UNIVERSIDADE ESTADUAL PAULISTA - UNESP
Faculdade de Ciências e Letras - Campus de Araraquara

LUCAS CONSOLIN DEZOTTI

FONTES TRADICIONAIS E RECURSOS COMPUTACIONAIS NA PRODUÇÃO DE UM DICIONÁRIO DIGITAL LATIM-PORTUGUÊS

os protótipos *eLatim* e *eLatim-vallex*

Araraquara
2025

LUCAS CONSOLIN DEZOTTI

FONTES TRADICIONAIS E RECURSOS COMPUTACIONAIS NA PRODUÇÃO DE UM DICIONÁRIO DIGITAL LATIM-PORTUGUÊS

os protótipos *eLatim* e *eLatim-vallex*

Tese apresentada à Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Araraquara, para obtenção do título de Doutor em Linguística e Língua Portuguesa.

Linha de Pesquisa: Ensino-Aprendizagem de Língua

Orientador(a): Profa. Dra. Anise D'Orange Ferreira

Araraquara

2025

D532f

Dezotti, Lucas Consolin
Fontes tradicionais e recursos computacionais na produção de um dicionário digital latim-português : os protótipos eLatim e eLatim-vallex / Lucas Consolin Dezotti. -- Araraquara, 2025
246 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Araraquara
Orientador: Anise D'Orange Ferreira

1. Lexicografia. 2. Enciclopédias e dicionários História e crítica. 3. Dicionários eletrônicos. 4. Software livre. 5. Língua latina. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

IMPACTO POTENCIAL DESTA PESQUISA

A pesquisa provê metodologias para a construção de dicionários bilíngues latim-português inovadores, com impacto direto sobre o acesso ao conhecimento do léxico latino e potencial de servir de modelo para outros projetos de dicionário. O emprego de dados e código abertos favorece seu reúso na criação de outras ferramentas gratuitas, promovendo o acesso de estudantes e pesquisadores ao conteúdo bem como estimulando sua participação em projetos colaborativos voltados ao desenvolvimento de recursos linguísticos.

POTENTIAL IMPACT OF THIS RESEARCH

The research provides methodologies for building innovative Latin-Portuguese bilingual dictionaries, with a direct impact on access to knowledge of the Latin lexicon and the potential to serve as a model for other dictionary projects. The use of open data and open source foster the creation of free tools upon data reuse, providing information access for students and researchers as well as stimulating their participation in collaborative projects related to the creation of linguistic resources.

LUCAS CONSOLIN DEZOTTI

FONTES TRADICIONAIS E RECURSOS COMPUTACIONAIS NA PRODUÇÃO DE UM DICIONÁRIO DIGITAL LATIM-PORTUGUÊS

os protótipos *eLatim* e *eLatim-vallex*

Tese apresentada à Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Araraquara, para obtenção do título de Doutor em Linguística e Língua Portuguesa.

Linha de Pesquisa: Ensino-Aprendizagem de Língua

Data da defesa: 19/05/2025

Banca Examinadora:

_____
Profa. Dra. Anise D'Orange Ferreira
UNESP - Faculdade de Ciências e Letras - Campus de Araraquara

_____
Profa. Dra. Adriana Zavaglia
USP - Faculdade de Filosofia, Letras e Ciências Humanas

_____
Profa. Dra. Giovanna Longo
UNESP - Faculdade de Ciências e Letras - Campus de Araraquara

_____
Prof. Dr. Oto Araújo Vale
UFSCar - Centro de Educação e Ciências Humanas

_____
Profa. Dra. Regiani Aparecida Santos Zacarias
UNESP - Faculdade de Ciências e Letras - Campus de Assis

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Agradeço ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da UNESP, câmpus de Araraquara, por financiar através do programa CAPES/PRINT um estágio doutorado-sanduíche de seis meses junto ao Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione (CIRCSE) da Università Cattolica di Milano.

Gostaria de agradecer, nominalmente, às pessoas que de diversas maneiras me ajudaram a realizar este trabalho:

À professora Anise D'Orange Ferreira, por me abrir as portas para as Clássicas Digitais e me guiar por este labirinto de novidades.

Às professoras Giovanna Longo, Adriana Zavaglia e Regiani Aparecida Santos Zacarias, bem como ao professor Oto Araújo Vale, pela leitura atenta e pelas contribuições valiosas oferecidas durante a defesa desta tese.

Às professoras Chiara Palladino, Maria Cristina Parreira e Adriane Ottaiano, pelas críticas e sugestões fornecidas durante os Seminários de Linguística (SELIN/UNESP).

A Marco Passarotti, Francesco Mambrini, Eleonora Litta, Flavio Cecchini, Federica Iurescia, Riccardo Ginevra, Claudia Corbetta, Giovanni Moretti, Rachele Sprugnoli, Paolo Ruffolo, Valeria Boano, Matteo Pellegrini, Martina Verdelli, por fazer com que minha estadia na Itália fosse plena de acolhimento e aprendizado.

À professora Ligeia Lugli, pelas aulas de programação em R imprescindíveis para o trabalho com dados empreendido nesta tese.

A Maria Celeste Consolin Dezotti, pela paciência de professora para ouvir minhas elucubrações e pelo carinho de mãe para confortar minhas angústias.

A José Dejalma Dezotti, pela inestimável contribuição do tradutor e pela estimada amizade de pai.

A Paula de Souza Paes, por estar ao meu lado nessa trajetória, sempre com amor, compreensão e apoio incondicional.

A toda a minha família,

MUITO OBRIGADO!

Horum igitur quispiam dicet, quaedam a me copiosius locupletiusque tractari potuisse. [...]
Quis tam iniquus vt exigat in huiusmodi scripti genere ne quis omnino praetereatur locus?
Vt nihil non legeris, nihil non annotaris, nihil non apparaueris, itane statim
in tam immensa rerum turba succurrit, quod oportuit quoque oportuit loco?

Haverá então quem diga que alguns assuntos poderiam ter sido tratados
com mais profundidade e riqueza de detalhes. [...] Mas quem seria tão injusto
a ponto de exigir que nesse gênero de escrito não se deixe escapar absolutamente nada?
Mesmo que se tenha lido tudo, anotado tudo, organizado tudo, por acaso nos ocorre assim
prontamente aquilo que convém e onde convém, em meio a tão grande massa de informações?

— Erasmo de Roterdã, *Adágios* 3.1.1 (*Herculei labores*)

RESUMO

A pesquisa parte da percepção de que há uma carência de recursos digitais em Língua Portuguesa voltados para o estudo de latim, e assume como objetivo investigar formas de construir um dicionário digital latim-português. Para tanto, o estudo encontra fundamentação nos princípios da lexicografia teórica e prática, buscando definir os tipos de dados lexicais mais comuns, as necessidades de informação típicas dos usuários de dicionários e os requisitos fundamentais de um projeto de dicionário digital, quais sejam, a criação de uma base de dados lexicográfica e de uma interface de consulta. Um panorama histórico da lexicografia do latim, abrangendo desde os repertórios manuscritos até as bases lexicais, passando pela tradição de dicionários impressos, é realizado a fim de identificar fontes de informação passíveis de incorporação à base de dados. Em seguida, descreve-se a metodologia utilizada para a compilação de diversos tipos de fonte em uma única base de dados integrada, que inclui métodos de obtenção de dados de dicionários tradicionais digitalizados e outras bases lexicais, bem como extração de dados de *corpus* (envolvendo, nesse último caso, a construção de um *corpus* paralelo de textos literários, igualmente descrito). Por fim, apresenta-se a interface de consulta, desenvolvida com o auxílio do programa de escrita de dicionários *lexicographR*) a fim de garantir a consistência do processo de transformação de dados legíveis por máquina em verbetes para leitura humana. O principal resultado é um protótipo de dicionário digital latim-português intitulado **eLatim**,[i] que fornece informações lexicais sobre os mil lemas latinos mais frequentes. Um segundo protótipo, intitulado **eLatim-vallex**,[ii] consiste em uma ferramenta de consulta experimental que permite ao usuário extrair frases de um *corpus* paralelo sintaticamente anotado de textos clássicos, utilizando como critério de seleção as propriedades sintático-semânticas dos termos dependentes.

Palavras-chave: Latim. Recursos lexicais. Lexicografia computacional. Corpus paralelo. Sintaxe de dependência.

[i] Cf. <https://lucascdz.github.io/eLatim/>.

[ii] Cf. <https://github.com/lucascdz/eLatim-vallex/>.

ABSTRACT

The research is motivated by the lack of free, open-source lexical resources in Portuguese for studying Latin. Its goal is to address this gap by investigating the development of a Latin-Portuguese digital dictionary. To achieve this, the study first examines the theoretical and practical principles of electronic lexicography, focusing on types of lexical data, user reference needs, and project requirements, such as creating a lexicographical database and designing a user interface. In compiling the database, a comprehensive history of Latin lexicography is reviewed to identify potentially reusable data sources. These sources include monolingual manuscript repertories, Latin-Portuguese bilingual printed dictionaries, and the latest computational lexical resources. The methodology for creating the lexicographical database from various sources is outlined. This methodology includes methods for obtaining data from legacy dictionaries and lexical databases, as well as for corpus data extraction that involves building a parallel corpus. Finally, the design of the user interface is presented. It employs a dictionary writing system, known as *lexicographR*, to ensure consistency in data compilation. The primary outcome is a dictionary prototype called **eLatim**,[iii] which provides lexical information for the 1,000 most frequent Latin lemmas. Additionally, a second prototype named **eLatim-vallex**[iv] offers a query tool that sorts sentences from a parallel corpus based on the grammatical features of dependent words.

Keywords: Latin. Lexical Resources. Computational Lexicography. Parallel corpus. Universal Dependencies.

[iii] Cf. <https://lucascdz.github.io/eLatim/>.

[iv] Cf. <https://github.com/lucascdz/eLatim-vallex/>.

LISTA DE ILUSTRAÇÕES

Figura 1 – Tipologia das decisões envolvidas no planejamento de um dicionário. 26
Figura 2 – Modelo de uso de dicionário. 30
Figura 3 – Captura de tela com o verbete *uenio* no *Diccionario Didáctico Digital de Latín* (Márquez Cruz; Fernández-Pampillón, 2021). 73
Figura 4 – Início da letra D do manuscrito ALC-286. 96
Figura 5 – Excerto do *Lexicon* de Nebrija. 100
Figura 6 – Excerto do *Dictionarium* de Calepino. 101
Figura 7 – Excerto do *Dictionarium latinogallicum* de Estienne. 104
Figura 8 – Página do *Dictionarium latino lusitanicum* de Jerônimo Cardoso. 108
Figura 9 – Página do *Index totius Artis* de Antonio Velez. 112
Figura 10 – Página de *Raizes da lingua Latina* de Amaro de Roboredo. 120
Figura 11 – Página da *Prosodia in Vocabularium bilingue Latinum et Lusitanum* de Bento Pereira 125
Figura 12 – Página do *Parvum Lexicon Latinum* de Pedro José da Fonseca. 131
Figura 13 – Página do *Magnum lexicon Latinum et Lusitanum* de Cabral. 135
Figura 14 – Página do *Novissimo diccionario* de Saraiva. 138
Figura 15 – Montagem comparativa dos verbetes relativos ao verbo *accuso* nos dicionários de Quicherat e Saraiva, respectivamente. 141
Figura 16 – Página do *Diccionário Latim-Portuguez* da coleção “Do Povo”. 143
Figura 17 – Página do *Dicionário Latino-Portugues* de Torrinha. 147
Figura 18 – Colagem comparativa dos verbetes para *calor* nos dicionários de Gaffiot e Faria, respectivamente. 151
Figura 19 – Fluxograma de processamento de dados do aplicativo *lexicographR*. 167
Figura 20 – Gráfico dinâmico do tipo *collapsibleTree* mostrando equivalentes de tradução sugeridos pela WordNet para o lema *uideo*, agrupados por domínio semântico. 172
Figura 21 – Diagrama de relações entre elementos do módulo *ontolex*. 178
Figura 22 – Visualização gráfica da representação LOD do verbete *iter* em Fonseca (1798). 183
Figura 23 – Amostra de verbete no modo de exibição, a partir da formatação HTML. . . . 184
Figura 24 – Gráfico da representatividade do TranslatinumCor-pt quanto à variação diacrônica por século (em milhares de *tokens*). 194
Figura 25 – Gráfico da representatividade do TranslatinumCor-pt quanto à variação por gênero textual (em milhares de *tokens*). 195
Figura 26 – Gráfico de barras indicando a frequência do lema *facinus* normalizada por século. 199

Figura 27 – Nuvem de palavras com os candidatos a colocado da classe dos substantivos para o lema *senatus*. 200
Figura 28 – Reprodução de tela da página inicial do **eLatim**, com amostra do prefácio. . 206
Figura 29 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído de dicionários tradicionais. 206
Figura 30 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído de bases lexicais. 207
Figura 31 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (tabela de candidatos a colocados). 208
Figura 32 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (nuvem de substantivos candidatos a colocados). 208
Figura 33 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (exemplos de uso). 209
Figura 34 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (gráfico de frequência). 209
Figura 35 – Reprodução de tela do **eLatim-vallex**, com exemplos de uso do verbo *lego*. 211
Figura 36 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com sujeito abstrato. 212
Figura 37 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento no acusativo. 213
Figura 38 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por *ad*+acusativo. 213
Figura 39 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por *in*+nome abstrato no acusativo. 214
Figura 40 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por nome no dativo. 214
Figura 41 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por *in*+nome abstrato no acusativo. 215

LISTA DE TABELAS

Tabela 1 – Amostra de itens lexicais com suas propriedades no LiLa Lemma Bank. . . 157
Tabela 2 – Amostra de dados extraídos do PrinParLat para o lema verbal *facio*. 158
Tabela 3 – Amostra de dados extraídos da versão da Latin WordNet publicada na LiLaKB para o lema verbal *video*. 161
Tabela 4 – Amostra de dados extraídos do LatinVallex 2.0 para o lema verbal *sum*. . . . 161
Tabela 5 – Amostra da planilha de dados que serve de base para a compilação do dicionário. 169
Tabela 6 – Amostra de dados adaptados do *PrinParLat* para o lema verbal *facio*. 170
Tabela 7 – Amostra da planilha com dados extraídos de uma combinação de WordNets para o lema *video*. 171
Tabela 8 – Amostra de verbetes segmentados segundo a análise posicional de Merrilees (1996). 174
Tabela 9 – Amostra de verbete segmentado em tabelas relacionais. 180
Tabela 10 – Número de indivíduos por classe do modelo OntoLex em cada dicionário. . 181
Tabela 11 – Resultados quantitativos do pareamento de lemas entre os dicionários e o LiLa Lemma Bank. 182
Tabela 12 – Amostra de verbetes compilados e codificados em HTML para utilização no aplicativo *lexicographR*. 184
Tabela 13 – Relação das obras que compõem o protótipo do TranslatinumCor-pt. 192
Tabela 14 – Tamanho dos *corpora* consultados, delimitados por período (em milhões de *tokens*). 194
Tabela 15 – Relação das obras que compõem o protótipo do TranslatinumCor-ptree. . . 196
Tabela 16 – Frequência absoluta e normalizada (por século) de lemas iniciados por FACI–. 198
Tabela 17 – Candidatos a colocados para o lema *senatus*. 200
Tabela 18 – Amostra da planilha com exemplos extraídos automaticamente segundo o modelo GDEx, acompanhados de tradução e referência. 202

LISTA DE ABREVIATURAS E SIGLAS

CLP	Corpus Lexicográfico do Português.
CNME	Campanha Nacional de Material de Ensino.
CONLL-U	Conference on Natural Language Learning Universal data format.
DLG	Dictionarium latinogallicum.
DWS	Dictionary Writing System.
FGD	Functional Generative Description.
HTML	Hyper Text Markup Language.
JNE	Junta Nacional de Educação.
LOD	Linked Open Data.
LWN	Latin WordNet.
MWE	Multiword Expression.
OCR	Optical Character Recognition.
OLD	Oxford Latin Dictionary.
PDF	Portable Document Format.
PLN	Processamento de Linguagem Natural.
PWN	Princeton WordNet.
RDF	Resource Description Framework.
RDFS	Resource Description Framework Schema.
SKOS	Simple Knowledge Organization System.
SPARQL	SPARQL Protocol and RDF Query Language.
TEI	Text Encoding Initiative.
UD	Universal Dependencies.
URI	Uniform Resource Identifier.
W3C	World Wide Web Consortium.

