology] in the DAF (AF)

Figure 69: Entry ‘futebol’ [football] in the DLPC (ACL)

Figure 70: Entries ‘fútbol/futbol’ [football] in the DLE (RAE)

Figure 71: Entry ‘football’ [football] in the DAF (AF)

Figure 72: Entries ‘fanerozóico’ and ‘fanerozoico’ [Phanerozoic] in the DLPC (ACL) and in the DLE (RAE)

Figure 73: Fragment of the entry ‘era’ [era] in the DLPC (ACL)

Figure 74: Fragment of the entry ‘era’ [era] in the DLE (RAE)

Figure 75: Fragment of the entry ‘ère’ [era] in the DAF (AF)

Figure 76: Entries ‘paleozóico’ [palaeozoic], ‘mesozóico’ [mesozoic], ‘cenozóico’ [cenozoic] in the DLPC (ACL)

Figure 77: Entries ‘paleozoico’ [Palaeozoic], ‘mesozoico’ [Mesozoic], ‘cenozoico’ [Cenozoic] in the DLE (RAE)

Figure 78: Entries ‘paléozoïque’ [palaeozoic], ‘mesozoico’ [mésozoïque], ‘cénozoïque’ [Cenozoic] in the DAF (AF)

Figure 79: Entry ‘carbonífero’ [Carboniferous] in the DLPC (ACL)

Figure 80: Entry ‘carbónico’ [Carboniferous] in the DLPC (ACL)

Figure 81: Entry ‘carbonífero’ [Carboniferous] in the DLE (RAE)

Figure 82: Entry ‘carbonifère’ [Carboniferous] in the DAF (AF)

Figure 83: Entry ‘águia’ [eagle; supporter of Sport Lisboa e Benfica sports club] in the DLPC (ACL)

Figure 84: Entry ‘chapéu’ [chip] in the DLPC (ACL)

Figure 85: Entry ‘grande penalidade’ [penalty kick] in the DLPC (ACL)

Figure 86: Entries ‘extremo’ [winger] and ‘lateral’ [back] in the DLPC (ACL)

Figure 87: Entries ‘extremo’ [winger] and ‘lateral’ [back] in the DLE (RAE)

Figure 88: Entries ‘ailier’ [winger] and ‘arrière’ [back] in the DAF (AF)

Figure 89: Entry ‘gilista’ [supporter of Gil Vicente Futebol Clube] in the DLPC (ACL)

Figure 90: Entry ‘leão’ [lion; supporter of Sporting Club de Portugal] in the DLPC (ACL)

Figure 91: Entry ‘portista’ [supporter of Futebol Clube do Porto] in the DLPC (ACL)

Figure 92: Entry ‘colchonero’ [supporter of Atlético de Madrid] in the DLE (RAE)

Figure 93: Entry ‘culé’ [supporter of Fútbol Club Barcelona] in the DLE (RAE)

Figure 94: Entry ‘merengue’ [Real Madrid Club de Fútbol] in the DLE (RAE)

Figure 95: Applying terminological methods when treating terms in general language dictionaries

Figure 96: International Chronostratigraphic Chart (Cohen et al., 2021)

Figure 97: Entries 'futebol/football/fútbol' (DLPC, DLE, DAF)

Figure 98: Football players occupy different positions on the field (Salgado & Costa, 2020)

Figure 99: Positions of football players on the field

Figure 100: Domains hierarchy

Figure 101: Dewey Decimal Classification System

Figure 102: Universal Decimal Classification System

Figure 103: UNESCO Thesaurus Classification System

Figure 104: EuroVoc Classification System

Figure 105: WordNet Domains Hierarchy

Figure 106: Domain labels within the EARTH SCIENCES superdomain showing geology as domain and identifying its subdomains

Figure 107: Olympic Sports

Figure 108: Domain labels within the SPORTS superdomain showing TEAM SPORTS, INDIVIDUAL SPORTS as domains and FOOTBALL as a subdomain

Figure 109: Validation grid template (DLP)

Figure 110: Representation of a generic relation using the concept of <GeochronologicUnit>

Figure 111: Representation of the relation the conceptual markers is_a and has_function established from <GeochronologicUnit>

Figure 112: Representation of a partitive relation using the concepts of <GeochronologicUnit> and <GeologicalTimeScale>

Figure 113: Representation of a generic relation using the concept of <GeologicalEra>

Figure 114: Representation of a mixed concept system with the concepts of <Back> and <Winger>

Figure 115: Representation of the relation of the conceptual markers is_a, part_of, and has_position established from <Winger>

Figure 116: Representation of the relation of the conceptual markers is_a, part_of and has_position established from <Winger>

Figure 117: Representation of the relation between the conceptual markers is_a, consists_of and formed_during established from <ChronostratigraphicUnit>

Figure 118: Representation of an associative relationship with the concepts of <ChronostratigraphicUnit> and <GeochronologicUnit> with generic and partitive relations – a mixed concept system

Figure 119: Conceptualising <Phanerozoic>

Figure 120: Entry 'era' [era] updated in the DLP (2021)

Figure 121: Entry ‘defesa’ [defence] updated in DLP (2021)

Figure 122: Different Views on Lexicographic Resources (Khan & Salgado, 2021)

Figure 123: XML Essential Changes – DLCP Original Encoding and DLP Conversion into TEI Lex-0

Figure 124: Entry ‘cristalografia’ [crystallography] in the DLPC (ACL)

Figure 125: Entry ‘cristalografia’ [crystallography] in the DLP (ACL)

Figure 126: Entry ‘paleozóico’ [palaeozoic] in the DLPC (ACL)

Figure 127: Entry ‘paleozoico’ [palaeozoic] in the DLP (ACL)

Figure 128: Entry ‘cratera’ [crater] in the DLPC (ACL)

Figure 129: Entry ‘estrelícia’ [strelitzia] in the DLPC (ACL)

Figure 130: Entry ‘defesa’ [defence] in the DLPC (ACL)

Figure 131: Entry ‘guarda-redes’ [goalkeeper] in the DLP (ACL)

Figure 132: Entry ‘trivela’ in the DLP (ACL)

LIST OF TABLES

Table 1: *Classifications of diasystematic information proposed by different researchers (retrieved from Salgado, Costa & Tasovac, 2019)*

Table 2: *Comparative typography of domain labels*

Table 3: *Domain labels in the three academy dictionaries*

Table 4: *Different abbreviations of the same domain labels in the DLPC, DLE and DAF*

Table 5: *Similar abbreviation labels and domains in the DLPC, DLE and DAF*

Table 6: *Domains (metalabels) with an exact correspondence (61)*

Table 7: *A portion of domain labels with a reLated correspondence*

Table 8: *A portion of domain labels without any correspondence, none*

Table 9: *Terms referring to positions occupied by football players on the field*

Table 10: *Conventional hierarchy of the chronostratigraphic/geochronologic units*

Table 11: *Comparison of academy dictionaries domain labels and classification systems (Salgado, Costa, & Tasovac, 2021)*

Table 12: *Comparison of definitions ‘éon’, ‘era’, ‘período’, ‘época’, ‘idade’ in DLPC (2001) and DLP (2021)*

Table 13: *Comparison of definitions ‘eonothem’, ‘erathem’, ‘system’, ‘series’, ‘stage’ in the DLPC (2001) and the DLP (2021)*

Table 14: *Comparison of ‘cenozoico’, ‘mesozoico’, ‘paleozoico’ definitions in the DLPC (2001) and the DLP (2021)*

Table 15: *Comparison of definitions of the concepts designated by the terms ‘carbónico’ and ‘carbonífero’ in the DLPC (2001) and the DLP (2021)*

Table 16: *Comparison of the definitions of the terms ‘ataque’, ‘defesa’, ‘meio-campo’ in the DLPC (2001) and the DLP (2021)*

Table 17: *Comparison of the definitions of the terms ‘guarda-redes’, ‘avançado’, ‘extremo’, ‘lateral’, ‘libero’, ‘defesa’, ‘médio’, ‘ponta de lança’ in the DLPC (2001) and the DLP (2021)*

Table 18: *Lexicographic/Terminological form of a term in a general language dictionary*

Table 19: *Domains and subdomains under study and their metalabel*

Table 20: *Domain label occurring at different levels of the entry’s hierarchy*

ANNEXES

INDICAÇÕES PARA CONSULTA DO DICIONÁRIO

Os artigos do Dicionário apresentam uma estrutura hierarquizada, em que a informação fornecida para cada palavra está devidamente identificada, com campos específicos bem assinalados graficamente, por forma a possibilitar uma consulta rápida e eficaz dos dados.