CONVENÇÕES TIPOGRÁFICAS

Maiúscula	inicia nomes próprios e nomes de classes gramaticais e ontológicas.
itálico	destaca formas lexicais em sua representação ortográfica (incluindo lemas e expressões multipalavra), bem como palavras estrangeiras e títulos de publicações impressas ou digitais.
negrito	destaca nomes dos produtos resultantes da pesquisa.
VERSALETE	destaca siglas, morfemas, segmentos de palavras e traços semânticos.
SansSerif	destaca termos técnicos.
‘aspas simples’	demarcam significados e traduções literais.
“aspas duplas”	demarcam expressões e sentenças usadas como exemplo, bem como citações[v] de outros autores.

[v] Nesse caso, as convenções acima não se aplicam, uma vez que a formatação do original é preservada.

SUMÁRIO

1	**INTRODUÇÃO**	**17**
2	**FUNDAMENTOS PARA UM PROJETO DE DICIONÁRIO**	**24**
2.1	PRÉ-LEXICOGRAFIA	26
2.1.1	Características gerais	28
2.1.2	Formatos	31
2.1.3	Fontes de evidência	34
2.2	MACROESTRUTURA	36
2.2.1	Itens lexicais	37
2.2.2	Seleção da nomenclatura	40
2.2.3	Arranjo das entradas	43
2.3	MICROESTRUTURA	45
2.3.1	Formas canônicas	46
2.3.2	Formas variantes	48
2.3.3	Classes e propriedades gramaticais	49
2.3.4	Pronúncia	50
2.3.5	Etimologia	51
2.3.6	Definição e equivalentes	54
2.3.7	Informação paradigmática	58
2.3.8	Informação sintagmática	61
2.3.9	Informações sobre uso	73
2.3.10	Ordenação das acepções	75
2.4	CONCLUSÃO DO CAPÍTULO	76
3	**PANORAMA HISTÓRICO DA LEXICOGRAFIA DO LATIM**	**78**
3.1	REPERTÓRIOS MANUSCRITOS	79
3.1.1	Antiguidade latina	80
3.1.2	Idade Média europeia	86
3.1.3	O manuscrito alcobacense ALC-286	94
3.2	DICIONÁRIOS IMPRESSOS	98
3.2.1	Inventores de dicionário: Nebrija, Calepino, Estienne	98
3.2.2	O *Dictionarium* de Jerônimo Cardoso	105
3.2.3	O *Index* de Antônio Velez	111
3.2.4	As *Raizes* de Amaro de Roboredo	116
3.2.5	A *Prosodia* de Bento Pereira	121
3.2.6	O *Parvum lexicon* de Fonseca	128

3.2.7 O *Magnum lexicum* de Manuel de Pina Cabral . . . 133
3.2.8 O *Novissimo diccionário* de F. R. dos Santos Saraiva . . . 136
3.2.9 O *Dicionário Latim-Português* "do Povo" . . . 142
3.2.10 O *Dicionário Latino-Português* de Francisco Torrinha . . . 145
3.2.11 O *Dicionário Escolar* de Ernesto Faria . . . 148
3.3 LÉXICOS COMPUTACIONAIS . . . 152
3.3.1 Dicionários tradicionais em formato digital . . . 153
3.3.2 Uma base de conhecimento interligado para o Latim . . . 156
3.3.3 Léxicos dinâmicos extraídos de *corpora* anotados . . . 162
3.4 CONCLUSÃO DO CAPÍTULO . . . 162

4 COMPILAÇÃO DE UMA BASE DE DADOS . . . 165
4.1 DADOS DE LÉXICOS COMPUTACIONAIS . . . 166
4.2 DADOS DE DICIONÁRIOS TRADICIONAIS . . . 170
4.2.1 Princípio de segmentação do verbete . . . 172
4.2.2 Representação da informação lexical . . . 174
4.2.3 Identificação dos lemas . . . 181
4.2.4 Formatação para publicação . . . 183
4.3 DADOS EXTRAÍDOS DE *CORPUS* . . . 185
4.3.1 Fundamentos para a compilação de *corpora* . . . 185
4.3.2 A compilação de um *corpus* paralelo latim-português . . . 188
4.3.3 Extração e visualização de dados: frequência, colocações, exemplos . . . 197
4.3.4 Uma base de dados sintático-semânticos . . . 201
4.4 CONCLUSÃO DO CAPÍTULO . . . 203

5 CONSTRUÇÃO DE INTERFACES DE CONSULTA . . . 204
5.1 ELATIM: DICIONÁRIO DIGITAL LATIM-PORTUGUÊS . . . 204
5.2 ELATIM-VALLEX: FERRAMENTA DE PESQUISA DE *CORPUS* . . . 210

6 CONCLUSÃO . . . 216

REFERÊNCIAS . . . 219

1 INTRODUÇÃO

A importância do latim na história cultural e linguística mundial é um fato conhecido, dados os papéis que essa língua desempenhou em diferentes períodos da história ocidental: na Antiguidade, como a língua do Império Romano; na Idade Média, como a língua da Igreja Católica; na Modernidade, como a língua franca do letramento e das ciências. Embora tenha perdido gradativamente espaço nas práticas culturais e pedagógicas na contemporaneidade, o interesse pelo estudo do latim como língua histórica permanece ativo nas universidades brasileiras.

A experiência de mais de uma década como professor do curso de Línguas Clássicas da Universidade Federal da Paraíba nos trouxe a percepção de que os alunos de latim raramente dispõem de recursos lexicais suficientes e acessíveis para desenvolverem autonomia em seus estudos. De fato, mesmo que atividades e métodos de ensino empregados pelos professores normalmente costumam vir acompanhados de seus respectivos vocabulários, tarefas mais realistas como interpretar e traduzir textos literários dependem de recursos mais abrangentes, que forneçam descrição detalhada do maior número possível de palavras e suas diversas acepções. Porém, quando somos confrontados pelos alunos com a pergunta ‘qual é o melhor dicionário latim-português?’ ou ‘qual das opções disponíveis em livrarias vale a pena comprar?’, é difícil dar uma resposta definitiva ou que não soe evasiva: qualquer um, nenhum, depende. De fato, não há um levantamento histórico-comparativo detalhado de dicionários bilíngues latim-português impressos; nesse sentido, *qualquer um* é capaz de fornecer ao menos uma noção básica do significado, geralmente na forma de um ou mais equivalentes. Por outro lado, há críticas gerais a certos aspectos legados pela tradição lexicográfica, que apontam insuficiências tanto de cobertura do léxico quanto do tipo de descrição; dessa perspectiva, *nenhum* dicionário será completamente adequado para as diversas necessidades informativas que o trabalho com os textos pode requerer. Além disso, a escolha também *depende* de fatores como investimento financeiro disponível e das situações em que se pretende utilizar o dicionário, ambos normalmente condicionados pelas características físicas do volume impresso: de um lado, grandes dicionários são mais abrangentes do ponto de vista informativo, porém são mais caros, mais volumosos e mais pesados, portanto mais recomendados para uso doméstico; de outro, dicionários menores são mais restritos do ponto de vista da informação, porém costumam ser menos custosos de adquirir e de transportar para locais de trabalho ou estudo.

Entretanto, vivemos uma fase de consolidação de mudanças culturais profundas, especialmente no modo como percebemos e nos relacionamos com as redes e produtos digitais. Esse processo pode ser caracterizado como uma eversão do ciberespaço (Jones, 2014), isto é, uma espécie de ‘inversão para fora’, no sentido de que algo que antes era limitado a ambientes específicos acessíveis por meio de equipamentos especiais passa a constituir a moldura geral para diversas atividades da nossa vida cotidiana, no campo das relações sociais, do consumo, do tra-

balho, do lazer. Se a década de 2000 testemunhou uma onda de tecnologias impactantes, como o rápido avanço da computação e comunicação portátil (*notebooks*, *tablets* e, principalmente, *smartphones*), a possibilidade de geolocalização em tempo real (GPS e GIS), a ampla difusão de redes sociais através da Internet, a substituição da realidade virtual (VR) pela realidade aumentada (AR), as experiências com novas mídias e formas de lazer (*videogames* e *e-books*), atualmente a disseminação e quase onipresença dos *smartphones* favorece o emprego amplo de padrões de acesso a informações em rede e verificação de identidade (e.g. *QR codes* e análise biométrica) nos mais diversos tipos de prestação de serviços, incluindo os governamentais (*Gov.br*), base da organização da sociedade.

Essa espécie de migração em massa para o ecossistema digital também se faz sentir no âmbito do ensino e da produção acadêmica. De um lado, as novas tecnologias da comunicação e informação têm causado mudanças estruturais nas formas de ensinar e aprender (Kenski, 2007); no caso do ensino-aprendizagem de línguas, isso favorece o emprego de formas participativas de ensino-aprendizagem baseadas em atividades exploratórias e aprendizado por descoberta centradas na participação do aprendiz tanto em situações reais de comunicação linguística quanto em situações reais de produção colaborativa de conhecimento, promovendo maior autonomia e engajamento (Dee; Foradi; Šarić, 2016). De outro lado, o uso de métodos computacionais têm impactado a pesquisa científica em praticamente todas as áreas do conhecimento; no caso das ciências humanas, as Humanidades Digitais emergem como uma nova área acadêmica interdisciplinar que busca estimular "a investigação, o ensino, a publicação e a divulgação das Humanidades através da incorporação de ferramentas, fontes e métodos digitais" (Alves, 2016, §2), favorecendo a preservação e o acesso ao conhecimento dos mais diversos objetos da cultura material (e.g. por meio de modelagem 3D de cidades e esculturas) e imaterial (e.g. por meio da digitalização e reedição de textos), ao passo que a utilização de métodos computacionais trazem abordagens e resultados inovadores para a análise linguística (Linguística Computacional) e a produção lexicográfica (Lexicografia Computacional).

A adesão ao ecossistema digital é observável mesmo em áreas de caráter supostamente tradicionalista, como a dos Estudos Clássicos. Os novos campos das Clássicas Digitais e da Filologia Digital têm empreendido projetos colaborativos voltados, por exemplo, para a criação de uma biblioteca digital de acesso aberto a textos clássicos (Ferreira; Blackwell; Palladino, 2020) que possa ampliar o público atendido de especialistas e interessados em geral e, com isso, desfazer a imagem dos estudos clássicos como um nicho elitista (Romanello; Bodard, 2016). Para tanto, investe-se na criação de uma infraestrutura de recursos linguísticos utilizados para o tratamento digital dos textos, que inclui diversos tipos de processamento computacional como lematização, anotação morfossintática e alinhamento de tradução (Almas, 2017; Bamman; Crane, 2015; Blackwell; Crane, 2009). Para os propósitos desta tese, importa entender o lugar do dicionário como componente fundamental dessa infraestrutura e, especialmente, a forma produtiva com que os dados lexicais extraídos de dicionários impressos podem ser integrados às ferramentas de trabalho com textos.

Nesse sentido, a história digital do dicionário latim-inglês compilado por Lewis e Short (1879) é exemplar. Desde que a primeira edição digital (Lewis; Short, 1997) foi publicada pelo Perseus Project seguindo os padrões de estruturação de dados da Text Encoding Initiative (TEI),[1] seu conteúdo tem sido utilizado como fonte de informação lexical por diferentes aplicações, plataformas e *websites*, por exemplo: portais de consulta lexicográfica simultânea, em que figura ao lado de outros dicionários de latim (e.g. *Latinitium*, *Logeion*);[2] ferramentas de análise e anotação de textos, em que serve de referência para o trabalho de lematização (e.g. *Collatinus*, *Arethusa*);[3] e ambientes enriquecidos de leitura de textos (e.g. *Diogenes*, *Beyond Translation*),[4] em que a interpretação pode ser diretamente auxiliada por dados suplementares de natureza contextual-enciclopédica, analítica e lexical, entre outras.[5] Mais recentemente, a mesma edição digital do Lewis e Short (1997) foi usada como fonte de dados para uma nova edição digital (Mambrini et al., 2021a), dessa vez publicada segundo os padrões de estruturação da Web semântica (*Linked Open Data* ou LOD), aumentando seu potencial de interação com outros recursos linguísticos lexicais e textuais no mesmo padrão LOD e permitindo que seu conteúdo seja utilizado como parâmetro de referência de desambiguação em uma ferramenta de publicação de novos recursos coligados (*LiLa TextLinker*).[6]

Se o dicionário digitalizado tem papel fundamental no trabalho de anotação de textos, o produto desse trabalho fornece material para uma nova geração de recursos lexicais baseados em extração de dados de grandes *corpora* anotados. Para ficar no âmbito das línguas históricas, já existem léxicos valenciais[7] produzidos a partir de *treebanks* de textos gregos e latinos (McGillivray; Passarotti; Ruffolo, 2009; McGillivray; Vatri, 2015; Passarotti; Saavedra; Onambele, 2016), bem como léxicos dinâmicos produzidos a partir de *corpora* paralelos com tradução alinhada no nível da palavra (Bamman; Crane, 2015; Yousef; Berti, 2015). Observa-se assim um ciclo virtuoso entre tradição e inovação em termos de conhecimento sobre o léxico: dicionários tradicionais digitalizados são usados como fontes de informação para a anotação de textos, os quais se tornam fontes de dados para a produção de novos recursos lexicais. Pesquisas na área da Lexicografia Pedagógica também apontam as vantagens trazidas pela mídia eletrônica, pelo potencial de viabilizar a produção de dicionários com características inovadoras, entre as quais se podem citar a multifuncionalidade – capacidade de atender a diferentes tipos de usuário[8]

[1] Cf. <https://tei-c.org/>.

[2] Cf. <https://latinitium.com/> e <https://logeion.uchicago.edu/>, respectivamente.

[3] Cf. <https://outils.biblissima.fr/en/collatinus-web/> e <https://www.perseids.org/tools/arethusa/app/>, respectivamente.

[4] Cf. <https://d.iogen.es/> e <https://beyond-translation.perseus.org/>, respectivamente.

[5] Nesses casos, pode-se dizer que o dicionário experimenta uma espécie de "retorno às origens" enquanto auxiliar imediato para a compreensão de textos, como o haviam sido as glosas na época da cultura manuscrita e deixaram de sê-lo com a condensação da informação em glossários e dicionários (Tarp, 2019, p.11-13).

[6] Cf. <https://lila-erc.eu/LiLaTextLinker/>.

[7] Léxicos valenciais são coleções de lemas (normalmente verbos) descritos em termos de seus argumentos sintáticos, os quais são extraídos automaticamente de *corpora* sintaticamente anotados, também conhecidos como *treebanks*.

[8] É oportuno esclarecer que, neste trabalho, preferiu-se o termo usuário para se referir ao consulente de dicionários.

em diferentes situações lexicograficamente relevantes, com base na modularização das informações (Buyse, 2006; Duran; Xatara, 2006; Tarp, 2008) – e a multidimensionalidade – capacidade de ser integrado a outros recursos educacionais, como gramáticas, lematizadores, analisadores morfológicos e consulta a *corpus* (Duran, 2008; Verlinde, 2008).

Se um tal ecossistema de recursos sobre o léxico latino está disponível para pesquisadores e aprendizes de latim competentes em línguas estrangeiras (em inglês principalmente, mas também em outras línguas[9]), o mesmo não se pode dizer dos recursos disponíveis em Língua Portuguesa. O que mais se aproxima de uma ferramenta de consulta de dicionários latim-português é a versão digitalizada do *Dicionário Escolar Latino-Português* (Faria, 1962); porém, foi produzida de forma anônima sem qualquer documentação descritiva nem estruturação de dados, portanto, sem possibilidades de processamento computacional.[10] Uma segunda fonte é o *Corpus Lexicográfico do Português* – CLP (Verdelho; Silvestre, 2002), um conjunto de antigos dicionários digitalizados, boa parte dos quais dicionários bilíngues latim-português; também neste caso, os dados não são estruturados e as ferramentas de busca disponíveis são mais orientadas para a pesquisa de *corpus* (leitura e concordanciamento do texto dicionarístico) do que para a consulta lexical propriamente dita. Por fim, a porção bilíngue latim-português da plataforma colaborativa *Wikcionario*[11] não possui as dimensões (pois possui apenas 3,5 mil verbetes) nem a autoridade (pois seu conteúdo é anônimo) necessária para ser usada como fonte de dados suficiente e confiável para a leitura de um texto latino.