Faz-se seguidamente, na coluna da esquerda, uma descrição da estrutura adoptada, que poderá servir de guia para o consultante, segundo a ordem linear e sequencial de apresentação das informações nos artigos do dicionário, e indicam-se, na coluna da direita, exemplos ilustrativos nos quais aparecem destacadas em sombreado as informações que se pretende ilustrar:

1. A cabeça do artigo identifica a palavra definida.

Esta surge grafada em letras minúsculas, em negro e está encostada com um espaço à esquerda, para maior destaque relativamente ao corpo do verbete, reservando-se a maiúscula inicial apenas para as palavras que ocorrem como nomes próprios. No caso de palavras que apresentem variação em género, masculino e feminino, a cabeça apresenta por extenso a forma masculina seguida de vírgula e da flexão feminina. Em palavras com variantes ortográficas, quer da flexão, quer do corpo da palavra, a informação será dada em linha com indicação das várias terminações ou das grafias variantes de uma mesma palavra, separadas por vírgula. Palavras com variantes ortográficas estão registadas em linha após a variante mais usual e apresentam a definição na entrada em linha. É dada entrada separada às restantes variantes sempre que se encontrem distantes na ordenação alfabética das diversas formas, com remissão para aquela que constitui a cabeça inicial do artigo.

2. No caso de existirem duas ou mais palavras homógrafas, etimológica, fonética ou sintacticamente distintas, recorre-se ao emprego de um índice numérico sobreposto à direita da cabeça por forma a distinguir as respectivas entradas.

3. Segue-se a transcrição fonética da palavra, de acordo com a pronúncia aproximada da norma de Lisboa e Centro, utilizando os símbolos do Alfabeto Fonético Internacional, entre parênteses

bonacheirão², o**ba** [bunɐ/ɐjɾɐw̃, -ónɐ]. *s.* (De *bom* + *-acho-* + *-eiro-* + suf. *-ão*). Pessoa de bondade e simplicidade excessivas; pessoa ingénua, paciente, pachorrenta. *Aquele bonacheirão não se rala quando os miúdos troçam dele.*

Abrilada [ɐbriláɐdɐ]. *s. f.* (De *Abril* + suf. *-ada*). **1.** *Hist.* Revolta ocorrida em 30 de Abril de 1824, na qual D. Miguel tenta depor seu pai, D. João VI. **2.** *Bras. Hist.* Revolução restauradora de Abril de 1832 em Pernambuco. **3.** (com minúsc.). *Pop.* Diz-se do período revolucionário português de 1974.

bailador², a, e**ira** [bajlɐdór, -ɐ, -ɛjɾɐ]. *s.* (De *bailar* + suf. *-dor*). **1.** Pessoa que baila. **2.** Pessoa que executa determinadas danças de carácter popular e tradicional. *Bailador de fandango.*

abetoira, abetoura [ɐbitóɾɐ], [ɐbitóɾɐ]. *s. f.* (Do lat. científico *botaurus*). *Zool.* Ave da família dos ardeídeos (*Botaurus stellaris*, Lin.), forte e de dorso pardo, semelhante à *garça*, que vive em lugares pantanosos e cujo macho tem voz estridente. ≈ **BETOIRA, BETOURA.**

aboiar¹ [ɐbojár]. *v.* (De *a-²* + *boi* + suf. *-ar*). **1.** *Region.* Levar a vaca ao boi tendo em vista a reprodução da espécie. ≈ **LEVAR À COBRIÇÃO.** **2.** *Region.* Cantar ou falar aos bois; guiar uma boiada, cantando. **3.** Cantar monotonicamente para serenar alguém. «*Casimiro Lopes levava o filho dela para o alpendre e embalava-o, aboindo, cantando*» (G. RAMOS, *S. Bernardo*, p. 189).

aboiar² [ɐbojár]. *v.* (De *a-²* + *bóia* + suf. *-ar*). **1.** *Náut.* Vir à tona de água. ≈ **BOIAR, FLUTUAR.** **2.** Assinalar, com bóia ou qualquer outro corpo flutuante, a posição de um objecto que se lançou à água, para, mais tarde, o recuperar.

Ca¹ *símb. Quím.* Símbolo do cálcio.

CA² [seá]. *s. f. Electr.* Sigla de Corrente Alternada.

expender [ɛ/pɛdɐɾ ~ ɨj/pɛdɐɾ]. *v.* (Do lat. *expēdĕre*). **1.** Expor pormenorizadamente; explicar em detalhe. **2.** Dar utilidade ou gastar. ≈ **DESPENDER. Expendere energia.**

rectos. Veja-se a lista dos símbolos apresentada na página ?? . Havendo variantes também usuais, estas serão apresentadas separadas pelo símbolo ~.

4. A **informação gramatical**, com indicação da categoria gramatical ou lexical, do género e do número, é apresentada em *itálico* no campo seguinte. Os símbolos, as abreviaturas, os elementos de formação, as contracções... também surgem identificados nesse campo.

5. A **etimologia** é indicada entre parênteses curvos depois da classificação gramatical, com informações sobre a origem e os elementos que entram na formação da palavra, assim como sobre o significado do étimo, quando ele não decorre imediatamente da definição que é dada a seguir.

6. A **definição**, em redondo, oferece uma descrição semântica das várias acepções da unidade lexical, devidamente numeradas com número em negro e ordenadas segundo uma ordem cronológica e/ou lógica. As combinatórias, as locuções e as expressões também são seguidas de definição em redondo; no caso de apresentarem mais do que uma acepção, recorre-se à sua numeração em redondo.

7. As **restrições de uso** sobre o registo de língua (e.g. *Fam.*, *Pop.*, *Gross.*, *Poét.*, etc.), o género ou o número, o uso geográfico (e.g. *Region.*, *Bras.*, *Angol.*, etc.), o campo científico (e.g. *Zool.*, *Bot.*, *Ling.*, etc.)... são apresentadas com as respectivas abreviaturas em *itálico* antes da primeira acepção, sendo válidas para todas as acepções, ou depois do número da acepção a que respeitam, sendo privativas da mesma.

apodídeos [epudídju]. *s. m. pl.* (Do gr. ἄπους, ὄδος 'sem pés' + suf. *-ídeo*). *Zool.* 1. Família (*Apodidae*) de aves com asas alongadas e pontudas, a que pertencem os andorinhões. 2. *s. m.* Ave dessa família.

Bi *simb. Quím.* Símbolo do *bismuto*.

bi- *elem. de form.* (Do lat. *bis* 'duas vezes'). Exprime as noções de *dois*, *duas vezes*. *Bianual*, *bimotor*.

A. D., a. D. *abrev.* Abrev. de *Anno Domini* ou *anno Domini*.

bombicídeo, a [bôbisídju, -v]. *adj.* (Do lat. *bombycis* < gr. βόμβυξ 'bicho-da-seda' + suf. *-ídeo*). *Zool.* Que pertence ou é relativo aos bombicídeos.

bombo [bôbu]. *s. m.* (Do lat. *bombus* < gr. βόμβος 'ruído'). Cf. *bumbo*. 1. *Mús.* Tambor de grandes dimensões, tocado em posição vertical, com maçanetas, sobre as duas faces de pele, usado em bandas, fanfarras e que entre os instrumentos de percussão representa o baixo. = *ZABUMBA*. *Abrindo o cortejo os zés-pereiras atroavam o ar com a música de bombos. bombo da festa*, pessoa sobre quem recai a agressividade de um grupo; aquele que é objecto de chacota. «A estudantada de Coimbra andou a gozar-te, palerma, foste um bombo de festa.» (A. TOJAL, *Putos*, p. 25). 2. Pessoa que toca esse instrumento. *O bombo da orquestra era um músico muito jovem. andar de bombo*, *Gross.*, estar grávida.

amor [umór]. *s. m.* (Do lat. *amor*, *-ôris*). 1. Predisposição da afectividade e da vontade, orientada para o objecto [...] **amor à camisola**. *Fam.* 1. Dedicção do desportista ou do adepto ao seu clube. «E por isso é que Zeca Sucateiro perdía a tramontana. Não que fosse mau homem, era amor à camisola, amor ao Beira-Rio. E também ao Benfica, deve-se acrescentar» (M. ALEGRE, *Alma*, pp. 106-107). 2. Dedicção desinteressada. [...]

abrigo [ebrigu]. *s. m.* (Do lat. *apricum* ou deriv. regres. de *abrigar*). 1. Protecção natural ou artificial contra o perigo [...] **ao abrigo de**. *loc. prep.* 1. Defendido de, fora do alcance de; protegido contra. *Ao abrigo dos ventos*. 2. Autorizado por; em conformidade com. *Ao abrigo do decreto n.º 3750*. [...]

betuláceas [bitulásje]. *s. f. pl.* (De *bétula* + suf. *-ácea*). *Bot.* 1. Família de plantas dicotiledóneas, que inclui árvores e arbustos das regiões frias e temperadas do Norte, de folhas simples e caducas, flores unissexuais, dispostas em amentilhos, frutos secos indeiscentes, a que pertence a *bétula*. 2. *s. f.* Planta dessa família.

avaria [everie]. *s. f.* (Do ár. *awarīya* 'produto deteriorado', pelo it. *avaria*). 1. *Náut.* Dano causado a um navio ou às mercadorias que ele transporta. 2. *Dir. Marít.* Despesa extra para segurança do navio e da sua carga, durante a viagem. 3. Dano, prejuízo, estrago. 4. Mau funcionamento ou falha num aparelho, num mecanismo, num veículo... = *DANO*, *DESARRANJO*, *FALHA*. *O motor do barco teve uma avaria. Avaria no carro. Avarias de telefone. Serviço de +s*

8. Após a definição da acepção é possível encontrar indicação de sinónimos e antónimos em versaletes, assinalados, respectivamente, pelos símbolos ≈ e ≠. Tratando-se de sinónimos e antónimos com pequenas variações, estilísticas, regionais ou outras, é dada a informação sobre a restrição de uso entre parênteses e em itálico.