A produção de um dicionário eletrônico no par latim-português a partir de um dicionário tradicional foi investigada por Somer (Somer, 2016). Todavia, sua proposta desenvolve recursos relativos apenas à análise morfológica (tabelas de flexão), restringindo-se ao eixo paradigmático da língua e conservando-se bastante fiel à abordagem tradicional. Isso se observa, por exemplo, na tradução de formas isoladas, desprovidas de contexto sintático, e no emprego de técnicas tomadas à Linguística de corpus "não para fornecer informações semânticas ... mas para fazer uma listagem das unidades lexicais usadas, assim como para fazer a contagem delas" (Somer, 2016, p.17). Outras propostas recentes de dicionários bilíngues latim-português (Gregis, 2018; Arruda, 2019) não contemplam o formato digital, além de serem construídas sobre *corpora* específicos de textos, vinculados não à tradição literária, mas a determinados métodos de ensino.

Constata-se, assim, a necessidade de um dicionário digital latim-português que atenda as necessidades informativas de quem estuda ou lê textos em latim e, ao mesmo tempo, seja capaz

[9] O portal de dicionários Logeion, por exemplo, agrega dicionários de latim monolíngues e bilíngues que contemplam os pares latim-inglês, latim-francês, latim-holandês.

[10] O recurso consiste em uma página Web bastante simples, constituída apenas por um campo de busca, acompanhado do prefácio reproduzido da edição impressa (Faria, 2015). Não há créditos ou informações sobre o trabalho de digitalização; a própria data de publicação da versão digital (19.nov.2015) só pode ser obtida por métodos indiretos (pelo registro no sistema de buscas Google). Também de forma indireta, é possível ler nos metadados a seguinte declaração de princípios: "Tentamos manter a experiência de busca de palavras exatamente igual àquela da procura em dicionário físico, pois assim o aluno não se vicía [sic] com as automações e facilidades que a informática pode trazer" – o que denota uma visão bastante negativa das possibilidades facilitadoras da aprendizagem que o processamento digital pode proporcionar.

[11] Cf. <https://pt.wiktionary.org/w/index.php?title=Categoria:!Entrada_(Latim)>

de instrumentalizar as novas formas de trabalho com textos no ecossistema digital. Preencher essa lacuna no campo dos Estudos Clássicos em Língua Portuguesa constitui a principal motivação deste trabalho, que tem na busca por um modo de realizá-lo seu principal desafio.

A pesquisa parte de dois pressupostos básicos. O primeiro encontra ancoragem na afirmação de Biderman (2001, p.178) de que "as gerações futuras de dicionaristas deveriam partir de um tesouro lexicográfico já existente, procedendo ao expurgo dos seus erros a fim de refazê-lo segundo critérios da moderna ciência lexicográfica" e se traduz na valorização do conhecimento acumulado sobre o léxico latino registrado nos dicionários bilíngues da tradição impressa. O segundo é a confiança na capacidade dos recursos computacionais de prover soluções para as dificuldades implícitas na questão sobre o "melhor dicionário latim-português", vista acima, quais sejam:

- aumentar a portabilidade e a acessibilidade, pela produção de uma ferramenta de consulta online e gratuita;
- aumentar a abrangência, pela constituição de uma base de dados formada pela soma de informações fornecidas pelo conjunto de dicionários da tradição impressa;
- aumentar a qualidade da informação, pela complementação da informação tradicional tanto com dados fornecidos pelas bases lexicais computacionais quanto com dados extraídos de *corpora*.

Em conjunto, tais ideias impõem múltiplas questões de forma e de fundo. As do primeiro tipo incluem: quais dicionários bilíngues latim-português constituem nosso patrimônio lexicográfico? Quais estão disponíveis em formato digital estruturado, isto é, passível de processamento computacional? Como estruturar os dados dos que não estão? As do segundo incluem: que necessidades informativas se presume que os usuários de dicionários bilíngues de latim possuam? Quais dessas necessidades são contempladas pelos dicionários tradicionais? Como obter as informações que não o são? Acrescentam-se, ainda, questões sobre a forma de integrar dados extraídos de diversas fontes e os tipos de seleção e visualização de dados que podem ser oferecidos ao usuário como resultado de uma consulta.

Coloca-se, assim, como problema de pesquisa, a definição de princípios, parâmetros e procedimentos metodológicos necessários para produzir um dicionário digital bilíngue latim-português para acesso online gratuito. Tal objetivo requer uma série de decisões e ações subsidiárias, a saber:

- definir as necessidades informativas típicas do usuário de dicionários de latim;
- definir os parâmetros lexicográficos básicos para a construção do dicionário;
- inventariar e descrever, do ponto de vista lexicográfico, o patrimônio dicionarístico bilíngue latim-português da tradição impressa;

- transformar informação textual digitalizada em dados estruturados;
- pesquisar recursos computacionais de natureza lexical e textual que possam servir como fontes de dados complementares;
- selecionar uma ferramenta de extração dados de *corpora*;
- organizar os dados coletados das diversas fontes em uma base de dados unificada;
- selecionar e utilizar um sistema de escrita de dicionários para a produção padronizada de verbetes digitais.

A realização de tais ações resultou nesta tese, que se organiza em quatro capítulos.

O primeiro apresenta fundamentos teóricos para a elaboração de um projeto de dicionário digital bilíngue latim-português, denominado **eLatim**. O capítulo se organiza em torno das diversas decisões que se impõem ao lexicógrafo na construção de tal projeto, desde a definição do tipo de dicionário e do formato de publicação até a escolha dos elementos descritivos que devem fazer parte dos verbetes, passando obviamente pela definição da abrangência da nomenclatura e das formas de acesso ao conteúdo. Logo, princípios gerais de lexicografia e lexicologia, pontuados por uma reflexão sobre as características específicas do latim enquanto língua histórica, são mobilizados para definir as situações típicas de consulta ao dicionário, as necessidades informativas mais comuns e, consequentemente, os tipos de informação lexical a serem compilados, bem como a forma adequada de indexá-los, sempre com o objetivo de aumentar a eficácia e a pertinência da resposta oferecida ao usuário.

O segundo capítulo faz um panorama histórico lexicografia do latim, a fim de identificar fontes de informação monolíngue e bilíngue (latim-português) sobre o léxico latino que possam ser reaproveitadas na constituição de um dicionário digital. O percurso cronológico tem início com os primeiros repertórios lexicais manuscritos produzidos na Roma antiga, acompanha seu desenvolvimento durante a Antiguidade Tardia e Idade Média, observa a intensa produção de dicionários na Idade Moderna, como resultado da invenção da imprensa, e, por fim, aponta as principais tendências da Lexicografia eletrônica contemporânea, de base computacional. Em cada fase, os diversos repertórios são apresentados de uma perspectiva macro e microestrutural, procurando relacionar, na medida do possível, as características internas mais marcantes ao seu contexto histórico de produção e recepção.

O terceiro capítulo apresenta os métodos empregados para a construção de uma base estruturada de dados lexicais, divididos conforme o tipo de fonte de informação a ser processada, quais sejam: bases de dados lexicais, dicionários tradicionais digitalizados e um *corpus* paralelo latim-português. Com efeito, cada tipo requer um conjunto de procedimentos específicos para que possam ser usadas na compilação dos verbetes, com diferentes níveis de complexidade e possibilidades de automatização. Assim, as bases de dados lexicais, por serem recursos computacionais já estruturados, normalmente demandam apenas a harmonização da forma de indexação

(i.e. dos lemas) e a tradução para o português da informação descritiva. Já os dicionários tradicionais digitalizados exigem estratégias de segmentação de texto, bem como de classificação e identificação de informações, para que seu conteúdo possa ser organizado em uma base de dados estruturada; logo, descrevem-se mais detidamente as etapas de realização desse trabalho, até então inédito no que se refere a dicionários bilíngues latim-português. O tratamento do *corpus* paralelo, por sua vez, envolve uma discussão sobre os requisitos, as fontes e os métodos utilizados para sua compilação (também inédita no que se refere ao par latim-português), bem como os procedimentos necessários tanto para a extração de dados sobre frequência, colocações e exemplos, quanto para a produção de visualizações gráficas a partir dos dados extraídos; nesse caso, destaca-se a utilização do *lexicographR*, um aplicativo de escrita de dicionários desenvolvido de forma simultânea (e, de certa forma, integrada a esta pesquisa).[12]

O quarto capítulo apresenta duas ferramentas digitais para a consulta de informações sobre o léxico latino, cuja elaboração constitui o resultado prático da pesquisa. Em um primeiro momento, descrevem-se os métodos empregados para a seleção e compilação padronizada de dados lexicais em verbetes eletrônicos, feitas com o auxílio do aplicativo *lexicographR*; desse processo resulta o **eLatim**,[13] um protótipo de dicionário digital latim-português que integra, em uma mesma microestrutura, os diversos tipos de informação disponíveis na base de dados, das entradas dos dicionários tradicionais aos dados extraídos do *corpus*. Em um segundo momento, descrevem-se os métodos utilizados na criação do **eLatim-vallex**,[14] um protótipo de aplicativo digital em formato ShinyApp que permite ao usuário não só observar diretamente o material linguístico reunido no *corpus* paralelo, mas também aplicar critérios sintático-semânticos para filtrar o material exibido na tela.

Por fim, as considerações finais trazem uma avaliação do que foi alcançado até o momento e elenca as possibilidades de aperfeiçoamento e de desenvolvimentos ulteriores. De fato, esta tese, longe de encerrar a pesquisa sobre as formas de registrar e veicular informações sobre o léxico latino, abre uma diversidade de portas e caminhos inovadores no que respeita a esse tema.

---

[12] O autor deste trabalho participou como *user tester* do projeto de desenvolvimento do aplicativo mencionado, intitulado *Democratizing Digital Lexicography*, coordenado por Ligeia Lugli e financiado pelo National Endowment for the Humanities (HAA-290402-23).

[13] Cf. <https://github.com/lucascdz/eLatim>.

[14] Cf. <https://github.com/lucascdz/eLatim-vallex>.

2 FUNDAMENTOS PARA UM PROJETO DE DICIONÁRIO

Existe, pelo menos desde Wiegand (1984), uma preocupação em considerar a Lexicografia como área do conhecimento claramente autônoma – por ter como objeto um produto tecnológico e social específico, o dicionário – e essencialmente prática, marcada tanto pela cientificidade quanto pela interdisciplinaridade. Restringindo o escopo aos dicionários de língua (em oposição aos dicionários especiais, terminológicos e enciclopédicos), o autor define Lexicografia linguística como "prática científica voltada para a produção de obras de referência sobre a língua […] que recorre, em suas atividades, aos resultados, métodos e teorias de diversas disciplinas acadêmicas, conforme o dicionário a ser produzido" (Wiegand, 1984, p.14, tradução própria).[1]

Em seu aspecto prático, as atividades lexicográficas podem ser divididas em três conjuntos, correspondentes às três fases do trabalho lexicográfico: (1) a elaboração de um projeto de dicionário; (2) a construção e o processamento de uma base de dados lexicográfica; (3) a escrita dos textos que formarão o dicionário. Em seu aspecto científico, Wiegand define os elementos que compõem uma teoria geral da prática lexicográfica,[2] destinada a fundamentar tal prática com as razões do conhecimento mobilizado e das escolhas feitas pelo lexicógrafo, materializadas no projeto de dicionário na forma de explicações a respeito:

a) dos propósitos do dicionário, geralmente derivados das necessidades comunicativas e cognitivas da sociedade;

b) do tipo de dicionário considerado adequado para satisfazer essas necessidades, incluindo seus componentes estruturais, isto é, os textos que materializam a descrição lexicográfica;

c) das conexões com outras teorias, especialmente, no caso da lexicografia linguística, das premissas tomadas à teoria lexical geral e à lexicologia específica das línguas objeto de descrição;

d) das conexões com a história da lexicografia, estabelecendo relações tanto com princípios tradicionais quanto com propostas inovadoras;

e) dos métodos utilizados para compilar uma base de dados lexicográfica, bem como para processar esses dados a fim de obter um arquivo lexicográfico, incluindo considerações sobre o papel da computação nesse trabalho.

---

1 No original: "Linguistic lexicography is scientific practice aimed at producing reference works on language, in particular dictionaries of language. Lexicographical activity has recourse to the results, methods and theories of various academic disciplines according to the type of reference work being produced."

2 Vale notar que essa teoria geral, juntamente com outras três sub-áreas de pesquisa – quais sejam, a história da lexicografia, o uso de dicionários e a crítica de dicionários – formam o que Wiegand (1984, p.13) denomina metalexicografia.

Um esquema tripartite semelhante se encontra em Atkins (2008). A autora divide o trabalho lexicográfico em três fases – planejamento, análise e síntese –, em que a primeira define as diretrizes teóricas e metodológicas que orientam a execução das outras duas. De um lado, para que o lexicógrafo possa investigar as propriedades e o comportamento das palavras e registrá-los na forma de um texto eletrônico estruturado (uma base lexicográfica), é preciso definir que fontes de evidência serão utilizadas, que podem variar de um amplo corpus eletrônico de textos ao conteúdo de dicionários já existentes. De outro, o lexicógrafo precisa ter um perfil de usuário e um tipo de dicionário previamente definido para extrair dessa coleção de fatos estruturados aqueles que considera particularmente relevantes para o dicionário que está produzindo. Em outras palavras, o projeto de dicionário deve não apenas identificar o tipo de dados a se registrar e como reconhecê-los, mas também definir em detalhes como o verbete deve ser construído e formatado. A propósito, importa notar que, com o advento das bases de dados eletrônicas, um mesmo conjunto de dados analíticos pode servir à elaboração de diversas sínteses, cada uma materializando um projeto de dicionário diferente voltado, por exemplo, para públicos específicos.

A autora descreve a elaboração desse projeto como um processo de tomada de decisão que incide sobre diversos tópicos, agrupados em três dimensões:

1. decisões preliminares (pré-lexicografia) a respeito das características gerais do dicionário e das fontes de evidência que serão utilizadas no trabalho;

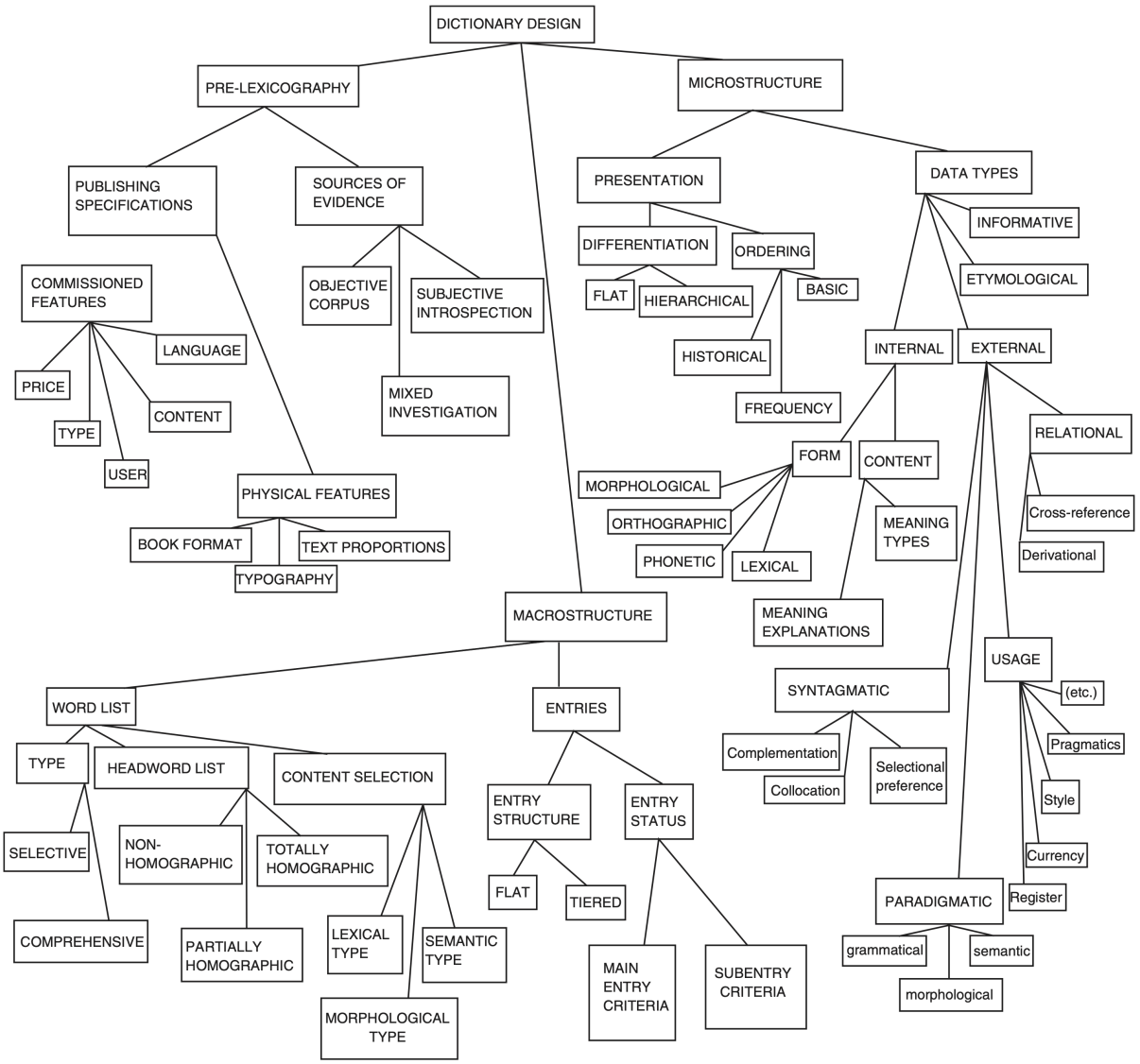
2. decisões macroestruturais, sobre os critérios para a composição e organização da lista de palavras a serem descritas;

3. decisões microestruturais, sobre o tipo de informação lexical fornecida pela descrição e a forma de organizar e apresentar tais informações.