9. Os exemplos, de uso opcional na estrutura de um artigo, são apresentados em itálico e oferecem informação sobre os contextos em que a unidade lexical pode ocorrer, e que mostram o funcionamento desta.

10. As abonações retiradas de textos literários, jornalísticos e outros preenchem funções idênticas às do exemplo. Aparecem citadas em itálico, entre aspas, com indicação da referência bibliográfica entre parênteses curvos (o autor em maiúsculas e com carácter redondo, a obra em itálico e a página em redondo).

11. É ainda possível encontrar em muitas acepções uma lista de co-ocorrentes privilegiados, que elucidam sobre os contextos em que a unidade frequentemente ocorre. A cabeça aparece representada nesse campo pelo sinal +, sendo toda a informação destacada em itálico.

12. As combinatórias fixas em que a unidade pode ocorrer e que formam locuções nominais ou adjetivais surgem em negro no fim de cada acepção com a definição dos respectivos significados em redondo devidamente numerados, se forem mais do que um. Poderão, tal como as acepções da entrada, ser ilustradas com exemplos e abonações. As combinatórias aparecem definidas na sua primeira palavra plena (geralmente, substantivo ou adjetivo). Estão também assinaladas em cada uma das palavras plenas que a constroem, com remissão, através do símbolo + sobreposto à direita da primeira palavra plena da combinatória, para a palavra onde se encontra a definição.

reparar a +. 5. *Fam.* Acto precipitado, irreflectido e de que pode ou não resultar prejuízo. ≈ HABILIDADE, PROEZA. *Disse-lhe que se deixasse de avarias, porque podia cair.* 6. *Fam.* Acto extravagante, ousado ou audaz. ≈ ESTRAGO, FAÇANHA, PROEZA. *Hoje fizemos uma avaria, jantámos no restaurante mais caro e mais chique da cidade.* 7. *Region. (Aleit.).* Milho de baixa qualidade ou avariado, utilizado na engorda dos porcos.

avarento², a [ɐvərətu, -ɐ]. s. (Do lat. hisp. **avarentus*, de *avārus* 'ávido'). 1. Pessoa que gosta de dinheiro ou de outros bens materiais acima de tudo e procura conservá-los ou aumentá-los cada vez mais, privando-se muitas vezes de comodidades ou prazeres; pessoa que tem avareza. ≈ AGARRADO, AVARO, FORRETA (*Fam.*), SOMÍTICO, SOVINA, ≠ DISSIPADOR, ESBANJADOR. *O povo costuma dizer: «O avarento rico não tem parente nem amigo» e «O dinheiro do avarento duas vezes vai à feira».* 2. Pessoa ávida de alguma coisa.

berma [bérmɐ]. s. f. (Do fr. *berme*). 1. *Fort.* Espaço entre a parte inferior da muralha e o fosso, destinado a deter as terras e pedras desprendidas pelos projecteis inimigos e a impedir que elas obstruíssem o fosso. 2. Caminho entre um molhe e a borda de um canal ou fosso. 3. Parte estreita e contígua à estrada, até à valeta. ≈ BORDA. *À medida que o carro passava, viam-se crianças a acenar na berma da estrada.* Ir pela +.

bom-serás [bõstráf]. s. m. sg. e pl. (De *bom* + forma do v. *ser*). *Fam.* Pessoa simplória, ingénua, que não faz mal a ninguém. ≈ BONACHEIRÃO, SIMPLÓRIO. *«Salram então dos caldeiros e fôrnlhas una meia dúzia de almas macarónicas, muito rotinhas, muito acanaveadas, umas tristes da maleita. Quem elas eram? Uma era o pai da Ana Malhada, um bom-serás, cose-dor de panelas, que levava a sua menina às costas de terra em terra com os utensílios do ofício.»* (AQUILINO, *Gavião*, p. 159).

descalçar [dĩ/kaʃár]. v. (Do lat. **discaleāre*). 1. Tirar aquilo que se tem calçado; despir os pés ou as mãos; tirar o calçado. ≠ CALÇAR, ENFIAR, PÔR. *Descalçou-se mal chegou a casa. Descalçou um chinelo e atirou-mo. Descalçou a criança e deitou-a.* + as botas, as luvas, as meias; + os sapatos. 2. Tirar aquilo que serve de apoio para que fique bem assente no chão; tirar o calço. *Descalçou a mesa e esta ficou a balançar.* 3. Tirar as pedras que cobrem o pavimento. ≈ DESCALCETAR, DESEMPEDRAR. ≠ CALCETAR, EMPEDRAR. 4. *Fam.* Tirar os recursos; deixar sem soluções, sem possibilidade de resolver alguma coisa; deixar descalço. ≈ DESAMPARAR, DESARMAR, DESPREVENTIR. **descalçar a bota**, resolver uma dificuldade.

bombeiro¹, a [bõbéjru, -ɐ]. s. (De *bomba²* + suf. *-eiro*). 1. Pessoa que faz parte de um corpo organizado de combate a incêndios; o que trabalha com bombas de incêndio. *Sapadores bombeiros. «Bombeiro, o pai, apagava incêndios nas matas, trazia as chamas na vista, televisão nem pensar, entretinha-se para adormecer com o Almanaque Cearense.»* (M. O. BRAGA, *Lua*, p. 95). bombeiro voluntário, o que pertence a uma corporação com a obrigatoriedade de acudir a incêndios, acidentes, unicamente por filantropia, corpo² de bombeiros. 2. *Bras.* Pessoa que faz ou conserta bombas, canos, bombeiro hidráulico, *Bras.*, canalizador. 3. *Bras. Fam.* Criança que, durante a noite, tem incontinência urinária.

besta² [béfɪɐ]. s. f. (Do lat. *bestia*). 1. Animal irracional, [...] besta de carga. 1. Animal usado para transporte de carregamentos. 2. *Deprec.* Pessoa que faz trabalhos que exigem grande esforço físico. [...]

13. As locuções (prepositivas, conjuncionais, adverbiais...) que contêm a palavra da entrada são apresentadas por ordem alfabética após a última acepção da mesma. Surgem destacadas em negro, com informação sobre a categoria gramatical em itálico. Incluem definição obrigatória e podem integrar, a título facultativo, sinónimos, antónimos, exemplos e abonações.

dali [dɐli]. *contr.* Contr. da prep. *de* com o adv. *ali*. 1. Indica ponto de partida de um movimento: daquele lugar, daquele ponto. *Fui dali para o médico, sem passar por casa.* 2. Indica pessoa, entidade ou situação acabadas de referir e que constituem a fonte ou origem de alguma coisa. *Eu bem insisti junto dos serviços, mas parece que dali não virá a solução do problema. O meu avô está tão mal que os médicos já não esperavam nada dali.* *dali a nada*, *loc. adv.*, muito pouco tempo depois. *Dali a nada estava ele a chamar-me.* *dali a pouco* (tempo), *loc. adv.*, pouco tempo depois. *Dali a pouco chegou o meu irmão.* *dali em diante*, *loc. adv.*, a partir de então, desde esse momento. *Devido a um esforço maior tirou boa nota: dali em diante foi sempre aluno aplicado.* *dali para a frente*, *loc. adv.*, o m. que dali em diante. *Dali para a frente nunca mais senti dificuldades.* *dali por diante*, *loc. adv.*, o m. que dali em diante. *Dali por diante resolveu-me sempre qualquer problema.*

14. Após as locuções, as expressões (idiomáticas, proverbiais...) — locuções verbais, frases e outras fórmulas fixas — que contêm a palavra da entrada estão listadas em negro por ordem alfabética e seguidas da respectiva definição.

dura [dúre]. *s. f.* (Deriv. regres. de *durar*). 1. Qualidade do que resiste ao tempo e ao uso. = DURABILIDADE. 2. Espaço de tempo entre o princípio e o fim de uma coisa. = DURAÇÃO. *ser de pouca dura*, *durar pouco tempo*; passar depressa. *Foi amor de pouca dura.* *ser sol de pouca dura*, *ser algo que, por ser bom ou agradável, dura pouco tempo.* *Aquele bom-humor do chefe foi sol de pouca dura, pois começou logo a resmungar.*