A Figura 1 fornece o mapa conceitual desse conjunto de decisões que se impõem ao lexicógrafo na elaboração do projeto de dicionário.

Este primeiro capítulo trata da concepção de um projeto para o dicionário digital bilíngue latim-português, denominado **eLatim**, com base em decisões teoricamente justificadas a respeito daquelas três dimensões. Para cada tópico, buscou-se a contribuição de autores especializados para aprofundamento das discussões, a fim de adequar cada decisão às características e necessidades da atual "cultura do latim" – expressão com que se pretende designar os participantes e as práticas comunicativas que são (re)estabelecidas sempre que uma pessoa lê ou traduz um texto escrito em latim, seja em casa, numa biblioteca ou numa sala de aula, a despeito da distância espaço-temporal entre emissor e receptor. Uma vez que boa parte das decisões tem caráter linguístico, tornam-se particularmente relevantes as contribuições das teorias linguísticas voltadas para o léxico, a exemplo da lexicologia e da semântica lexical; outra parte depende de questões históricas e questões relacionadas com as tecnologias da informação. Assim, partindo de uma concepção de língua, procuram-se associar conhecimentos relativos ao processo

Figura 1 – Tipologia das decisões envolvidas no planejamento de um dicionário.



Fonte: Atkins (2008, p.36).

de ensino-aprendizagem, à análise linguística, à descrição lexicográfica, à pesquisa de corpus e ao uso de métodos computacionais e tecnologias digitais.

2.1 PRÉ-LEXICOGRAFIA

Segundo Atkins (2008), a pré-lexicografia diz respeito ao conjunto de determinações e especificações sobre o dicionário a ser produzido, que servirão de parâmetros para o trabalho de compilação e organização dos dados. Elas incluem:

a) características gerais, como língua(s), público-alvo, tipo de dicionário, tipo de conteúdo e preço final;

b) especificações editoriais, como formato do volume, cores e tipografia;

c) especificações quanto ao tipo de material utilizado como fonte de evidência linguística.

Antes de apresentar as decisões tomadas nesta fase de planejamento, é importante marcar algumas diferenças contextuais entre o modelo de Atkins (2008) e a produção do **eLatim**. O roteiro estabelecido pelo modelo pressupõe a publicação de um volume impresso produzido com motivações declaradamente comerciais[3]. Nesse contexto, o projeto de dicionário tem início com a identificação de uma lacuna no mercado de dicionários pelo departamento de *marketing* de uma editora (*publisher*), que define as características gerais do dicionário e contrata um lexicógrafo para atuar como coordenador da equipe de produção (*editor*). Tais definições são elaboradas tendo como fator principal o preço estimado para o produto final: "o preço condiciona todas as decisões relativas às características físicas do volume e, em particular, delimita o espaço disponível para os editores" (Atkins, 2008, p.36, tradução própria)[4].

O trabalho aqui desenvolvido, por sua vez, se insere em um contexto acadêmico, com financiamento público[5] e, portanto, orientado para a produção de conhecimento aberto, livre e gratuito[6]. Assim, todo o processo decisório fica a cargo do pesquisador-lexicógrafo, a começar pela identificação de uma lacuna nos recursos lexicais bilíngues publicados sob licença aberta: a carência de um dicionário eletrônico bilíngue latim-português. Como vimos, os dicionários disponíveis apresentam insuficiências ora ligadas à forma de acesso (digitalização em formatos de imagem, que não permitem utilizar texto como *input*), ora ligadas ao conteúdo acessado (não apresentam atualizações em relação à lexicografia tradicional). Além disso, o fator preço definido como zero implica na utilização de recursos igualmente gratuitos em todas as fases da produção, seja quanto às fontes de informação, seja quanto aos recursos técnicos necessários para publicar o dicionário na Web.[7]

Desse modo, algumas determinações gerais são estabelecidas *a priori* pelo pesquisador: trata-se de produzir um dicionário eletrônico de latim em formato acessível pela Web para um público falante de português. Daí decorrem as demais definições, condicionadas principalmente pela caracterização de um perfil de usuário padrão e de suas necessidades informativas em situa-

[3] A autora distingue explicitamente entre dicionários 'acadêmicos/históricos' (*scholarly and historical dictionaries*) e dicionários comerciais (*trade dictionary*), ressaltando sua grande *expertise* na produção destes e, de certo modo, questionando a viabilidade daqueles, na medida em que são planejados "com menos limitações de espaço e de tempo de conclusão, com uma tendência de parar na letra C ou demorar 50 anos para chegar lá" (Atkins, 2008, p.31). Em certa medida, demonstrar que um dicionário acadêmico/histórico pode ser feito em tempo hábil é um dos objetivos deste trabalho.

[4] No original: "[the price] constrains all decisions on the physical features of the book and, in particular, delimits the space available to the editors".

[5] Esta pesquisa foi realizada em uma universidade pública (UNESP/Araraquara) e foi possível graças a uma licença remunerada obtida pelo pesquisador junto à Universidade Federal da Paraíba, bem como ao apoio do programa CAPES/PrInt para um estágio de seis meses junto ao CIRCSE – Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione (Università Cattolica di Milano, Itália).

[6] Pinfield et al. (2021) mencionam dois relatórios recentes (de 2014 e 2019, respectivamente) que listam o Brasil entre os países que mais disponibilizam pesquisa acadêmica em regime de acesso aberto. Os autores apontam que parte desse protagonismo se deve a um modelo de comunicação acadêmica predominantemente baseado em periódicos financiados por instituições públicas e geridos pela comunidade, a partir de uma abordagem que se poderia chamar de "comunidade de conhecimento" (*knowledge commons*), em contraste com o "mercado de conhecimento" (*knowledge market*) predominante nos contextos europeu e norte-americano.

[7] Vale notar que boa parte desses recursos é publicada sob licenças que vedam seu uso em contexto comercial, de modo que sua utilização é uma vantagem praticamente exclusiva do contexto acadêmico.

ções de uso consideradas típicas. Dada a ausência de pesquisas de uso de dicionário específicas, essas informações serão construídas por raciocínio dedutivo, com base em pesquisas sobre uso de dicionários bilíngues em geral e, principalmente, em considerações sobre os contextos em que o latim é usado na atualidade.

2.1.1 Características gerais

O latim pode ser caracterizado de duas perspectivas. Do ponto de vista do sistema, o latim é uma língua natural, que pode ser definida como instrumento humano de cognição e comunicação, de natureza social e abstrata, que se realiza concretamente em atos linguísticos individuais com base em um sistema complexo dotado de pelo menos dois graus diferentes de abstração – o sistema funcional, como conjunto de possibilidades representativas e associativas válidas em uma língua, e o sistema normal, como conjunto de variantes de realização marcadas e não-marcadas, aparentemente facultativas (estilísticas), mas nunca indiferentes para o sentido, como sinônimos, expressões idiomáticas, frases feitas (Coseriu, 1967). Do ponto de vista do uso, o latim é uma língua histórica, porque desprovida de falantes nativos: suas realizações se resumem ao corpus de registros escritos que sobreviveram à ação do tempo. Nesse sentido, nas palavras de (Prado, 1992b, p.35), "o texto se nos oferece, arqueológica e metalinguisticamente, como enunciado de uma enunciação (viva) do passado, […] prova de que houve uma enunciação. […] O texto está para o estudante de latim assim como a ânfora de vinho está para o arqueólogo."

Esse *status* duplo tem pelo menos três implicações.

A primeira é a dificuldade de se recriar na língua escrita o que Coseriu chama de entornos da fala – "las circunstancias del hablar […] que pueden agruparse en cuatro tipos: situación, región, contexto y universo de discurso" (Coseriu, 1967, p.309), em oposição às operações de determinação, empregadas para "decir algo acerca de algo con los signos de la lengua" (Coseriu, 1967, p.291) –, com consequências não só para a interpretação literária, mas também nos planos teórico, descritivo e analítico da língua. Para uma língua de que só restaram signos escritos, torna-se particularmente importante basear ao máximo a interpretação em determinantes linguísticos: "apenas uma correta visão do sistema gramatical subjacente à língua dos romanos permite compreender outros níveis da expressividade".

A segunda reside no fato de que o conhecimento de uma língua sem falantes só pode ser obtido em contextos de ensino-aprendizagem. No Brasil, a principal instância de ensino-aprendizagem de latim são os cursos de graduação em Letras, particularmente os que oferecem habilitação em Latim ou Línguas Clássicas. Assim, sem excluir outras possibilidades que apenas uma enquete abrangente poderia revelar, pode-se estabelecer, como público-alvo de dicionários de latim, o seguinte perfil: pessoa adulta, segundo grau completo, cursando graduação em Letras ou profissional graduado, atuante quer como professor-pesquisador, quer como tradutor.

A terceira é que, na atualidade, o uso comunicativo do latim praticamente se restringe à leitura, interpretação e, eventualmente, tradução de textos escritos em seus diversos níveis (lin-

guístico, discursivo, histórico, estético, literário e cultural) (Lima, 1995; Prata; Fortes, 2015).[8] Com isso, mesmo quando se usa conversação simulada em sala de aula como estratégia didática (Leite, 2015), os objetivos gerais do ensino e da aprendizagem de latim se voltam especialmente para o desenvolvimento da competência receptiva na modalidade escrita (Bruno, 1992; Dezotti, 1992; Prado, 1992b). Tal competência, segundo (Widdowson, 1991), envolve duas habilidades complementares: (1) compreender, no sentido de reconhecer orações e seus significados enquanto elementos linguísticos (forma); e (2) ler, no sentido de reconhecer a força das frases em seu funcionamento enquanto partes de um discurso (uso). Assim, pode-se definir, como situação de uso de dicionário mais comum, uma atividade real ou simulada (isto é, para aprendizado) que envolva interpretação de signos e/ou tradução de textos escritos em latim.

Dicionários bilíngues são ferramentas reconhecidamente úteis para a resolução de problemas em atividades receptivas de língua estrangeira. Hartmann (2001) se refere a estudos que mostram vantagens de se usar esse tipo de dicionários tanto em atividades de aprendizado, memorização e retenção de vocabulário quanto em tarefas receptivas como leitura. Mesmo que, no segundo caso, o emprego de técnicas de dedução com base em contexto possam ser mais efetivas para a retenção de informações, a qualidade da informação retida tende a ser maior mediante o uso de dicionário, segundo o autor.

Nesse tipo de atividade, segundo Hannay (1996), o problema que normalmente se coloca ao usuário é a incompreensão do significado de certo item lexical da L2 em determinado contexto, de modo que a consulta ao dicionário segue um roteiro que vai do desconhecido (o item da L2) para o conhecido (sua língua materna). Não obstante a tradução ser uma atividade produtiva em língua materna, pode-se afirmar que o usuário de dicionários bilíngues orientados para recepção espera encontrar informações sobre os itens lexicais da língua estrangeira, mais do que sobre a sua própria. Por isso, um dicionário bilíngue voltado para a recepção deve prover "o retrato mais abrangente possível das propriedades fonéticas, semânticas, gramaticais e estilísticas do item lexical em questão" (Hannay, 1996, p.148).[9]

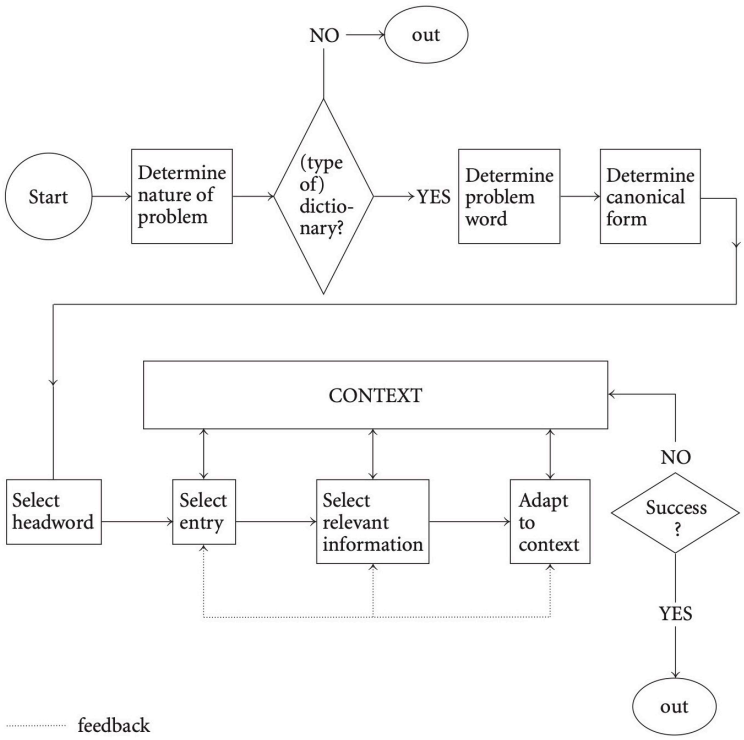
Contudo, como qualquer ferramenta, o dicionário exige habilidades específicas para ser usado de modo eficiente. Bogaards (1996) descreve o ato de consulta ao dicionário como parte de um processo dinâmico que tem início no momento em que um problema relacionado com a linguagem é encontrado e se desenvolve em quatro etapas, cada uma envolvendo uma tomada de decisão por parte do usuário (ver Figura 2).[10]. A primeira etapa é determinar a natureza do problema (se é uma questão conceitual, sintática, lexical, entre outras) e decidir se o dicionário (e que tipo de dicionário) é a melhor ferramenta para resolvê-lo. A segunda e a terceira consistem, respectivamente, em determinar a palavra problemática e deduzir sua forma canônica (isto

[8] Note-se que o uso do termo 'comunicativo' aqui diz respeito ao fato de haver um emissor-escritor e um receptor-leitor de uma mensagem escrita, não tendo qualquer relação com as chamadas abordagens comunicativas de ensino, baseadas sobretudo em práticas oralizadas em situações cotidianas.

[9] Detalhes sobre os tipos de informação requeridos por um dicionário bilíngue serão apresentados adiante, quando se tratar das decisões relacionadas à microestrutura.

[10] Cf. Hartmann (2001) para um modelo de consulta descrito em termos semelhantes.

Figura 2 – Modelo de uso de dicionário.



Fonte: Bogaards (1996, p.29)

é, a forma usada como entrada pelo dicionário), a fim de localizar no dicionário o verbete correspondente; caso existam duas ou mais entradas homógrafas, a seleção da entrada correta requer que se considere o contexto do item problemático. A quarta etapa, mais complexa, é encontrar a informação relevante para a solução do problema; nesse caso, a comparação de contextos será especialmente requisitada, visto que a aplicabilidade da informação encontrada no dicionário depende do grau de similaridade com o contexto do problema, exigindo maior ou menor esforço de adaptação. A consulta termina quando o usuário considera que a informação obtida é satisfatória – uma avaliação que pode não coincidir entre todos os usuários: "Usuários podem ficar satisfeitos com determinados resultados que um especialista na língua (e.g. um professor ou um lexicógrafo) saberia não ser correto" (Bogaards, 1996, p.30, tradução própria).[11]

Varantola (1994) ressalta o papel do usuário como "tomador de decisão" (*decision maker*), especialmente quando o dicionário é usado para fins de tradução: "O principal problema para os tradutores não é simplesmente encontrar respostas para suas pesquisas, mas ter certeza de que suas decisões, tomadas com base na informação encontrada, é correta e aceitável para o contexto" (Varantola, 1994, p.607).[12] É no sentido de evitar julgamentos equivocados por parte do usuário que se exige dos dicionário o maior número de informações contextuais possíveis, seja na forma de explicações, seja na forma de exemplos.

[11] No original: "Users may be satisfied with a particular result, whereas the expert (an adult native speaker, a foreign language teacher or a lexicographer) may know that the solution found is not correct".

[12] No original: "The main problem for the translators, in addition to finding answers to their queries, is how to be sure that the decision they make, on the basis of the information provided, is correct and acceptable for the context.".

A discussão apresentada sobre uso de dicionários de latim serve de guia para se pensar as demais decisões relativas ao plano do dicionário. O fato de se desenvolver num contexto acadêmico voltado para a produção de conhecimento aberto e gratuito orienta a escolha de um formato digital acessível através da Web e que funcione independente do dispositivo utilizado (celulares, *tablets*, computadores). A designação da recepção de textos escritos como a função principal do dicionário norteia a escolha de fontes de informação pertinentes ao *corpus* de textos escritos em latim, com preferência para os registros datados do período da Antiguidade Clássica, quando o latim tinha *status* de língua materna. Por fim, a forma de indexar as palavras no dicionário (decisões macroestruturais) bem como a forma de selecionar e organizar os elementos da descrição lexical em cada verbete (decisões microestruturais) são definidas tendo em vista o percurso de consulta estabelecido pelo modelo de Bogaards (1996).

2.1.2 Formatos

A revolução digital trouxe novas formas de produzir dados lexicográficos bem como novas formas de acessá-los e utilizá-los. Tarp (2019) resume as quatro grandes tendências da Lexicografia nesse novo contexto:

1. transformação do dicionário impresso em dicionário *digital* (especialmente de internet);
2. transformação do dicionário padronizado em dicionário *personalizado*, que se adapta às necessidades de seu usuário em cada consulta concreta;
3. transformação do dicionário para consulta isolada em dicionário *integrado* a outras ferramentas; e
4. substituição do dicionário como tal por *dados lexicográficos*, que se usam para diferentes fins, não necessariamente adquirindo a forma de dicionário.

O último ponto merece destaque, segundo o autor:

> a lexicografia deve se conscientizar de que seu produto por excelência são os dados lexicográficos e que estes podem ser usados para uma infinidade de aplicações que vão desde sua apresentação em dicionários até sua integração em diferentes tipos de ferramentas, assistentes, plataformas, serviços, etc. (Tarp, 2019, p.15, tradução própria).[13]

Para Steurs et al. (2020), a viabilização desse enorme potencial de reuso dos dados lexicográficos depende de uma mudança de objetivos da lexicografia acadêmica, de uma disciplina focada na produção de dicionários específicos para uma disciplina focada na construção de bases de dados e infraestruturas centralizadas que possam servir tanto para a produção de dicionários

[13] No original: "la lexicografía antes de todo debe tomar consciencia de que su producto por excelencia son los datos lexicográficos y que estos pueden usarse para un sinfín de aplicaciones que van desde su presentación en diccionarios hasta su integración en diferentes tipos de herramientas, asistentes, plataformas, servicios, etc."

e ferramentas lexicais para consumo humano quanto para serem usadas em processamento automático de linguagem natural (PLN). Em se tratando, nesta pesquisa, da produção de um dicionário eletrônico em regime de acesso aberto acessível pela Web, a definição do formato a ser utilizado envolve decisões sobre dois aspectos: o formato da base de dados e o formato da interface gráfica de consulta.

Trata-se, no primeiro caso, de um recurso destinado ao processamento por máquina, que possibilite a extração e compilação de dados de modo automatizado. Para garantir essa interação entre a base de dados e suas eventuais aplicações, adotam-se alguns princípios da Web Semântica (Berners-Lee; Lassila, 2001) e do chamado paradigma *Linked Open Data* (LOD) (Berners-Lee, 2006), ambos propostos com o objetivo de promover a interatividade entre a informação produzida para consumo humano e a produzida para processamento por máquina. São eles:

- estar disponível na Web sob licença aberta (e.g. Creative Commons);
- ser estruturado e legível por máquina (e.g. uma planilha, não uma imagem ou fotografia);
- não ser um formato patenteado (e.g. texto separado por caractere, e não um arquivo Microsoft Word ou Excel);
- identificar os dados de acordo com uma ontologia, isto é, um vocabulário especializado para representação do conhecimento (e.g. OntoLex-lemon (Cimiano; McCrae; Buitelaar, 2016)).

Uma opção que atende a esses requisitos é estruturar a base de dados em torno de planilhas de texto separado por tabulação (armazenadas em arquivos de extensão TSV), cujas colunas são nomeadas de acordo com modelos de ontologia especializada na representação de conhecimento lexical.

O formato da interface gráfica, por sua vez, diz respeito ao modo e à forma como os valores extraídos da base de dados são acessados e convertidos em informação visual (isto é, para consumo humano) através da Web. Em outras palavras, trata-se de dotar o dicionário de uma estrutura de acesso, no sentido de "um conjunto estruturado de indicadores lexicográficos que direcionam o usuário para a informação desejada" (Bergenholtz; Tarp; Duvå, 1995, p.219). Essa estrutura é formada por dois tipos de indicadores: os de acesso externo auxiliam o usuário a encontrar determinado verbete; os de acesso interno auxiliam o usuário a encontrar categorias de informação específicas no interior do verbete selecionado.

Nos dicionários impressos, a estrutura de acesso externo consiste basicamente na ordenação das entradas (alfabética ou temática), auxiliada pelo destaque tipográfico aplicado sobre os lemas (mais comum) e pela presença de índices suplementares (menos comum). A estrutura de acesso interna, por sua vez, se baseia fundamentalmente em recursos tipográficos. São exemplos: o uso de caracteres delimitadores de seção (numeração nas acepções, símbolos especiais apontando para um exemplo lexicográfico ou demarcando uma seção dedicada à fraseologia); e

o uso de formatação especial para distinguir categorias de informação colaterais ao significado (classes gramaticais, marcas de uso, padrão sintático, etc.).

Dicionários eletrônicos permitem combinar soluções tradicionais, baseadas em tipografia, com inovações exclusivas do ambiente digital. A primeira e mais evidente é a lista das palavras-entrada ou lemas, isoladas dos seus respectivos verbetes, que serve como principal estrutura de acesso externo; para acessar o verbete que procura, o usuário percorre a lista, geralmente em ordem alfabética, e clica no lema respectivo. Essa lista pode ser combinada com um campo de busca textual, em que o usuário digita a palavra (ou apenas uma parte dela) e encontra o verbete respectivo; caso o campo de busca seja direcionado para o conteúdo dos verbetes, é possível procurar ocorrências de determinada forma em todos os exemplos do dicionário e, com isso, obter uma visão ampliada do item lexical em relação ao conjunto do léxico. A organização interna dos verbetes se beneficia da ausência de restrições quanto ao espaço gráfico disponível, que permite dispor o conteúdo dos verbetes de modo extensivo, sem ter de recorrer aos diversos mecanismos de condensação (Wiegand, 1996) comuns à tradição impressa (em especial, abreviaturas e repetidores). Assim, diferentes categorias de informação podem não apenas receber formatação específica, mas também ocupar espaços diferentes na tela, ganhando em clareza e potencial de sua identificação pelo usuário.

Por envolver a conversão de formatos (isto é, da estrutura da base de dados para a visualização como página Web), a compilação da interface de consulta requer, sobretudo, consistência. Um modo de garanti-la é utilizar um programa de escrita de dicionários (em inglês, *Dictionary Writing Systems* ou DWS). Nesta pesquisa, adotamos para este fim o programa desenvolvido em código aberto *lexicographR* (Lugli, 2024b),[14] que oferece duas possibilidades de formato de publicação:

a) QuartoBook, que se apresenta como página estática, acessível a qualquer dispositivo conectado à Web e dotado de aplicativo navegador, com possibilidade de hospedagem gratuita através de sites que oferecem este serviço (e.g. GitHub Pages, Google Sites, WordPress, entre outros); e

b) ShinyApp, que consiste em um aplicativo para Web com características dinâmicas, capaz de proporcionar ao usuário maior interatividade na seleção e filtragem dos dados, mas com a desvantagem de depender de um servidor atualmente pago.

O objetivo de manter o custo material de produção próximo de zero[15] faz com que o formato preferível de publicação do dicionário seja o QuartoBook; não obstante, um experimento em

[14] Entre as razões para a escolha deste programa estão: (1) o fato de ser escrito em código aberto e em regime de reuso gratuito, ao contrário de DWS mais conhecidos como o ecossistema *Sketch Engine*; e (2) o fato de o autor desta pesquisa ter participado de sua concepção enquanto colaborador (*user tester*) do projeto *Democratizing Digital Lexicography*, coordenado por Ligeia Lugli e realizado em 2023 e 2024 (https://doi.org/10.6084/m9.figshare.c.7207656), com financiamento do National Endowment for the Humanities (HAA-290402-23).

[15] Vale notar que o custo de produção nunca será zero, uma vez que sempre haverá custos relacionados aos recursos humanos – por enquanto, restrito ao pesquisador autor deste trabalho – e ao consumo energético dos computadores utilizados no processamento de dados e publicação das ferramentas de consulta.

formato ShinyApp foi realizado com fins exploratórios. Os detalhes técnicos sobre os formatos definidos nesta seção serão apresentados no Capítulo 4.

2.1.3 Fontes de evidência

Wiegand (1984, p.14) define a base do dicionário como "o conjunto de materiais linguísticos que formam a base empírica para a produção de um dicionário de língua". Para compor essa base, o autor separa as fontes em dois tipos, de acordo com um critério simples: fontes *primárias* são aquelas que não consistem em um dicionário de língua (e.g. um *corpus* de textos); fontes *secundárias* são todos os dicionários de língua consultados.

Atkins (2008, p.38) parece contemplar as fontes primárias quando divide os tipos de evidências em três grupos, segundo o grau de objetividade do método de obter informações. No polo subjetivo estão as evidências obtidas por introspecção, limitadas ao conhecimento individual do próprio lexicógrafo. Em um ponto intermediário da escala estão as evidências obtidas por meio de investigação, realizada a partir de um arquivo de citações coletadas manualmente pelo editor ou por uma equipe de assistentes, ou então com base em pesquisas com informantes; tais técnicas, porém, se tornaram ultrapassadas desde que a maior parte dos textos é encontrada em formato digital (Atkins; Rundell, 2008, p.53). No polo objetivo estão as evidências extraídas da consulta a um *corpus* eletrônico, isto é, um conjunto selecionado de textos que provê ao lexicógrafo uma porção representativa do uso linguístico real (ainda que não totalmente objetiva, uma vez que os critérios de seleção de textos podem enviesar os dados).

Quanto às fontes secundárias, não mencionadas por Atkins (2008)[16], outros autores validam o uso de informação extraída de dicionários impressos na composição da base de dados. Biderman (2001, p.178), por exemplo, defende a construção de bases digitais de conhecimento lexical a partir de dicionários tradicionais, desde que reformulados segundo critérios da moderna ciência lexicográfica, com base no entendimento do que constitui a competência lexical de um falante nativo (por exemplo, o conhecimento das propriedades e do estatuto sintático-semântico dos verbos). A autora adverte que uma mera transposição do dicionário tradicional para a forma digital é insuficiente por duas razões: (a) porque dicionários tradicionais foram elaborados em torno de conhecimentos implícitos da estruturação do léxico, que devem ser explicitados na edição eletrônica para que se tornem legíveis pela máquina e passíveis de processamento automático; (b) porque, citando Coseriu, "os grandes monumentos lexicográficos não deixam de ser, em sua maior parte, simples repertórios em que as palavras se consideram como entidades

[16] Tal omissão se deve provavelmente ao ambiente competitivo próprio ao contexto comercial, muito embora a dependência dos novos dicionários em relação a seus predecessores seja maior do que se costuma admitir. Em contexto acadêmico, ao contrário, referenciar a informação é um requisito básico da produção de conhecimento e, portanto, logo reconhecer e esclarecer o grau de dependência dos materiais disponíveis é fundamental para o avanço científico, como afirma Hartmann (2001, p.39, tradução própria): "a maioria dos dicionários têm seus precursores e todos têm seus imitadores: a compreensão dos fundamentos históricos da elaboração de dicionários é, portanto, um dos pré-requisitos para o progresso da lexicografia acadêmica". No original: "Most dictionaries have forerunners, and all have imitators; an understanding of the historical foundations of dictionary-making is therefore one of the preconditions of further progress in academic lexicography."

isoladas, e não como elementos de um sistema, organicamente opostos e associados". Hanks (2022), por sua vez, também aponta o descompasso entre o avanço tecnológico e a renovação do conteúdo no âmbito da lexicografia computacional ao constatar que dicionários eletrônicos normalmente oferecem "pouco mais do que uma ligeira adaptação do texto dicionarístico tradicional" (p.473, tradução própria). O autor menciona, entre os problemas de se usar diretamente o conteúdo de dicionários tradicionais em atividades computacionais, o fato de que seres humanos (não importa se aprendizes, tradutores ou pesquisadores) são tolerantes a variações na forma de expressar a mesma informação, ao passo que a máquina exige uma maior consistência na apresentação dos dados.

Em se tratando de lexicografia do latim, a decisão por um ou mais tipos de fontes de informação para a base de dados está atrelada, antes de tudo, à sua condição de língua histórica.

No caso das fontes primárias, métodos subjetivos são inviáveis devido à inexistência de falantes nativos que pudessem descrever as propriedades de um item lexical com base em conhecimento internalizado, obtido por aquisição. Por definição, todo e qualquer conhecimento atual sobre o léxico latino se desenvolve em contextos de aprendizado, geralmente com base em descrições anteriores (dicionários, gramáticas, métodos de ensino) e na prática com os textos.

O uso de *corpora* textuais como fontes de evidência é mais promissor, embora não isento de dificuldades. A principal delas é a necessidade de contar com um *corpus* linguisticamente anotado para que a extração de dados possa ser eficaz. Para além da lematização – que pode ser considerada um requisito mínimo, dada a complexidade e variedade das formas flexionadas que constituem a Língua Latina –, anotações de nível morfológico, sintático e semântico são importantes para se extrair informações a respeito do comportamento da palavra em contexto. Já a informação contrastiva, elemento constitutivo do dicionário bilíngue, depende da existência de *corpora* paralelos alinhados no nível da palavra.

Um *corpus* de textos latinos anotados com esse grau de completude não parece existir. Há, sim, os resultados de diversas iniciativas independentes de transformar o patrimônio textual latino em recursos linguísticos computacionais, que produziram *corpora* linguisticamente anotados de dimensões e níveis de anotação diversos, mas ainda insuficientes para uma renovação radical da prática lexicográfica, especialmente a lexicografia bilíngue latim-português. A comparação da quantidade de textos latinos linguisticamente anotados disponível na Web com o total de textos digitalizados fundamenta essa afirmação. Tomando como base o projeto *Corpus Corporum* (Roelli, 2014),[17] que agrega recursos textuais lematizados produzidos por quinze projetos diferentes, o universo de textos latinos não anotados soma aproximadamente 160 milhões de *tokens*, dos quais menos de 500 mil possuem anotação sintática; destes, apenas 140 mil são de textos de autores do período clássico, isto é, datados entre os séculos II a.C. e II d.C. Quanto à informação contrastiva, essencial em projetos de dicionário bilíngue, a quantidade de *tokens* alinhados com tradução para o português é quase irrisória, contabilizando aproximadamente 7

[17] Cf. <https://mlat.uzh.ch/home>.

mil palavras na plataforma de alinhamento *Ugarit*).[18]

As dificuldades de se obter informação de fontes primárias enfatizam o valor do patrimônio lexicográfico construído ao longo de vinte séculos de estudo ininterrupto da língua latina, dos repertórios lexicais produzidos na Antiguidade Tardia e na Alta Idade Média aos léxicos computacionais do século XXI, passando, é claro, pelos diversos dicionários impressos entre os séculos XV e XX. Neste vasto *corpus* de fontes secundárias, importam especialmente para esta pesquisa os dicionários bilíngues latim-português e vice-versa, cada um representando, a seu modo, uma tentativa de oferecer ao público lusófono informação sobre (e, principalmente, equivalentes para) o significado da expressão latina.