15. Informações especiais, de ordem flexional ou derivacional, são apresentadas no fim, tais como indicações sobre aumentativos, diminutivos, plurais dos compostos hifenizados, plurais irregulares, com transcrição fonética do plural em palavras que registem alterações vocálicas no plural, formação de advérbios derivados em *-mente*, com remissão em (V.) quando os mesmos têm direito a entrada própria no dicionário.

abreviado¹, **a** [ɐbrɨvjádu, -ɐ]. *adj.* (Do part. pas. do v. *abreviar*). 1. Que tem pouca duração ou extensão. = CURTO, REDUZIDO. ≠ LONGO. *Executou a obra em abreviado tempo.* 2. Que se tornou mais curto, retirando partes consideradas menos importantes; de que se fez um resumo, que se abreviou. = CONCISO, RESUMIDO, SUCINTO. ≠ ALARGADO, DESENVOLVIDO. *Eles apresentaram uma versão abreviada dos acontecimentos. Relato abreviado.* **edição**¹ **abreviada**. 3. Que foi escrito com abreviaturas. *Escrita abreviada.* Adv. abreviadamente (V.).

amoroso², **a** [ɐmurózu, -óze]. *s.* (De *amor* + suf. *-oso*). 1. Pessoa que está apaixonada, que se entrega ao amor. = APAIXONADO, ENAMORADO. «*são os amorosos, enlaçados, misteriosos*» (J. DANTAS, *Ao Ouvido*, p. 86). 2. P. Us. Galã que desempenha o papel de apaixonado. *O filme acaba com a partida do amoroso.* Pl. amorosos, as [ɐmurózu f., -v f.].

bom-bocado [bõbukádu]. *s. m. Cul.* Doce feito de açúcar, gemas de ovos, coco ou amêndoas... Pl. bons-bocados.

16. Todos os símbolos e abreviaturas utilizados no dicionário se encontram listados, respectivamente nas pp. XXXVII e XXXI-XXXIII.

ARTÍCULOS DE MUESTRA

Lema (palabra estudiada), con moción de género.	cómico, ca. (Del lat. <i>comicus</i> , y este del gr. <i>κωμικός</i> <i>kōmikos</i>). adj. 1. Que divierte y hace reír. <i>Situación cómica</i> . 2. Perteneciente o relativo a la comedia. 3. Dicho de un actor: Que representa papeles cómicos . U. t. c. s. 4. Dicho de un autor antiguo: Que escribía comedias. U. t. c. s. ● m. y f. 5. comediante (actor). ○ f. 6. <i>Pan. historieta</i> (serie de dibujos). U. m. en pl. 7. <i>Pan. dibujos animados</i> . ■ ~ de la legua. m. y f. cómico itinerante que hacía sus representaciones de pueblo en pueblo. □ dar, o poner, la ~ . locs. verbs. coloqs. <i>Ven.</i> Hacer el ridículo. ponerse ~ . loc. verb. coloq. <i>Ven.</i> Contrariar los deseos o aspiraciones de alguien. > tira ~, vis ~.	Separación entre el bloque de acepciones y el de formas complejas (véase SIGNOS).
Categoría gramatical.		Separación entre el bloque de combinaciones estables del lema y el resto de las formas complejas (véase SIGNOS).
Signo de palabra que representa al lema (véase SIGNOS).		
Inicio de los envíos a otros lemas (véase SIGNOS).	vis ¹ . (Del lat. <i>vis</i> 'fuerza', 'potencia'). ~ cómica . f. Capacidad de una persona, especialmente de un actor, para hacer reír.	Envío a un artículo que recoge una combinación estable del lema.
Número de la acepción.	ballenero, ra. adj. 1. Perteneciente o relativo a la ballena o a su industria. 2. Dicho de un barco: Destinado a la captura de ballenas. U. m. c. s. m. ● m. 3. Pescador de ballenas. ○ f. 4. Bote o lancha auxiliar que suelen llevar los barcos balleneros para arponear las ballenas.	Forma compleja (combinación estable) a la que remite el envío anterior.
Separación de las acepciones en bloques según su categoría gramatical (véase SIGNOS).		Separación de las acepciones en bloques según su subcategoría gramatical (véase SIGNOS).
Superíndice numérico diferenciador de palabras homónimas.	champán ¹ . (Del malayo <i>campán</i> , y este del chino <i>san pan</i> 'tres tablas'). m. Embarcación grande, de fondo plano, que se emplea en China, el Japón y algunas partes de América del Sur para navegar por los ríos.	
Definición sinónima que remite a un lema con superíndice.	champán ² . (De <i>Champagne</i> , comarca francesa). m. Vino espumoso blanco o rosado, originario de Francia.	Etimología.
Información geográfica y gramatical complementaria.	champaña . m. champán ² . En Col., Méx. y Ven., u. c. f.	
Glosa explicativa de una voz (en negrita) de la misma familia que la definida.	malar . (Del lat. <i>mala</i> 'mejilla' y -ar). adj. 1. <i>Anat.</i> Perteneciente o relativo a la mejilla. ● m. 2. <i>Anat.</i> pómulo (hueso y prominencia de la mejilla).	Definición sinónima seguida de glosa (enunciado abreviado de la acepción a la que se remite).
	pómulo . (Del lat. <i>pomulum</i> 'manzanita', por la forma). m. 1. Hueso y prominencia de cada una de las mejillas. 2. Parte del rostro correspondiente al pómulo .	Acepción íntegra a la que remite la glosa anterior.
Variante del lema.	verdugazo . m. Golpe dado con el verdugo (azote).	
	ozonósfera . (Tb. ozonósfera , <i>Am.</i>). f. <i>Meteor.</i> capa de ozono .	Información complementaria sobre la extensión geográfica de la variante.
	bikini . (Tb. biquini . ♦ Del ingl. <i>bikini</i> , y este de <i>Bikini</i> , nombre de un atolón de las Islas Marshall, con infl. de <i>bī</i> 'bi-', por alus. a las dos piezas). m. Prenda femenina de baño compuesta de un sujetador y una braga. En Arg., u. c. f.	

Envío de una variante a su lema.	biquini. <i>V. bikini.</i>	
	pispás. en un ~. (Tb. en un pis pas. ♦ De or. expr.). loc. adv. coloq. en un santiamén.	Variante de forma compleja.
	pis². en un ~ pas. <i>V. pispás.</i>	
Envíos contiguos agrupados.	substancia..., sustantivo, va. <i>V. sustancia..., sustantivo.</i>	
Separación de enunciados dentro del paréntesis informativo (véase signos).	poeta, tisa. (Del lat. <i>poēta</i> , y este del gr. ποιητής <i>poiētēs</i> ; para la forma f., cf. fr. mediev. <i>poétisse</i> . ♦ Para el f., u. t. la forma <i>poeta</i>). m. y f. 1. Persona que compone obras poéticas. 2. Persona dotada de gracia o sensibilidad poética.	Transliteración, en la etimología, de una voz griega.
		Información morfológica.
Información complementaria sobre uso antonomástico.	inmaculado, da. (Del lat. <i>immaculātus</i> . ♦ Escr. con may. inicial en acep. 2). adj. 1. Que no tiene mancha. ● f. 2. por antonom. La Virgen María. La Inmaculada.	Información ortográfica referida a una acepción del artículo.
Contorno verbal (sujeto).	congresional. adj. <i>Ant., Col., EE. UU. y Nic.</i> Perteneciente o relativo al Congreso de una nación.	Información sobre usos de América y de España.
	famoseo. m. despect. coloq. <i>Esp.</i> Conjunto de gente famosa o de personas que buscan la fama a toda costa apareciendo en los medios de comunicación.	
	recesar. tr. 1. <i>Perú.</i> Clausurar una cámara legislativa, una universidad, etc. ○ intr. 2. <i>Cuba, Méx., Nic. y Perú.</i> Dicho de una corporación: Cesar temporalmente en sus actividades.	
Contorno de adjetivo (sustantivo del que se predica).	combi. (Del ingl. <i>combi</i> , y este acort. de <i>combination</i> 'combinación'). adj. 1. Dicho de una furgoneta o de un furgón: Que tiene un volumen interno dividido en espacio con asientos y espacio para carga. U. t. c. s. 2. Dicho de un electrodoméstico: Que tiene dos puertas y combina frigorífico y congelador. U. t. c. s. m.	
Ámbito o circunstancia a que una acepción se restringe.	sudista. adj. 1. En la guerra de Secesión de los Estados Unidos de América, partidario de los Estados Confederados del Sur. Apl. a pers., u. t. c. s. 2. Perteneciente o relativo a los sudistas o a los estados que defendían.	Información gramatical complementaria.
	biomecánico, ca. (La forma f., del fr. <i>biomécanique</i> , de bio- 'bio-' y <i>mécanique</i> 'mecánica'). adj. 1. Perteneciente o relativo a la biomecánica. ● f. 2. Estudio de la aplicación de las leyes de la mecánica a la estructura y el movimiento de los seres vivos.	Empleo del lema (en negrita) referido a acepciones anteriores o posteriores del mismo artículo.