Sem fechar os olhos para as inadequações detectadas pela crítica (Lima, 1995; Longo, 2006), não há como ignorar ou descreditar *a priori* esse conhecimento lexical herdado. O que é preciso é estimar seu valor relativo pelo confronto produtivo entre si e com outras fontes de conhecimento. Para Humblé (2001, p.52–3), viabilizar esse confronto é uma das principais inovações possibilitadas pelo dicionário em formato eletrônico:

> Um novo dicionário eletrônico deveria tirar proveito das capacidades quase infinitas de armazenamento e lançar mão de todos os diferentes tipos de conhecimento lexicográfico e enciclopédico acumulados ao longo dos séculos. De uma perspectiva eletrônica, diversas fontes de informação podem ser disponibilizadas ao mesmo tempo.[19]

Para além do patrimônio lexicográfico, uma base de dados pode agregar igualmente dados extraídos de fontes primárias[20], permitindo à ferramenta de consulta disponibilizar ordenadamente em um único espaço informações oriundas de fontes diversas inclusive quanto à sua natureza. Com isso, o usuário tem a oportunidade de ele próprio avaliar a pertinência de cada uma ao contexto do problema que motivou sua consulta ao dicionário e fundamentar sua decisão. Nesse sentido, Varantola (1994) ressalta os benefícios, para o usuário de dicionários, de contar com dados extraídos de *corpus* em complemento à informação lexicográfica tradicional, argumentando que a informação altamente contextualizada que estes dados fornecem pode auxiliar o usuário a se certificar de que sua escolha interpretativa é correta e aceitável em determinado contexto particular.

2.2 MACROESTRUTURA

Em lexicografia, o termo macroestrutura se refere ao conjunto organizado de entradas em uma obra de referência, que permite a localização da informação pelo compilador e pelo usuário

[18] Cf. <https://ugarit.ialigner.com/>. Mais detalhes sobre os corpora de textos latinos disponíveis e a compilação de um corpus paralelo para fins lexicográficos são apresentados no Capítulo 3.

[19] No original: "A new electronic dictionary should take advantage of the almost unlimited capacities of electronic storage and put to use all the different kinds of lexicographic and encyclopaedic knowledge accumulated over the centuries. From an electronic point of view, several sources of information can be available at the same time."

[20] A ideia de integrar em uma mesma uma base lexicográfica dados extraídos de *corpus* e dados extraídos de dicionários tradicionais se encontra em Teubert (1984) e Kruyt (1996).

(Hartmann, 2001; Hartmann; James, 2002).

Essa definição contém dois aspectos que requerem decisões por parte do lexicógrafo no planejamento de um dicionário (Atkins, 2008, p.39): de um lado, a elaboração da lista de itens lexicais que serão objeto de descrição (também chamada de nomenclatura), baseada em critérios de seleção previamente definidos; de outro, o arranjo desses itens, incluindo a forma de ordená-los (por ordem alfabética, temática, frequencial, etc.), o *status* de cada item no conjunto (como entrada principal ou como subentrada de outro item) e os modos de lidar com fenômenos de homonímia e polissemia.

Tais escolhas devem responder ao roteiro de consulta atribuído ao usuário pelo modelo de Bogaards (1996), tendo particularmente em vista o segundo e terceiro passos, isto é, a identificação da palavra que causa problemas de compreensão e a determinação de sua forma canônica, pela qual o usuário localizará a informação no dicionário.

2.2.1 Itens lexicais

Item lexical pode ser "qualquer palavra, abreviação, elemento de composição ou expressão passível de figurar em um dicionário como objeto de algum tipo de descrição lexicográfica (normalmente, sua definição ou tradução)" (Atkins; Rundell, 2008, p.163, tradução própria).[21]. Essa definição abrangente inclui itens de diversas dimensões, quais sejam:

- palavras simples de todas as classes, sejam estas de natureza lexical (substantivos, adjetivos, verbos e advérbios) ou gramatical (preposições, conjunções, pronomes, auxiliares e determinantes).
- abreviaturas e contrações, como alfabetismos (e.g. CNPq), acrônimos (e.g. CAPES) e fusões de palavras (como as que ocorrem em português entre preposição e artigo, e.g. *das*, *no*).
- elementos de composição de palavras, especialmente afixos (como IM– em *impossível* ou –MENTO em *melhoramento*).
- expressões multipalavra, em que a co-ocorrência de determinadas formas constitui uma unidade com significado específico, não transparente (i.e. não correspondente à soma dos significados das unidades), como é o caso de certas colocações, substantivos compostos, frases fixas e semi-fixas, provérbios e citações.

Note-se que palavra é um conceito-chave nessa discussão, na medida em que os demais itens são definidos de forma relativa, seja como componentes de palavras (no caso dos afixos), seja como compostos de palavras (no caso das abreviaturas, contrações e expressões multipalavra). Não obstante a conhecida dificuldade de se definir palavra em termos linguísticos,[22] o fato

[21] No original: "A 'lexical item' is any word, abbreviation, partial word, or phrase which can figure in a dictionary (often as the headword of an entry) as the 'target' of some form of lexicographic description, most commonly a definition or a translation".

[22] Cf. a discussão em Biderman (2001, p.109ss) e Welker (2004, p.17ss).

de o latim sobreviver apenas como língua escrita e, especialmente, o pressuposto de que o dicionário de latim será usado principalmente em atividades que envolvem compreensão de textos escritos justifica a adoção de uma definição baseada em critérios gráficos de segmentação, aliás utilizada tanto em semântica quanto em linguística computacional. Assim, para nossos propósitos, palavra é "sequência de letras que, na prática tipográfica normal, é limitada de ambos os lados por um espaço" (Lyons, 1980, p.24).[23]

Igualmente necessária é a distinção ulterior de três entidades a que o termo palavra pode se referir. Em sentido concreto, palavra pode designar individualmente cada ocorrência (*token*) de um grupo de letras delimitado por espaços em branco em um texto, assim como pode designar em conjunto cada tipo (*type*), isto é, cada uma das possibilidades de combinação de letras que ocorrem em um texto.[24] Em sentido abstrato, palavra pode se referir ao lexema, que consiste na compreensão unitária de um item lexical independente de suas diversas possibilidades de realização morfológica (e.g. dizer que *cantarei* e *cantando* são a mesma palavra *cantar*). Neste trabalho de pesquisa, dada sua natureza lexicográfica, utiliza-se palavra normalmente como sinônimo de lexema; seus tipos (as diversas realizações morfológicas) são designados como formas da palavra; e suas ocorrências, como ocorrências ou *tokens*.

A classificação das palavras pode ser feita com base em critérios semânticos ou formais.

Do ponto de vista do significado,[25] o léxico pode ser dividido em dois grandes conjuntos: o das palavras lexicais e o das palavras gramaticais ou funcionais (Borba, 2003, p.46). No primeiro grupo, mais extenso, estão os itens que relacionam o sistema de noções da língua com o mundo exterior, capazes de representar realidades extralinguísticas como substâncias (*criança*, *sujeira*), atributos (*infantil*, *sujo*), ações (*brincar*, *sujar*), estados (*doente*, *ter*), processos (*confecção*, *descansar*); é um conjunto aberto e sempre vulnerável a influências externas. O segundo grupo, mais restrito, compreende as palavras que estabelecem as relações intra e interdiscursivas, por traduzirem "um conjunto de conceitos abstratos […] que só encontram representação no sistema gramatical", como quantificação, intensificação, referenciação, identificação, modalização, além das relações espaciais, temporais, entre outras.

Grosso modo, isto é, desconsiderando as diversas superposições e outros traços definitórios,[26] esses traços semânticos podem ser usados para subdividir ambos os conjuntos em classes de palavras. Assim, as classes lexicais compreendem os Nomes, que designam substâncias (e.g. seres, objetos, sentimentos, qualidades, fenômenos); os Verbos, que expressam ações, processos,

[23] O uso dessa definição em linguística computacional é atestado em Ježek e Sprugnoli (2023, p.37).

[24] Segundo Lyons (1980, p.20), "a relação entre ocorrências e tipos será referida como uma relação de instanciação; diremos que as ocorrências actualizam o seu tipo".

[25] O termo significado é aqui entendido como valor individual potencial e paradigmático que todo item léxico tem à disposição do falante, ocupando um lugar no sistema delimitado pelos demais significados pertencentes ao mesmo conjunto, por contraste (Borba, 2003, p.19); em oposição, os termos significação ou sentido designam um valor sintático-semântico resultante da combinação de unidades.

[26] Pode-se citar, por exemplo, o uso de verbos suporte em formas analíticas para indicar noções aspectuais (e.g. iteratividade em "tem trabalhado", progressividade em "ir andando") ou modais (e.g. possibilidade em "pode acabar").

estados relacionados com os nomes; os Adjetivos, que indicam atributos dos nomes; e os Advérbios, que indicam circunstâncias relacionadas aos verbos (e.g. lugar, tempo, modo). As classes gramaticais, por sua vez, compreendem os Pronome (e.g. usados para indicar referenciação), os Determinantes (e.g. usados para indicar identificação), os Possessivos (e.g. usados para indicar pertencimento), os Numerais (e.g. usados para indicar quantificação), as Preposições e as Conjunções (e.g. relações espaciais e temporais).

Todavia, a relação entre forma e traço semântico muitas vezes é imprecisa. O léxico de uma língua é marcado por ambiguidades que podem ser contrastivas, quando uma forma lexical tem acidentalmente dois ou mais significados distintos e não relacionados entre si (fenômeno chamado de homonímia, e.g. em português, *canto* 'produção melódica de sons' e *canto* 'ângulo formado por duas paredes'), ou complementares, quando uma forma lexical possui diferentes significados ou acepções que correspondem a manifestações de um mesmo significado básico em diferentes contextos (fenômeno chamado polissemia, e.g. a *asa* de um pássaro, de um avião, de uma xícara, etc.).[27] Ainda que possa ser visto como um fenômeno acidental, parecem existir esquemas regulares de polissemia, isto é, alternâncias sistemáticas de significado entre classes inteiras de palavras a partir de processos metonímicos e metafóricos (Ježek; Sprugnoli, 2023, p.43). Podem-se citar, como exemplos do primeiro tipo, nomes de autores que passam a significar suas obras (e.g. "Cícero foi um orador romano" *vs.* "passei a manhã traduzindo Cícero"), nomes de frutos que designam as respectivas plantas (e.g. "espremer um limão" *vs.* "plantar limão no quintal"), ou ainda a chamada polissemia inerente[28], quando, em virtude de uma complexidade ontológica do objeto denotado, seus diversos aspectos podem ser ativados simultaneamente em uma expressão (e.g. em "o livro ali na estante é chato", a palavra *livro* denota ao mesmo tempo um objeto físico posicionado em determinado local e um conteúdo que provoca certo estado mental no leitor). No segundo tipo estão incluídos os usos figurados, baseados no processo cognitivo da semelhança (e.g. "chover protestos", em que *chover* é reinterpretado por uma semelhança entre a quantidade de gotas de chuva e a quantidade de vozes em protesto).

Do ponto de vista da forma, as palavras podem ser classificadas com base no tipo de morfema que participa de sua formação. [29] Morfemas são as unidades mínimas significativas de uma língua, cujo significado é incompleto e abstrato (Biderman, 2001, p.125); podem ser de três tipos: lexicais, flexivos e derivacionais. Morfemas lexicais estão presentes em todas as palavras, pois constituem a base (ou raiz) de seu significado e permitem organizar o léxico em famílias (e.g. o morfema CANT- define a família de *cantar*, *cantor*, *cantoria*, *cantiga*, etc.). Morfemas flexivos portam significação gramatical, particularmente as categorias de número, gênero, caso, modo, tempo, pessoa (e.g. *cant<u>ar</u>*, *cant<u>ei</u>*, *cant<u>ávamos</u>*, *cantigas<u></u>*). Morfemas derivacionais fazem inci-

[27] Do ponto de vista psicolinguístico, considera-se que palavras homônimas constituem itens diferentes no léxico mental (Ježek; Sprugnoli, 2023, p.42).

[28] O termo é de Pustejovksy – cf. Ježek e Sprugnoli (2023, p.43).

[29] Seguimos a exposição de Ježek e Sprugnoli (2023, p.37–61), que utiliza palavra no sentido exposto acima. Outros autores usam o termo lexia quando apresentam essa classificação – cf. Welker (2004), Biderman (2001), Borba (2003).

dir sobre a base lexical significados abstratos como a possibilidade (*variável*, 'que pode variar'), oposição (*desrespeito*, 'que não tem respeito'), negação (*impossível*, 'que não é possível'), relação ou movimento espacial (*sobrepôr*, 'pôr por cima'; *importar*, 'trazer para dentro'), entre outros.[30] Assim, palavras podem ser:

- variáveis ou invariáveis, conforme a presença ou ausência, respectivamente, de morfemas flexivos em sua constituição;
- simples ou complexas, conforme a ausência ou a presença, respectivamente, de morfemas derivativos e/ou mais de um morfema lexical em sua constituição.

Palavras simples são constituídas de um único morfema lexical; no caso das palavras variáveis, suas diversas formas apresentam também um ou mais morfemas flexivos (e.g. *cantaram*, *cantos*).

Palavras complexas podem ser de tipo morfológico ou sintático. No primeiro grupo entram as palavras derivadas, constituídas de um morfema lexical unido a um morfema derivacional (e eventuais morfemas flexivos, e.g. *cantor*, *cantáveis*), e as palavras compostas, formadas por dois ou mais morfemas lexicais (e eventuais morfemas derivacionais ou flexivos, e.g. *aguardente*, *pernilongo*, *saca-rolhas*). O segundo grupo é o das expressões multipalavra (em inglês, *multiword expression* ou MWE), termo usado em linguística computacional para se referir aos fenômenos linguísticos que ultrapassam a fronteira das palavras mas não constituem sintagmas livres, isto é, são "sequências de palavras que possuem uma coesão interna típica das palavras simples, que bloqueia modificações lexicais e sintáticas" (Ježek; Sprugnoli, 2023, p.39). Combinações como *mesa redonda*, *sala de espera*, *jogar fora* não podem sofrer substituições ou inserções sem que o significado seja alterado (e.g. "mesinha redonda", "quarto de espera", "sala suja de espera", "jogar lá fora"). Expressões multipalavra podem ter várias dimensões, indo do sintagma nominal (e.g. *faca de dois gumes*) ao predicado verbal (e.g. *chover no molhado*) e daí ao enunciado completo (e.g. *quem com ferro fere, com ferro será ferido*).[31]

2.2.2 Seleção da nomenclatura

A determinação dos tipos de itens que hão de figurar como entradas no dicionário deve ser compatível com o roteiro de consulta do usuário, a partir dos critérios que ele presumidamente adotará tanto para a segmentação do texto e quanto para a lematização das formas encontradas.

Assim, tendo-se definido a recepção escrita como situação principal de uso de dicionário de latim, parte-se do princípio de que o usuário buscará ali informações a respeito de objetos linguísticos delimitados segundo o critério gráfico – entre espaços em branco ou sinais de pontu-

[30] Sobre os morfemas derivacionais com interesse lexicográfico, ver Jackson (2002, p.8).

[31] Uma descrição mais detalhada será apresentada adiante, quando se tratar das informações sintagmáticas que dicionários podem incluir na microestrutura.

ação –, conforme a segmentação do texto oferecida pela edição utilizada.[32] Por tal critério, ficam excluídas da nomenclatura as unidades menores do que a palavra, como é o caso dos afixos e outros elementos de composição. Mesmo que haja alguma validade em o usuário tentar depreender o significado de uma palavra a partir de seus elementos formadores tomados isoladamente, esse comportamento não deve ser atribuído a todo usuário, que pode ou não ter o conhecimento necessário ou não confiar em tal método dedutivo, já que, em se tratando de linguagem, o significado do todo pode não corresponder exatamente à soma das partes. Informações sobre a formação das palavras costumam ser fornecidas pelas gramáticas, o que não impede, vale dizer, que o dicionário inclua uma lista de elementos formadores (tanto afixos quanto bases lexicais) em forma de apêndice.

O mesmo critério leva igualmente à exclusão da nomenclatura – mas não do dicionário – das unidades maiores que a palavra (as MWE), pelo motivo inverso. Partindo do pressuposto de que o usuário de dicionário bilíngue procura respostas para fatos linguísticos desconhecidos, pode-se presumir que o desconhecimento prévio da existência de uma coesão especial entre as palavras de determinadas expressões induza o usuário à busca pelo significado de cada item individualmente, conforme a segmentação oferecida pelo texto. Logo, indexar expressões multipalavra na nomenclatura parece pouco eficiente enquanto estrutura de acesso à informação, o que não significa isentar o dicionário de fornecer informações a respeito desses itens. Ao contrário, o formato digital permite criar uma estrutura mais adequada, que consiste em vincular cada unidade fraseológica às entradas de todas as palavras que participam de sua composição, de modo que o usuário possa encontrá-las consultando a entrada relativa a qualquer de seus elementos constituintes ou, alternativamente, utilizando um campo de busca textual.[33]

Passam, portanto, pelo critério gráfico: palavras simples de formação primitiva ou derivada, bem como abreviaturas e contrações. O critério da lematização permite excluir da nomenclatura as duas últimas.

Em linguística computacional, lematização é o processo de anotação do lema das palavras de um dado texto (Ježek; Sprugnoli, 2023). Lema é o termo usado em lexicografia e em linguística computacional para se referir à forma canônica de citação de uma palavra, convencionalmente utilizada em contextos metalinguísticos para representá-la como uma unidade abstrata, independente de suas realizações morfológicas concretas.[34] Nesse sentido, corresponde ao terceiro passo do roteiro de consulta de Bogaards (1996), donde se presume que o uso de dicionários requer um grau mínimo de conhecimento linguístico e lexicográfico – para saber analisar morfologicamente a forma ocorrente no texto e para deduzir dela a forma canônica que consta da nomenclatura, respectivamente. Esse conhecimento é especialmente importante no caso de

[32] Vale lembrar que a segmentação do texto latino é intervenção editorial moderna, podendo apresentar diferentes possibilidades de leitura a partir de um manuscrito que, via de regra, é caracterizado pela ausência de segmentação no nível das palavras, geralmente apresentando ligaduras contínuas entre os caracteres.

[33] Detalhes técnicos de como esse vínculo é estabelecido são apresentados no Capítulo 3.

[34] Nesse sentido, não se confunde com o lexema, definido anteriormente, já que a forma de citação constitui uma de suas formas, eleita por convenção para representá-lo, não o lexema em si.

uma língua morfologicamente rica como o latim, em que as possibilidades formais das palavras variáveis podem ser extremamente numerosas.

Tradicionalmente, a regra para se determinar a forma canônica das palavras latinas varia conforme a classe:

- para substantivos, usa-se a forma do nominativo singular (exceção feita a certas palavras lematizadas pelo nominativo plural, seja por não ocorrerem no singular, seja por terem sentidos específicos em cada forma);
- para adjetivos e determinantes, usa-se a forma do nominativo masculino singular (com exceção dos *pluralia tantum*, igualmente mencionados para os substantivos);
- para verbos, utiliza-se a primeira pessoa singular do presente do indicativo (com exceção de verbos impessoais, lematizados pela terceira pessoa do mesmo tempo e modo, e dos verbos defectivos, que só possuem formas de pretérito perfeito, nesse caso mantendo-se a pessoa e o modo).

No **eLatim**, tal sistema é mantido por duas razões: (1) pelo fato de ser um padrão convencionado, segundo o qual a maior parte dos usuários desenvolveu suas habilidades lexicográficas; (2) pelo fato de ser amplamente utilizado em linguística computacional tanto na elaboração de léxicos quanto na anotação de *corpora* de textos.

A complexidade morfológica do latim pode dificultar o processo de dedução do lema a partir de uma forma flexionada, especialmente em casos de alternância de radicais para indicar aspecto verbal, em que a relação entre forma flexionada e lema é menos transparente. Por isso, muitas vezes a nomenclatura de dicionários de latim inclui formas flexionadas para indicar o radical do *perfectum* de alguns verbos (e.g. *dixi*, perfeito de *dico*; *peperi*, perfeito de *pario*), bem como o radical do *supino*, utilizado em diversas formas nominais derivadas (e.g. *satus*, particípio de *sero*; *disertus*, particípio de *dissero*). Esse recurso, certamente útil em um dicionário impresso, torna-se desnecessário em um dicionário em formato eletrônico, dada a possibilidade de se utilizar um campo de busca textual como estrutura de acesso externo. Com isso, estas e quaisquer outras formas flexionadas podem ser localizados no interior do verbete, em seção específica da microestrutura.

O mesmo recurso pode ser usado para abreviaturas e contrações entre palavras, por também constituírem formas alternativas de representação escrita. No primeiro caso, sua ocorrência nos textos se dá na forma de um segmento isolado por espaços como uma palavra, eventualmente marcado graficamente (via de regra, por um espaço-branco à esquerda e um ponto-final à direita). Nas edições de textos latinos as mais frequentes são formas abreviadas de nomes próprios (e.g. *M.* para *Marcus*, *D.* para *Decimus*, *C.* para *Gaius*, etc.), numerais (e.g. *I* para *unus*, *X* para *decem*, *L* para *quinquaginta*), símbolos financeiros (*HS* por *sestertius*) e expressões de data (e.g. "A. d. III Kal. Mai." por "ante diem tertius Kalendas Maias", correspondente a 29 de abril no calendário atual) e fórmulas de saudação ("S.P.D." por "salutem plurimam dicit", algo como

'saúda efusivamente'), sobretudo nas correspondências. No segundo caso, trata-se de formas isoladas, em número limitado. Algumas têm função claramente formular, como *sis* (contração de *si uis*, literalmente 'se queres'), usada com função semelhante a "por favor" em português; ou então *satin* (contração do advérbio *satis* com a partícula enclítica interrogativa *-ne*), usada para introduzir perguntas retóricas do tipo: "não basta…[isso ter acontecido]?". Outras representam elisões fonéticas que remetem à linguagem falada, como *meumst*, *factumst*, *molestumst* (formas contratas de *meum est* 'é meu', *factum est* 'está feito' e *molestum est* 'é um incômodo', respectivamente). Tanto formas abreviadas quanto formas contratas podem figurar no interior do verbete, em seção destinada às formas alternativas do lema, representado na nomenclatura por sua forma completa. O acesso às formas alternativas pode ser feito por busca textual.[35]

Com isso, a nomenclatura do **eLatim** passa a ser constituída, do ponto de vista da forma, por lemas de palavras (simples, derivadas e compostas por aglutinação). Por se tratar de um dicionário geral, não se aplicam critérios de exclusão ulteriores, sejam eles baseados em propriedades semânticas especiais (e.g. nomes coletivos, nomes próprios de pessoas ou lugares) ou em contextos marcados de uso (e.g. arcaísmos, regionalismos, estrangeirismos, gírias, jargões, insultos, palavras tabus). Considerando o pressuposto de que o usuário busca no dicionário informação sobre uma palavra que ele desconhece, uma lista separada de nomes próprios ou qualquer outro item seria contraproducente.

2.2.3 Arranjo das entradas

Para cumprir o papel de estrutura de acesso externo – isto é, de entrada para a descrição lexicográfica –, os itens que constituem a nomenclatura devem obedecer a um arranjo sistemático que oriente o usuário a encontrar o lema desejado. Três aspectos devem ser considerados: ordem, estatuto e mecanismos de diferenciação.

A ordem é o recurso mais utilizado para auxiliar o usuário a percorrer as entradas; em especial, a ordem alfabética pode ser considerada como que uma expectativa consensual quando se trata de dicionários gerais utilizados em atividades de leitura de textos.[36]

Quanto ao estatuto ou *status* de cada item no conjunto, Atkins (2008, p.40) explica que a estrutura das entradas pode ser plana (*flat*), quando cada item tem sua entrada própria, ou estratificada (*tiered*), em que certos itens são posicionados como subentradas (visualmente marcadas por recuo ou outro recurso gráfico) de uma entrada principal ou de outra subentrada, normalmente com base em critérios de formação (e.g. palavras derivadas como subentradas das palavras primitivas, típica de dicionários organizados por famílias). Em um dicionário latim-português como o **eLatim**, uma estrutura plana, com todos os itens lexicais da nomenclatura dispostos no nível de entrada principal, parece mais indicada principalmente por duas razões:

[35] Também é possível fornecer, em apêndice, uma lista de abreviaturas acompanhadas de suas formas por extenso e significado.

[36] Repertórios lexicais com finalidades diversas (e.g. aprendizagem de vocabulário) podem se valer de outras formas de organização (e.g. temática, frequencial, por famílias, etc.).

1. a indisponibilidade de dados suficientes para uma organização por famílias, dado que a maior parte das fontes de dados (dicionários impressos e léxicos eletrônicos) não fornece informações sistemáticas sobre a composição morfológica de todos os itens;

2. a necessidade de equiparar o dicionário com os recursos textuais (*corpora*) disponíveis, considerando que o processo de lematização computacional tende a neutralizar morfemas flexivos, mas não derivativos, produzindo uma lista de lemas individualizados para palavras simples e derivadas.

Por fim, resta considerar os mecanismos de diferenciação utilizados para lidar com palavras polissêmicas e, especialmente, lemas homógrafos. Nesse aspecto, Atkins (2008) elenca três possibilidades para organização da lista de entradas:

1. totalmente não homográfica, em que cada entrada é constituída por uma forma ortográfica exclusiva como lema, com todas as acepções relacionadas com essa forma constituindo uma única entrada polissêmica;

2. totalmente homográfica, em que cada entrada corresponde a uma unidade tanto de forma quanto de conteúdo, de modo que diversas entradas podem compartilhar uma mesma forma ortográfica do lema mas nenhuma pode ser polissêmica;

3. parcialmente homográfica, em que pode ocorrer tanto identidade de lemas para entradas diferentes quanto entradas eventualmente polissêmicas.

Nenhuma das três possibilidades é isenta de problemas.

Na primeira, pode-se citar como desvantagem o fato de impor uma unificação da descrição lexical para palavras com lemas homógrafos mas propriedades morfológicas e sintáticas evidentemente distintas (e.g. a forma *dico* como lema dos verbos *dicĕre* ‘dizer’ e *dicāre* ‘dedicar’), o que torna inviável uma descrição formal válida para ambas praticamente e inconsistente a consideração de suas diversas acepções como componentes de uma mesma polissemia.

Na segunda, a desvantagem se dá pelo motivo inverso, na medida em que criar uma entrada para cada significado distinto da palavra resultaria numa multiplicação de lemas não só mais difícil de gerenciar (tanto pelo compilador quanto pelo usuário), mas também capaz de causar a impressão de que cada entrada se refere a uma de uma palavra diferente, prejudicando a percepção da existência da polissemia pelo usuário e da necessidade de interpretação tendo em vista o contexto.

A terceira opção, menos radical, se apresenta como preferível ao permitir agrupar os diversos significados de uma palavra em entradas únicas e, ao mesmo tempo, distinguir palavras acidentalmente homográficas quanto a seus lemas, mas evidentemente distintas quanto às demais propriedades definidoras.[37] Nesse caso, a exigência é dispor de critérios objetivos que

[37] Reforça essa posição o fato de que, do ponto de vista psicolinguístico, palavras homônimas constituem itens diferentes no léxico mental, o que não ocorre com palavras polissêmicas (Ježek; Sprugnoli, 2023, p.42).

determinam quais sejam essas propriedades, bem como de um mecanismo para distinguir lemas homógrafos na nomenclatura. Para responder à primeira exigência, o **eLatim** se baseia nos critérios propostos por Passarotti et al. (2020): cada lema possui uma única base lexical (ou raiz), uma única classe de palavras e um único padrão flexional (paradigma morfológico). Para responder à segunda, utiliza-se o recurso, bastante usual em lexicografia, de acrescentar um algarismo após a forma do lema (e.g. *dico1*, *dico2*), atribuído com base na frequência do item na base de dados, em ordem decrescente.

2.3 MICROESTRUTURA

O quarto passo do modelo de Bogaards (1996) prevê que o usuário, após localizar na macroestrutura o lema que lhe interessa, busque no interior do verbete informação relevante para a resolução de seu problema linguístico. O grau de relevância da informação depende do tipo de problema a ser resolvido, mas sempre implica um confronto de contextos. Em se tratando de um problema de compreensão, o objetivo geral da consulta é saber sobre o significado da palavra. Como palavras frequentes possuem normalmente mais de um significado, o usuário necessita igualmente de informações contextuais para validar suas escolhas – para se certificar de que selecionou tanto a entrada correta quanto a acepção mais condizente com o problema.

Do ponto de vista da produção do dicionário, o sucesso dessa operação depende de duas decisões: (1) quanto aos tipos de informação que serão fornecidos; e (2) quanto ao modo de organizar os elementos no interior dos verbetes. Esses dois aspectos são os componentes da chamada microestrutura do dicionário (Hartmann, 2001, p.65).

Atkins e Rundell (2008, p.152–8) definem três grupos de informações consideradas relevantes para compor uma base de dados lexicais:

1. propriedades inerentes, como classe, formas, comportamento gramatical e tipo semântico (e.g. em inglês *argue* é um verbo, tem como formas *argues, argued, arguing*, costuma constituir núcleo de predicados e, apesar de polissêmico, é essencialmente um verbo de comunicação);

2. propriedades contextuais, extraídas da observação de concordâncias, que permitem distinguir diferentes sentidos para uma palavra com base nos padrões de combinação observados (sejam estes de complementação ou coocorrência de termos, e.g. "argue with X [umpire, father] about Y" remete a uma discussão com dois participantes e um tópico, "argue for" a uma motivação com um participante humano e um tema, "argue that Z [the law, the system]" a uma afirmação, e assim por diante); e

3. propriedades do texto-fonte em que se observa cada ocorrência, que incluem autor, título, gênero discursivo, registro, entre outras informações fornecidas pelos metadados do corpus que são relevantes para emoldurar cada uso da palavra a um contexto sócio-discursivo, bem como calcular sua maior ou menor dispersão.

Vale lembrar que muitas dessas informações podem dizer respeito a cada unidade lexical (*lexical unit*) vinculada ao lema, e não ao lema como um todo.[38].

Quanto à apresentação dessas informações para consulta, Hartmann (2001, p.65) divide o verbete em duas partes: comentário formal e comentário semântico.

O comentário formal inicia o verbete e traz uma descrição das características gerais sobre o item lexical representado pelo lema, normalmente válidas para todas as acepções, independente do grau de polissemia do lema, isto é, do número de unidades lexicais que ele comporta.[39] Welker (2004) identifica seis tipos de informação consideradas relevantes sobre a forma do lema: ortografia, pronúncia, classe gramatical, padrão flexional e etimologia.

O comentário semântico congrega as informações a respeito do(s) significado(s) das palavras, normalmente divididos em acepções. São consideradas como potencialmente relevantes na descrição do significado (Welker, 2004, p.108): informações explicativas (normalmente na forma de uma definição), informações paradigmáticas (grupos e categorias semânticas, sinônimos e antônimos, hiperônimos e hipônimos, etc.), informações sintagmáticas (padrões de combinação, colocações, exemplos), informações sobre variação e registro (marcas de uso), indicações de derivação semântica e ilustrações (desenhos, gráficos).

A discussão a seguir aborda os tipos de informação formal e semântica elencados, tendo em vista o contexto de uso de dicionários de latim.