Ejemplos (en cursiva).	carente. (Del lat. <i>carens</i>, -<i>entis</i>). adj. Que carece de algo. <i>Una obra carente DE originalidad.</i> equidistar. (De <i>equi-</i> y <i>distar</i>). intr. Dicho de dos personas o cosas con relación a otra, o de una persona o cosa con relación a otras: Estar a la misma distancia. <i>Los puntos de una parábola equidistan DE su directriz.</i> U. t. en sent. fig. <i>Un planteamiento que equidista DEL idealismo y DEL pragmatismo.</i>	Información sobre las preposiciones (en versalita) con las que se suele combinar la palabra.
Entrada propia para un superlativo irregular.	recentísimo, ma. adj. sup. de reciente.	
Información sobre los superlativos.	reciente. (Del lat. <i>recens</i>, -<i>entis</i>. ♦ Sup. irreg. recentísimo; reg. recentísimo). adj. 1. Nuevo, fresco o acabado de hacer. 2. Que ha sucedido hace poco. ● m. 3. <i>And.</i> levadura (hongo unicelular).	Información sobre uso regional de España.
Información acerca de la vigencia de la acepción.	punción. (Del lat. <i>punctio</i>, -<i>ōnis</i> 'acción de punzar'). f. 1. <i>Med.</i> Introducción de un instrumento agudo, como un trocar o una aguja, en un tejido, órgano o cavidad. 2. desus. punzada (dolor agudo). nártex. (Del gr. bizant. <i>νάρθήξ nárthēx</i>). m. <i>Arg.</i> Atrio o vestíbulo situado a la entrada de las iglesias paleocristianas y bizantinas.	Especialidad a la que corresponde la palabra o la acepción.
Información acerca de la intención del hablante.	titulitis. f. despect. coloq. Valoración desmesurada de los títulos y certificados de estudios como garantía de los conocimientos de alguien. sartén. (Del lat. <i>sartāgo</i>, -<i>īnis</i>). f. 1. Recipiente de cocina, generalmente de metal, de forma circular, poco hondo y con mango largo, que sirve sobre todo para freír. <i>En muchos lugares de Am. y Esp., u. c. m.</i> 2. Lo que se fríe de una vez en la sartén. ■ tener alguien la ~ por el mango. loc. verb. coloq. Ser dueño de la situación, poder decidir o mandar. > fruta de ~.	Información geográfica y gramatical complementaria.
Nivel de uso.	mor. (Afér. de <i>amor</i>). por ~ de . loc. prepos. cult. Por causa de.	
Definición sinonímica.	noctívago, ga. (Del lat. <i>noctivāgus</i>). adj. poét. noctámbulo. Apl. a pers., u. t. c. s.	Ámbito de la voz.
Sujeto de la locución verbal (sin negrita).	vistazo. (Der. de <i>vista</i>). m. Mirada superficial o ligera. ■ dar, o echar, alguien un ~ a algo. locs. verbs. 1. Cuidar de ello mirándolo de cuando en cuando, generalmente por encargo de otro. <i>Échale un vistazo al asado.</i> 2. Examinarlo, reconocerlo superficialmente.	Complemento de la locución verbal (sin negrita).

Información sobre irregularidades de conjugación del verbo.	predecir. (Del lat. <i>praedicere</i> . ♦ V. conjug. en APÉNDICE; part. irreg. predicho). tr. Anunciar por revelación, conocimiento fundado, intuición o conjetura algo que ha de suceder.	Envío al APÉNDICE de modelos de conjugación.
	predicho, cha. part. irreg. de predecir .	Entrada propia para un participio irregular.
	preterir. (Del lat. <i>praeterire</i> 'pasar adelante'. ♦ Conjug. c. <i>pedir</i> . ♦ U. solo las formas cuya desinencia empieza por -i). tr. 1. Hacer caso omiso de alguien o algo. 2. Der. Omitir en la institución de herederos a los que son forzosos, sin desheredarlos expresamente en el testamento.	Modelo de conjugación.
Extranjerismo (en cursiva).	kilobyte. (Voz ingl., de <i>kilo</i> - 'kilo-' y <i>byte</i>). m. <i>Inform.</i> Unidad equivalente a 1024 (2 ¹⁰) bytes. (Símb. kB).	Símbolo asociado a una acepción.
Locución latina (en cursiva).	ab initio. (Loc. lat.). loc. adv. 1. Desde el principio. 2. Desde tiempo inmemorial o muy remoto.	
	bici. (Acort.). f. coloq. bicicleta .	Acortamiento.
Afijos (prefijos y sufijos).	pre- . (Del lat. <i>prae-</i>). pref. Indica anterioridad local o temporal, prioridad o encarecimiento. <i>Prefijar, prehistoria, propósito, preclaro</i> .	
	-aico, ca. (Del lat. <i>-aicus</i>). suf. Forma adjetivos que indican pertenencia o relación. <i>Incaico, altaico, voltaico, algebráico</i> .	
Elementos compositivos (prefijos y sufijos).	hidro- . (Del gr. ὑδρο- <i>hydro-</i>). elem. compos. Significa 'agua'. <i>Hidroavión, hidrofobia</i> .	
	-logía. (Del gr. -λογία <i>-logía</i>). elem. compos. Significa 'tratado', 'estudio', 'ciencia'. <i>Mineralogía, lexicología</i> .	
	ONG. (Sigla de <i>organización no gubernamental</i>). f. Organización de iniciativa social, independiente de la Administración pública, que se dedica a actividades humanitarias, sin fines lucrativos.	Sigla.

protéger

<
>

PROTÉGER conjugaison **verbe transitif** (se conjugue comme *Alléger*).

xiv^e siècle. Emprunté du latin *protegere*, « couvrir, abriter ; protéger, garantir », lui-même composé de *pro*, « en avant, devant », et *tegere*, « couvrir ; cacher, protéger ».

1. Veiller sur quelqu'un, lui prêter secours et appui. *Que Dieu les protège ! Les chevaliers avaient le devoir de protéger les plus faibles. Protéger la veuve et l'orphelin, secourir les malheureux, les opprimés. Des policiers. La loi protège la personne aliénée.*



DICTIONNAIRE DE
L'ACADÉMIE
FRANÇAISE

Q

×

se protège en émettant de l'encre. Se protéger de la pluie, du froid ou contre la pluie, le froid. Au participe passé, adjectivement. Une espèce protégée, qu'on cherche à préserver de l'extinction. Passage protégé, passage délimité réservé aux piétons pour la traversée de la chaussée.

- Spécialement. Couvrir, gagner un objet, une chose pour qu'ils ne puissent constituer un danger. *En escrime, une mouche protège la pointe des fleurets, des épées, des sabres.*

2. Soutenir une personne, contribuer à sa fortune, s'intéresser à sa réussite ; veiller au développement, au progrès d'une entreprise. *Les Médicis protégèrent de grands artistes. Protéger les lettres, les sciences.*

- Au participe passé, subst. *User de son influence en faveur d'un protégé. Il appuie la candidature de sa protégée.*

Orthographe i

◇ Se conjugue ***je protégerai, je protégerais*** (futur et conditionnel), selon les rectifications orthographiques de 1990.

– [règle §3] Les accents et le tréma • accent grave (verbes en -é.er).

Voir aussi ^

- [FranceTerme] :
 - (Aménagement et urbanisme) **résidence protégée**

List of abbreviations of DLPC (2001) – domain labels

CLASSIFICAÇÃO DO VOCABULÁRIO QUANTO À REPARTIÇÃO POR CIÊNCIAS,
TÉCNICAS E FORMAS DE ACTIVIDADE**a**

<i>Acúst.</i>	= Acústica.
<i>Aeron.</i>	= Aeronáutica.
<i>Agr.</i>	= Agricultura.
<i>Alg.</i>	= Álgebra.
<i>Alveit.</i>	= Alveitaria.
<i>Alven.</i>	= Alvenaria.
<i>Anat.</i>	= Anatomia.
<i>Antr.</i>	= Antropologia.
<i>Apic.</i>	= Apicultura.
<i>Arit.</i>	= Aritmética.
<i>Arm.</i>	= Armaria.
<i>Arqueol.</i>	= Arqueologia.
<i>Arquit.</i>	= Arquitectura.
<i>Artilh.</i>	= Artilharia.
<i>Astr.</i>	= Astronomia.
<i>Astronáut.</i>	= Astronáutica.
<i>Astrol.</i>	= Astrologia.
<i>Autom.</i>	= Automobilismo.

b

<i>Bact.</i>	= Bacteriologia.
<i>Balist.</i>	= Balística.
<i>B.-Art.</i>	= Belas-Artes.
<i>Biol.</i>	= Biologia (Citologia, Histologia).
<i>Bot.</i>	= Botânica.
<i>Bromat.</i>	= Bromatologia.

c

<i>Carn.</i>	= Carniçaria.
<i>Carp.</i>	= Carpintaria.
<i>Cartog.</i>	= Cartografia.
<i>Cerâm.</i>	= Cerâmica.
<i>Chapel.</i>	= Chapelaria.
<i>Cibern.</i>	= Cibernética.
<i>Cineg.</i>	= Cinegética.
<i>Cinem.</i>	= Cinema, cinematografia.
<i>Cirurg.</i>	= Cirurgia.
<i>Comérc.</i>	= Comércio.
<i>Constr.</i>	= Construção.
<i>Contab.</i>	= Contabilidade.
<i>Corr.</i>	= Correios.
<i>Cosmol.</i>	= Cosmologia.
<i>Cristalog.</i>	= Cristalografia.
<i>Cronol.</i>	= Cronologia.
<i>Cosmol.</i>	= Cosmologia.
<i>Cul.</i>	= Culinária.
<i>Curt.</i>	= Curtumes.
<i>Cutel.</i>	= Cutelaria.

d

<i>Des.</i>	= Desenho.
<i>Desp.</i>	= Desporto.
<i>Diplom.</i>	= Diplomática.
<i>Dir.</i>	= Direito.
<i>Dir. Can.</i>	= Direito Canónico.
<i>Dir. Civil</i>	= Direito Civil.
<i>Dir. Comerc.</i>	= Direito Comercial.
<i>Dir. Fiscal</i>	= Direito Fiscal.
<i>Dir. Intern.</i>	= Direito Internacional.
<i>Dir. Marít.</i>	= Direito Marítimo.

e

<i>Ecl.</i>	= Eclesiástico.
<i>Econ.</i>	= Economia.
<i>Econ. Pol.</i>	= Economia Política.
<i>Electr.</i>	= Electricidade.
<i>Electrotéc.</i>	= Electrotécnica.
<i>Embr.</i>	= Embriologia.
<i>Encad.</i>	= Encadernação.
<i>Eng.</i>	= Engenharia.
<i>Equit.</i>	= Equitação.
<i>Esc.</i>	= Escolar.
<i>Escol.</i>	= Escolástica.
<i>Escult.</i>	= Escultura.
<i>Esg.</i>	= Esgrima.
<i>Espir.</i>	= Espiritualismo.
<i>Estát.</i>	= Estática.
<i>Ética.</i>	
<i>Etnog.</i>	= Etnografia.

f

<i>Farm.</i>	= Farmácia.
<i>Filol.</i>	= Filologia.
<i>Filos.</i>	= Filosofia.
<i>Fin.</i>	= Finanças.
<i>Fís.</i>	= Física.
<i>Fís. Atóm.</i>	= Física Atómica.
<i>Fisiol.</i>	= Fisiologia.
<i>Fonét.</i>	= Fonética.
<i>Fort.</i>	= Fortificação.
<i>Fot.</i>	= Fotografia.
<i>Fut.</i>	= Futebol.
<i>Futur.</i>	= Futurologia.

g

<i>Geneal.</i>	= Genealogia.
<i>Genét.</i>	= Genética.
<i>Geod.</i>	= Geodesia.
<i>Geog.</i>	= Geografia.
<i>Geol.</i>	= Geologia.

Geom. = Geometria.
Gram. = Gramática.

h

Heráld. = Heráldica.
Hidrául. = Hidráulica.
Hidrogr. = Hidrografia.
Hig. = Higiene.
Hist. = História.
Hist. Nat. = História Natural.
Histol. = Histologia.
Hortic. = Horticultura.

i

Ind. = Indústria.
Inform. = Informática.

j

Jard. = Jardinagem.
Jur. = Jurídico, jurisprudência.

l

Ling. = Linguística.
Literat. = Literatura.
Liturg. = Liturgia.
Lóg. = Lógica.

m

Magn. = Magnetismo.
Marc. = Marcenaria.
Mat. = Matemática.
Mecân. = Mecânica.
Med. = Medicina.
Med. Leg. = Medicina Legal.
Metal. = Metalurgia.
Meteor. = Meteorologia.
Métr. = Métrica.
Mil. = Militar.
Miner. = Mineralogia.
Mitol. = Mitologia.
Mús. = Música.

n

Náut. = Náutica.
Neur. = Neurologia.
Numism. = Numismática.

o

Ocult. = Ocultismo.
Olar. = Olaria.
Ópt. = Óptica.
Ouriv. = Ourivesaria.

p

Paleog. = Paleografia.
Paleont. = Paleontologia.

Parasit. = Parasitologia.
Pat. = Patologia.
Pec. = Pecuária.
Pedag. = Pedagogia.
Pesc. = Pesca, pescaria.
Pint. = Pintura.
Pisc. = Piscicultura.
Políc. = Polícia.
Polít. = Política.
Pré-Hist. = Pré-História.
Psicol. = Psicologia.
Psicofisiol. = Psicofisiologia.
Psiqu. = Psiquiatria.

q

Quím. = Química.
Quím. Org. = Química Orgânica.
Quím. Biol. = Química Biológica.

r

Radiol. = Radiologia.
Rel. = Religião.
Retór. = Retórica.

s

Sal. = Salinas.
Sapat. = Sapataria.
Seric. = Sericicultura.
Serr. = Serralharia.
Sider. = Siderurgia, siderotecnica.
Silv. = Silvicultura.
Sociol. = Sociologia.

t

Tanoar. = Tanoaria.
Taurom. = Tauromaquia.
Teat. = Teatro.
Tecnol. = Tecnologia.
Telecom. = Telecomunicações.
Telef. s. fios = Telefonia sem fios.
Teleg. = Telegrafia.
Teol. = Teologia.
Teos. = Teosofia.
Terat. = Teratologia.
Tip. = Tipografia.
Topog. = Topografia.
Trigon. = Trigonometria.

v

Veloc. = Velocipedia.
Ven. = Venatório.
Veter. = Veterinária.
Vin. = Vinificação.
Vit. = Viticultura.

z

Zool. = Zoologia.
Zootec. = Zootecnia.

List of abbreviations of DLE, 23rd edition (2014)

Abreviaturas y signos empleados

-
- Las abreviaturas escritas con mayúscula inicial aparecen siempre así en el Diccionario (por ej., «*Burg.*», «*Fórm.*»). Las que figuran con minúscula inicial, por el contrario, pueden aparecer escritas de este modo o con mayúscula («*apóc.*» / «*Apóc.*»).
 - Las abreviaturas que desarrollan palabras con variación de género pueden sustituir tanto al masculino como al femenino (por ej., «*it.*» puede equivaler a ‘italiano’ o a ‘italiana’).
 - Toda abreviatura a la que se posponga una *-s* pasa a designar el mismo concepto, pero expresado en plural (por ej., «*exprs.*» ‘expresiones’, «*locs.*» ‘locuciones’, etc.).
 - Algunas formas abreviadas corresponden a dos desarrollos posibles (por ej., «*fr.*», que puede equivaler a ‘francés’ o a ‘frase’). El contexto en que se presentan estas abreviaturas despeja cualquier duda acerca de su interpretación.
 - No se recogen en esta tabla abreviaturas y símbolos que aparecen en el Diccionario y son de uso general. Este tipo de abreviaciones se explican adecuadamente en la Ortografía publicada por la Real Academia Española.
-

Abreviaturas

a.	alto	afect.	afectivo
abl.	ablativo	afér.	aféresis
abrev.	abreviación	<i>Agr.</i>	agricultura
acep.	acepción	aim.	aimara
acort.	acortamiento	al.	alemán
acrón.	acrónimo	<i>Ál.</i>	Álava
act.	activo	<i>Alb.</i>	Albacete
acus.	acusativo	<i>Alm.</i>	Almería
<i>Acús.</i>	acústica	<i>Alq.</i>	alquimia
adapt.	adaptación; adaptado	alterac.	alteración
adj.	adjetivo	alus.	alusión
adv.	adverbio; adverbial	<i>Am.</i>	América
advers.	adversativo	<i>Am. Cen.</i>	América Central
<i>Aer.</i>	aeronáutica	amer.	americano

<i>Am. Mer.</i>	América Meridional	<i>Constr.</i>	construcción
<i>Anat.</i>	anatomía	contracc.	contracción
<i>And.</i>	Andalucía	copulat.	copulativo; verbo copulativo
ant.	anticuado; antiguo	<i>Córd.</i>	Córdoba
<i>Ant.</i>	Antillas	<i>C. Real</i>	Ciudad Real
antonom.	antonomasia	<i>C. Rica</i>	Costa Rica
<i>Antrop.</i>	antropología	<i>Cuen.</i>	Cuenca
apl.	aplicado	cult.	culto
apóc.	apócope	dat.	dativo
apos.	aposición	deformac.	deformación
<i>Ar.</i>	Aragón	dem.	demonstrativo
ár.	árabe	<i>Dep.</i>	deportes
arag.	aragonés	der.	derivado
<i>Arg.</i>	Argentina	<i>Der.</i>	derecho
<i>Arq.</i>	arquitectura	desc.	desconocido
<i>Arqueol.</i>	arqueología	despect.	despectivo
art.	artículo	desus.	desusado
ast.	asturiano	deter.	determinado
<i>Ast.</i>	Asturias	dialect.	dialectal
<i>Astrol.</i>	astrología	dim.	diminutivo
<i>Astron.</i>	astronomía	disc.	discutido
atóm.	atómico	distrib.	distributivo
aum.	aumentativo	disyunt.	disyuntivo
aux.	auxiliar; verbo auxiliar	<i>Ec.</i>	Ecuador
<i>Áv.</i>	Ávila	<i>Ecd.</i>	ecdótica
b.	bajo	<i>Ecol.</i>	ecología
<i>Bad.</i>	Badajoz	<i>Econ.</i>	economía
<i>Bal.</i>	Islas Baleares	<i>EE. UU.</i>	Estados Unidos
berb.	bereber	<i>Electr.</i>	electricidad; electrónica
<i>Bil.</i>	Bilbao	elem.	elemento
<i>Biol.</i>	biología	<i>El Salv.</i>	El Salvador
<i>Bioquím.</i>	bioquímica	<i>Equit.</i>	equitación
<i>Bol.</i>	Bolivia	<i>Esc.</i>	escultura
<i>Bot.</i>	botánica	escri.	escrito
<i>Burg.</i>	Burgos	<i>Esgr.</i>	esgrima
c.	como	esp.	español
<i>Các.</i>	Cáceres	<i>Esp.</i>	España
<i>Cád.</i>	Cádiz	<i>Estad.</i>	estadística
<i>Can.</i>	Canarias	estud.	estudiantil
<i>Cantb.</i>	Cantabria	etim.	etimología
<i>Carp.</i>	carpintería	eufem.	eufemismo; eufemístico
<i>Cast.</i>	Castilla	excl.	exclamativo
cat.	catalán	expr.	expresión; expresivo
<i>Cat.</i>	Cataluña	ext.	extensión
celtolat.	celtolatino	<i>Ext.</i>	Extremadura
cf.	<i>confer</i>	f.	femenino; nombre femenino
cient.	científico	fest.	festivo
<i>Cineg.</i>	cinagética	fig.	figurado
<i>Cinem.</i>	cinematografía	<i>Fil.</i>	filosofía
clás.	clásico	<i>Filip.</i>	Filipinas
<i>Col.</i>	Colombia	<i>Fís.</i>	física
coloq.	coloquial	<i>Fisiol.</i>	fisiología
<i>Com.</i>	comercio	<i>Fon.</i>	fonética; fonología
comp.	comparativo	Fórm.	fórmula
compos.	compositivo	<i>Fotogr.</i>	fotografía
conc.	concesivo	fr.	francés
condic.	condicional	fr.	frase
conj.	conjunción	frec.	frecuentativo
conjug.	conjugación	frec.	frecuentemente
conjunt.	conjuntivo	fut.	futuro

<i>Gal.</i>	Galicia	m.	masculino; nombre masculino
gall.	gallego	[u.] m.	[usado] más
gallegoport.	gallegoportugués	m. or.	mismo origen
galolat.	galolatino	<i>Mad.</i>	Madrid
genit.	genitivo	<i>Mál.</i>	Málaga
<i>Geogr.</i>	geografía	malson.	malsonante
<i>Geol.</i>	geología	<i>Man.</i>	La Mancha
<i>Geom.</i>	geometría	<i>Mar.</i>	marina
ger.	gerundio	<i>Mat.</i>	matemáticas
germ.	germania	may.	mayúscula
germ.	germánico	<i>Mec.</i>	mecánica
gót.	gótico	<i>Med.</i>	medicina
gr.	griego	metapl.	metaplasmo
<i>Gram.</i>	gramática	metát.	metátesis
<i>Gran.</i>	Granada	<i>Meteor.</i>	meteorología
<i>Gran Can.</i>	Gran Canaria	<i>Métr.</i>	métrica
<i>Guad.</i>	Guadalajara	<i>Méx.</i>	México
guar.	guaraní	<i>Mil.</i>	milicia
<i>Guat.</i>	Guatemala	<i>Mit.</i>	mitología
<i>Guin.</i>	Guinea Ecuatorial	mod.	moderno
<i>Guip.</i>	Guipúzcoa	mozár.	mozárabe
hebr.	hebreo	<i>Mur.</i>	Murcia
<i>Heráld.</i>	heráldica	<i>Mús.</i>	música
hisp.	hispánico	n.	neutro
<i>Hond.</i>	Honduras	n. p.	nombre propio
<i>Huel.</i>	Huelva	<i>Nav.</i>	Navarra
<i>Hues.</i>	Huesca	neerl.	neerlandés
ilat.	ilativo	neg.	negación
imit.	imitación; imitativo	negat.	negativo
imper.	imperativo	<i>Nic.</i>	Nicaragua
imperf.	imperfecto	nórd.	nórdico
impers.	impersonal; verbo impersonal	núm.	número
<i>Impr.</i>	imprenta	<i>Numism.</i>	numismática
inc.	incierto	<i>occid.</i>	occidental
incoat.	incoativo	occit.	occitano
indef.	indefinido	onomat.	onomatopeya; onomatopéyico
indet.	indeterminado	<i>Ópt.</i>	óptica
indic.	indicativo	or.	origen
infant.	infantil	<i>orient.</i>	oriental
infinit.	infinitivo	<i>Ortogr.</i>	ortografía
infl.	influencia; influido; influjo	<i>Pal.</i>	Palencia
<i>Inform.</i>	informática	<i>Pan.</i>	Panamá
<i>Ingen.</i>	ingeniería	<i>Par.</i>	Paraguay
ingl.	inglés	<i>Parapsicol.</i>	parapsicología
intens.	intensivo	part.	participio
interj.	interjección; interjetivo	pas.	pasivo
interrog.	interrogativo	perf.	perfecto
intr.	intransitivo; verbo intransitivo	pers.	persona
inus.	inusual	person.	personal
irl.	irlandés	peyor.	peyorativo
irón.	irónico	<i>Pint.</i>	pintura
irreg.	irregular	pl.	plural
it.	italiano	poét.	poético
jap.	japonés	ponder.	ponderativo
jerg.	jerga; jergal	pop.	popular
lat.	latín; latino	port.	portugués
leng.	lenguaje	poses.	posesivo
leon.	leonés	pref.	prefijo
<i>Ling.</i>	lingüística	prep.	preposición
loc.	locución	prepos.	preposicional

pres.	presente	suf.	sufijo
pret.	pretérito	sup.	superlativo
<i>P. Rico</i>	Puerto Rico	sust.	sustantivo
prnl.	pronominal; verbo pronominal	t.	terminación
pron.	pronombre	[conj.] t.	[conjunción] temporal
pronom.	pronominal	[u.] t.	[usado] también
prov.	provenzal	<i>Taurom.</i>	tauromaquia
<i>Psicol.</i>	psicología	Tb.	también
<i>Psiquiatr.</i>	psiquiatría	<i>Tecnol.</i>	tecnologías
p. us.	poco usado	<i>Telec.</i>	telecomunicación
<i>P. Vasco</i>	País Vasco	<i>Ter.</i>	Teruel
<i>Quím.</i>	química	<i>T. lit.</i>	teoría literaria
<i>R. Dom.</i>	República Dominicana	<i>Tol.</i>	Toledo
ref.	referido	<i>Topogr.</i>	topografía
refl.	reflexivo	tr.	transitivo; verbo transitivo
reg.	regular	trad.	traducción
[marca] reg.	[marca] registrada	<i>Transp.</i>	transportes
regres.	regresivo	TV.	televisión
<i>Rel.</i>	religión	u.	usado
relat.	relativo	<i>Ur.</i>	Uruguay
<i>Ret.</i>	retórica	<i>Urb.</i>	urbanismo
rur.	rural	V.	véase
s.	sustantivo	<i>Val.</i>	Valencia
<i>Sal.</i>	Salamanca	<i>Vall.</i>	Valladolid
sánschr.	sánscrito	var.	variante
<i>Seg.</i>	Segovia	<i>Ven.</i>	Venezuela
sent.	sentido	verb.	verbal
<i>Sev.</i>	Sevilla	<i>Veter.</i>	veterinaria
Símb.	símbolo	<i>Vizc.</i>	Vizcaya
sínc.	síncopa	vocat.	vocativo
sing.	singular	vulg.	vulgar
<i>Sociol.</i>	sociología	<i>Zam.</i>	Zamora
<i>Sor.</i>	Soria	<i>Zar.</i>	Zaragoza
subj.	subjuntivo	<i>Zool.</i>	zoología

Signos

- * Forma hipotética
- ® Marca registrada (véase «La vigesimotercera edición del Diccionario...»).
- ~ Signo que sustituye al lema (véase «Advertencias»).
- || Separación entre acepciones dentro de un mismo bloque, separación entre formas complejas, separación entre acepciones de una misma forma compleja.
- (||) Contiene un enunciado que expresa condensadamente determinada acepción de otro artículo.
- ◆ Separación entre enunciados, dentro del paréntesis que subsigue al lema.
- Comienzo de un bloque de acepciones que comparten categoría gramatical.
- Comienzo de un bloque de acepciones que comparten subcategoría gramatical.
- Comienzo del bloque de formas complejas.
- Separación entre las dos secciones que puede haber dentro de las formas complejas (véase «La vigesimotercera edición del Diccionario...»).
- Comienzo, al final del artículo, del bloque de envíos a otros lemas.

List of abbreviations of DAF, 9th edition (2021)

Tableau des abréviations

Tableau des abréviations
utilisées dans le Dictionnaire (9^e édition)

Dans cette édition, le signe* indique les entrées nouvelles ;
le signe ◇ renvoie aux recommandations de 1990.
Note : de nombreuses abréviations ont été développées (mises au long) dans l'édition numérique.

Absolt.	Absolument.
Abusivt.	Abusivement.
ACOUST.	Acoustique.
Adj.	Adjectif ou adjective.
Adjt.	Adjectivement.
ADM.	Administration.
Adv.	Adverbe.
Advt.	Adverbialement
AÉRON.	Aéronautique.
AGRIC.	Agriculture.
AGRON.	Agronomie.
ALCH.	Alchimie.
ALG.	Algèbre.
ANAT.	Anatomie.
Anciennt.	Anciennement.
ANTHROPOL.	Anthropologie.
ANTIQ.	Antiquité.
APIC.	Apiculture.
Apr.	Après (une date ou un nom).
ARBOR.	Arboriculture.
ARCHÉOL.	Archéologie.
ARCHIT.	Architecture.
ARITHM.	Arithmétique.
ARTILL.	Artillerie.
ASTROL.	Astrologie.
ASTRON.	Astronomie.
Auj.	Aujourd'hui.
Av.	Avant (une date ou un nom).
BALIST.	Balistique.
BÂT.	Bâtiment.
BIOL.	Biologie.
BOT.	Botanique.
BX-ARTS.	Beaux-Arts.
CATHOL.	Catholique.
CH. DE FER.	Chemins de fer.
CHIM.	Chimie.
CHIR.	Chirurgie.
CHORÉGR.	Chorégraphie.
Class.	Classique.

Compl.	Complément.
COMPT.	Comptabilité.
CONFIS.	Confiserie.
Conj. de coord.	Conjonction de coordination.
Conj. de sub.	Conjonction de subordination.
COURSES	Courses de chevaux.
COUT.	Couture.
CUIS.	Cuisine.
D.	Densité.
Dém.	Démonstratif.
DÉMOGR.	Démographie.
Didact.	Didactique.
ÉBÉN.	Ébénisterie.
ÉCON.	Économie.
ÉLECTR.	Électricité.
ÉLECTRON.	Électronique.
Ellipt.	Elliptiquement.
ENSEIGN.	Enseignement.
ENTOM.	Entomologie.
ÉQUIT.	Équitation.
ESCR.	Escrime.
ESTH.	Esthétique.
ETHNOL.	Ethnologie.
ÉTHOL.	Éthologie.

Ex.	Exemple.
Exclam.	Exclamatif.
Expr.	Expression.
F.	Féminin.
Fam.	Familier, familièrement.
FÉOD.	Féodal ou féodalité.
Fig.	Au figuré.
FISC.	Fiscalité.
Fut.	Futur.
GASTRON.	Gastronomie.
Généralt.	Généralement.
GÉOD.	Géodésie.
GÉOGR.	Géographie.
GÉOL.	Géologie.
GÉOM.	Géométrie.
GRAMM.	Grammaire.
GRAV.	Gravure.
GRECQ.	Grecque.
HÉMAT.	Hématologie.
HÉRALD.	Héraldique.
HIPPOL.	Hippologie.
HIST.	Histoire.
HISTOL.	Histologie.
HORT.	Horticulture.

HYDROGR.	Hydrographie.
HYDROL.	Hydrologie.
HYG.	Hygiène.
Imp.	Imparfait.
Imper.	Impératif.
Impers.	Impersonnel ou en emploi impersonnel.
Ind.	Indirect.
Indéf.	Indéfini.
Indic.	Indicatif.
Inf.	Infinitif.
INFORM.	Informatique.
Interj.	Interjection.
Interr.	Interrogatif.
Intr.	Intransitif.
Intranst.	Intransitivement.
Inv.	Invariable.
Iron.	Ironique ou ironiquement.
Irrég.	Irrégulier.
J.-C.	Jésus-Christ.
JOAILL.	Joaillerie.
lat.	Latin ou latine.
LÉGISL.	Législation.
LINGUIST.	Linguistique.
Litt.	D'un emploi littéraire.

LITURG.	Liturgie.
Loc.	Locution.
Loc. adj.	Locution adjective.
Loc. adv.	Locution adverbiale.
Loc. conj.	Locution conjonctive.
Loc. prép.	Locution prépositive.
Loc. verb.	Locution verbale.
M.	Masculin.
M. at.	Masse atomique.
MATH.	Mathématiques.
MÉCAN.	Mécanique.
MÉD.	Médecine.
MÉTALL.	Métallurgie.
MÉTÉOR.	Météorologie.
MICROBIOL.	Microbiologie.
MILIT.	Art militaire ou vie militaire.
MINÉR.	Minéralogie.
MUS.	Musique.
MYTH.	Mythologie.
N.	Nom.
Néol.	Néologisme.
N. f.	Nom féminin.
N. m.	Nom masculin.
N° at.	Numéro atomique.

Numér.	Numéral.
NUMISM.	Numismatique.
OCCULT.	Occultisme
OCÉAN.	Océanographie.
ŒNOL.	Œnologie.
OPHTALM.	Ophtalmologie.
OPT.	Optique.
ORFÈVR.	Orfèvrerie.
OROGR.	Orographie.
PALÉONT.	Paléontologie.
Par affaibl.	Par affaiblissement de sens.
Par anal.	Par analogie.
Par exag.	Par exagération d'emploi.
Par ext.	Par extension.
Par méton.	Par métonymie.
Par opp.	Par opposition.
PATHOL.	Pathologie.
PÂTISS.	Pâtisserie.
PEINT.	Peinture.
Péj.	Péjorativement.
Pers.	Personne.
P. ET CH.	Ponts et Chaussées.
PHARM.	Pharmacie.
PHIL.	Philosophie.

PHON.	Phonétique.
PHOT.	Photographie.
PHYS.	Physique.
PHYS. NUCL.	Physique nucléaire.
PHYSIOL.	Physiologie.
Pl.	Pluriel.
Poét.	D'un emploi poétique.
POLIT.	Politique.
Pop.	D'un emploi populaire.
Poss.	Possessif.
P. p.	Participe passé.
P. présent	Participe présent.
Pr.	Pronom.
Pr. dém.	Pronom démonstratif.
Pr. indéf.	Pronom indéfini.
Pr. interr.	Pronom interrogatif.
Pr. pers.	Pronom personnel.
Pr. rel.	Pronom relatif.
Préf.	Préfixe.
PRÉHIST.	Préhistoire.
Prép.	Préposition.
Prés.	Présent (dans la conjugaison).
Pron.	Forme pronominale ou emploi pronominal.
PROS.	Prosodie.
Prov.	Proverbe.

PSYCHAN.	Psychanalyse.
PSYCHIATR.	Psychiatrie.
PSYCHOL.	Psychologie.
PSYCHOPATHOL.	Psychopathologie.
PSYCHOSOCIOL.	Psychosociologie.
RADIOGR.	Radiographie.
RADIOPH.	Radiophonie.
RELIG.	Religion.
RHÉTOR.	Rhétorique.
ROM.	Romain ou romaine.
s.	Siècle.
SC.	Sciences.
SC. NAT.	Sciences naturelles.
SCOLAST.	Scolastique.
SCULPT.	Sculpture.
Sing.	Singulier.
SOCIOL.	Sociologie.
Spécialt.	Spécialement.
SPÉLÉOL.	Spéléologie.
STAT.	Statistique.
Subj.	Subjonctif.
Subst.	Substantif ou substantivement.
Suff.	Suffixe.
SYLVIC.	Sylviculture.

Symb.	Symbole.
Syn.	Synonyme.
TAUROM.	Tauromachie.
TECHN.	Technique.
TÉLÉV.	Télévision.
TÉLÉCOMM.	Télécommunications.
THÉOL.	Théologie.
Tr.	Transitif.
Transt.	Transitivement.
TRAV. PUBL.	Travaux publics.
Triv.	Trivial.
TYPOGR.	Typographie.
URBAN.	Urbanisme.
V.	Verbe.
V. déf.	Verbe défectif.
V. impers.	Verbe impersonnel.
V. intr.	Verbe intransitif.
V. pron.	Verbe pronominal.
V. tr.	Verbe transitif.
VÈN.	Vènerie.
VÉTÉR.	Vétérinaire.
VITIC.	Viticulture.
Vulg.	Vulgaire.
ZOOL.	Zoologie.