2.3.1 Formas canônicas

Tanto em lexicografia quanto em linguística computacional, o lema das palavras variáveis costuma corresponder a uma de suas formas, tradicionalmente escolhida como sua representante canônica. Nos dicionários de latim, porém, a tradição é que a forma canônica da entrada seja composta pelo lema propriamente dito acompanhado de indicações (normalmente abreviadas) de uma ou mais formas flexionadas, cujo número varia dependendo da classe e do paradigma flexional a que a palavra pertence. Substantivos, por exemplo, são normalmente representados por duas formas casuais (nominativo e genitivo, e.g. *animus, -i* ou *res, rei*); adjetivos são representados por duas ou três formas de nominativo, conforme a distinção mais ou menos marcada de gênero gramatical (e.g. masculino, feminino e neutro em *magnus -a -um* ou masculino/feminino e neutro em *omnis -e*); a apresentação dos verbos varia, chegando a ser representados por até

[38] O termo é de Cruse (1986), que o define como sendo um "complexo de forma-significado dotado de propriedades semânticas (relativamente) estáveis e discretas" (p.49, tradução própria do original: "form–meaning complexes with (relatively) stable and discrete semantic properties") ou, mais simplesmente, como "a união de uma forma lexical e um único significado" (p.77, tradução própria do original: "the union of a lexical form and a single sense") –entendendo por forma lexical a unidade abstrata que representa o conjunto de formas flexionadas de uma palavra, isto é, o lema. Welker (2004, p.22) vincula o interesse desse recorte com o fato de cada acepção ter suas características sintagmáticas e paradigmáticas específicas.

[39] Há casos, todavia, em que a descrição do significado pode exigir determinações formais específicas ou restritas; um exemplo em latim é a palavra *aedis*, que no singular significa 'templo' e, no plural, 'casa'; nesse caso, pode-se dizer que o lema comporta duas unidades lexicais, para cuja descrição a oposição singular/plural é significativa.

cinco formas para contemplar a flexão verbal em todos os seus aspectos (e.g. as entradas *dico -is -xi -ctum -ere* ou *do das dedi datum dare* contêm formas da 1ª e 2ª pessoas do presente do indicativo, da 1ª pessoa do pretérito perfeito do indicativo, do supino e do infinitivo presente ativo, respectivamente).

A presença de diversas formas acompanhando o lema se justifica pelo fato de o latim ser uma língua extremamente rica do ponto de vista da morfologia, sendo o dicionário responsável por prover informação completa sobre as possibilidades de flexão de cada palavra, de modo a permitir a identificação e análise de qualquer forma encontrada em um texto (Celano, 2020, p.22–23). Com efeito, no caso de 'dico', por exemplo, o conhecimento dos padrões flexionais latinos permite depreender do conjunto de formas fornecido: a classe de palavra (verbo), o paradigma flexional (3ª conjugação) e as três bases morfológicas da flexão verbal (a saber, do *infectum*, do *perfectum* e das formas nominais do verbo).

Todavia, a forma tradicional de apresentar essa informação por meio de abreviações tem sido alvo de críticas por criar dificuldades para os usuários. Segundo Lima (1995, p.106), "nada há de mais impróprio à descrição da língua do que essas lacerações praticadas por dicionaristas no corpo dos nomes latinos em sua representação escrita", pois tal prática, segundo o autor, "deforma a imagem visual da palavra e não confere ao próprio tema o relevo que lhe é devido por sua força unificadora e de étimo".

Na mesma linha, Longo (2006) menciona, entre as barreiras que o usuário deve superar ao consultar dicionários bilíngues latim-português, a arbitrariedade e a inconsistência observadas no modo de apresentar as formas canônicas secundárias por meio de segmentos de palavras. Mesmo quando se considera apenas sua função primária, de remeter ao paradigma flexional conforme é descrito pelas gramáticas tradicionais, a falta de um padrão constante de segmentação pode levar o usuário a resultados incoerentes em sua tentativa de reconstituição das formas completas:

> O corte aplicado fora do lugar naturalmente esperado resulta em coisas do tipo *frigor, ōris*; *gurges, ĭtis*; *caput, ĭtis*, o que possivelmente levaria um iniciante à conclusão lógica de que o genitivo singular para cada um desses vocábulos fosse, **frigororis*, **gurgesitis* e **caputitis* ao invés de *frigoris*, *gurgitis* e *capitis*. (Longo, 2006, p.44)

Pode-se acrescentar como exemplo o lema *dico* mencionado acima: a segunda forma canônica (abreviada *-is*) representa a forma completa *dicis* e sua reconstituição se dá por seu acréscimo ao segmento inicial do lema (DIC-); já a terceira forma canônica (abreviada *-xi*) aproveita um segmento diferente do lema (DI-) para se reconstituir na forma completa *dixi*.

O modo tradicional de apresentar informações lexicais se insere no contexto dos mecanismos de condensação textual típicos de dicionários impressos (Wiegand, 1996) e pode ser explicado por razões de economia de espaço; de fato, duas ou três dezenas de caracteres são suficientes para veicular diversos tipos de informação (classe, paradigma, bases lexicais). Porém,

são informações implícitas, que exigem do usuário uma capacidade de decodificação metalinguística nem sempre presente em aprendizes, especialmente nas fases iniciais da aprendizagem.

Em um dicionário eletrônico como o **eLatim**, essa necessidade de economia espacial não se impõe, permitindo não apenas apresentar as formas canônicas da palavra em versões integrais (não fragmentadas) mas também acrescentar informação explícita (e propositalmente redundante) sobre suas propriedades flexivas, por meio de recursos como descrição morfológica de cada forma canônica e destaque tipográfico sobre o segmento da palavra a ser reutilizado para a declinação (e.g. dicere : infinitivo presente ativo), acompanhada da indicação explícita do *tema* da palavra (tema do *infectum* : **dic-**).[40]

Quanto à informação sobre o paradigma flexional que se encontra implícita nas formas canônicas, ela é fornecida no **eLatim** de modo explícito, por meio de metalinguagem, em seção específica (cf. 2.3.3, adiante).

2.3.2 Formas variantes

Além das formas flexionadas, são igualmente frequentes em dicionários de latim informações sobre variações na forma do lema. Sua existência pode ser atribuída à longa diacronia em que o latim foi falado e, principalmente, manuscrito, como possível reflexo de influências regionais, do nível de instrução do autor, de escolhas estilísticas ou, ainda, de mudanças no padrão ortográfico estabelecido (Passarotti et al., 2017, p.26), sendo observadas tanto entre textos de diferentes épocas quanto entre diferentes versões editadas de um mesmo texto.

Suas ocorrências podem ser divididas em dois tipos: variantes gráficas e variantes lexicais.

Por variante gráfica se entende uma forma alternativa de representação escrita do item lexical, normalmente causada por diferenças (ou ausência) de convenção tipográfica.[41] Entre os exemplos mais comuns estão a alternância entre U/V (e.g. *voluptas/uoluptas* ‘prazer’), I/J (e.g. *iaceo/jaceo* ‘estou deitado’), I/Y/U (e.g. *lacrima/lacruma/lacryma* ‘lágrima’, *optimus/optumus* ‘o melhor’) e a presença ou ausência de H (e.g. *sulphur/sulfur* ‘enxofre’, *erus/herus* ‘proprietário’), entre outras possibilidades.

Variantes lexicais são aquelas em que se verifica uma mudança (ou restrição) de paradigma morfológico entre as formas. Ocorre, por exemplo, com certos estrangeirismos, devido à influência persistente do paradigma da língua fonte (e.g. palavras gregas como *cometes/cometa* ‘cometa’ ou *diameter/diametros/diametrus* ‘diâmetro’). Outras refletem um trânsito interno entre paradigmas semelhantes do latim (e.g. *domus* ‘casa’ tem formas flexionadas com vogal temática alter-

[40] Temas são segmentos de palavra formados por uma raiz e uma vogal temática ou consoante ou semiconsoante. Constituem uma maneira alternativa e econômica de representar a base lexical usada para a flexão nominal e verbal segundo critérios linguísticos. Para Longo (2006), a importância do tema está em se opor, no interior do sistema, à flexão; porém, afirma que “só faz sentido pensar em temas como parte de um método linguístico” de ensino-aprendizagem (Longo, 2006, p.48). Nesse sentido, oferecer informação sobre o lema em versões alternativas permite que o dicionário seja uma ferramenta de consulta à informação em princípio menos específica com relação aos diferentes métodos utilizados.

[41] A possibilidade de que essas diferenças de grafia representem formas alofônicas não pode ser descartada, tampouco comprovada com segurança, de modo que não é aqui utilizada como critério de distinção.

nando entre -O- e -U-). Algumas são lematizadas por uma forma canônica distinta do padrão de citação, que dá relevo a apenas uma parcela do paradigma flexional por ser mais frequente; são exemplos os verbos depoentes, lematizados pela forma passiva (e.g. *sequo/sequor* 'seguir'), os verbos impessoais, lematizados pela terceira pessoa (e.g. *pudeo/pudet* 'envergonhar'), e alguns substantivos lematizados pela forma do plural, normalmente envolvendo alteração de significado (*aedes* 'casa', em contraste com *aedis* 'templo').

Ambos os tipos de variantes devem ser registrados pelo dicionário do modo mais completo possível, para contemplar o usuário independente do padrão utilizado pelo texto que ele tem em mãos. Esse registro pode ser feito no interior da microestrutura, por meio de listas específicas para cada tipo e acessíveis por busca textual. Vale notar que, no caso das variantes lexicais, a diferença de paradigma flexional é um critério suficiente para determinar sua inclusão também como item da macroestrutura (Passarotti et al., 2020). Já no caso das variantes gráficas, apenas uma das formas consta da nomenclatura, selecionada de preferência segundo um critério filológico (a forma considerada mais correta), ou então pelo critério de frequência (aquela com mais ocorrências no *corpus*).

2.3.3 Classes e propriedades gramaticais

As classificações gramaticais têm importância dupla em lexicografia, seja como descritores de propriedades inerentes das palavras (Atkins; Rundell, 2008), seja como critério distintivo entre lemas semelhantes (Passarotti et al., 2020).

O nível mais geral de classificação é atribuição a uma classe de palavras, que incide sobre todos os itens lexicais e contém informações básicas sobre seu comportamento sintático (Borba, 2003). Outras categorias são específicas de certas classes, como o pertencimento a um paradigma flexional, que se aplica apenas às classes flexivas (isto é, verbos, substantivos, adjetivos e pronomes) e o gênero gramatical (determinado apenas no caso dos substantivos).

Malgrado sua importância, muitos dicionários "economizam espaço com a omissão dessa informação" (Welker, 2004, p.115). Até mesmo uma indicação quase onipresente nos dicionários modernos como a classe de palavras pode ser omitida em alguns dicionários impressos de latim, dada a possibilidade de tais informações serem deduzidas das formas canônicas, como dito acima; nesses casos, apenas as classes invariáveis, cujo lema consiste em uma forma única, recebem essa etiqueta.[42]

Em um dicionário eletrônico como o **eLatim**, em que a economia de espaço não é uma exigência, toda informação relevante sobre as propriedades gramaticais do lema pode ser veiculada de forma explícita; isso significa atribuir a todos os itens lexicais tanto uma classe gramatical quanto um paradigma flexional, incluindo a indicação de invariável para itens de classes não flexivas. Além disso, outras categorias se atribuem de modo específico, como gênero (todos os substantivos), número (itens que só ocorrem no plural) ou restrições formais como depoentes

[42] Uma visão detalhada dos dicionários impressos latim-português é feita no Capítulo 2.

(verbos que só ocorrem em formas da passiva) ou defectivo (verbos que só ocorrem em formas do perfeito).

2.3.4 Pronúncia

Na prática lexicográfica, a informação sobre pronúncia é tratada de modos diversos, variando nos dicionários da presença sistemática à completa omissão. Welker (2004, p.112–3) nota a importância especial concedida à ortoépia pela metalexicografia bilíngue, em geral justificada por uma presumida necessidade de auxílio especial da parte de falantes não nativos; porém, o autor argumenta que, mesmo para estes, "a pronúncia só precisa ser indicada nos casos em que não existam regras claras", normalmente aprendidas "nos primeiros semestres de aprendizado".

Em se tratando de uma língua histórica, o conhecimento sobre a pronúncia é sempre presumido e convencionado *a posteriori*, dada a ausência de uma comunidade ativa de falantes. No caso do latim, assim como se observou no âmbito da grafia, a longa diacronia em que essa língua se manteve em uso teve repercussões sobre o modo de realizar fonemas e pronunciar palavras. Faria (1958, p.23–24) menciona sucessivos esforços de se "corrigir a má pronúncia do latim", desde Alcuíno de York no século VIII, passando pelos humanistas do século XVI como Erasmo de Roterdã e Pedro Ramo[43], até a Conferência Internacional de Instrução Pública, ocorrida em Genebra em 1938, que recomendou que "a pronúncia do latim fosse unificada, na medida do possível, em função das descobertas científicas recentes" (entenda-se: de acordo com a pronúncia que teria vigorado em Roma no período clássico, *circa* I a.C., conforme a reconstituição feita por estudos de fonética latina realizados à luz da Gramática Comparativa e da Linguística então nascente). No contexto brasileiro, segundo o autor, a recomendação ao uso dessa pronúncia "reconstituída" ou "restaurada" aparece já nos programas de ensino da reforma Campos[44] e foi ampla e progressivamente adotada pelos professores desde então, sendo considerada atualmente a pronúncia *standard* do latim em contextos acadêmicos.

As regras da pronúncia acadêmica são bem definidas e costumam ser apresentadas junto com o alfabeto latino nas primeiras páginas de gramáticas – como em Faria (1958, p.15–25) — e métodos de ensino – por exemplo, Amarante (2018, p.39–48).

Há dois conjuntos de regras: as que definem o modo de pronunciar letras e sílabas e as que definem o modo de acentuar as palavras.

O primeiro conjunto contém basicamente instruções sobre o modo de articular os sons representados por cada letra ou combinação, como ditongos e consoantes duplicadas. A identidade do alfabeto usado em latim e em português permite que a maior parte das regras se volte

[43] Este, aliás, conhecido como introdutor dos grafemas J e V, também conhecidos como letras ramistas, para auxiliar os aprendizes a distinguir as ocorrências de I e U como semivogais, respectivamente (Araújo, 2007, p.24).

[44] Referência ao Decreto no 19.851, de 11 de abril de 1931, também conhecido como Lei Francisco Campos em virtude do nome do Ministro da Educação e Saúde Pública à época, que reformou o ensino secundário universitário e criou a Faculdade de Educação, Ciências e Letras para formação de professores licenciados (Santos Sobrinho, 2013).

para as diferenças de valor fonético em relação às regras de pronúncia do português – e.g. letras com valor ambíguo em português (como X, C, G, R, entre outras) têm especificado seu valor único e constante em latim: X sempre tem som de /ks/; R é sempre vibrante; C e G são sempre oclusivas velares, etc.). Tais especificações têm caráter universal e independem do contexto de sua ocorrência no item lexical.

Já o segundo trata de definir o acento da palavra em termos de sílaba tônica. Nesse caso, a diferença de sistema fonológico entre o latim (baseado em duração/quantidade) e o português (baseado em timbre/intensidade) exige que se aplique uma regra de conversão para que a distinção possa ser realizada pelo falante lusófono. Essa regra se baseia em um único critério, com duas possibilidades: palavras com a penúltima sílaba longa são pronunciadas como paroxítonas; palavras com a penúltima sílaba breve são pronunciadas como proparoxítonas. Sua aplicação, portanto, depende da configuração específica de cada item lexical, o que faz do dicionário o instrumento apropriado para registrar a informação sobre as quantidades silábicas de forma sistemática: no mínimo, a quantidade da penúltima sílaba, suficiente para se determinar a pronúncia da palavra isolada; de preferência, as quantidades de todas as sílabas da palavra, considerando sua utilidade tanto como critério de distinção de homógrafos[45] quanto como fundamento de análises métricas de versos e cláusulas oratórias.

Por isso, a forma mais comum de apresentar a informação sobre a pronúncia em dicionários de latim consiste em especificar a quantidade silábica por meio de sinais gráficos específicos aplicados sobre as vogais: o macro, um traço retilíneo que indica vogal longa (e.g. *ā ē ī ō ū*); e a braquia, traço curvilíneo que indica vogal breve (e.g. *ă ĕ ĭ ŏ ŭ*). Por seu caráter tradicional, que pressupõe um conhecimento pelo usuário, esse padrão é mantido no **eLatim**. Não obstante, existem outras formas utilizadas modernamente para representar a pronúncia tendo em vista outros públicos – e.g. o alfabeto fonético internacional para uso de linguistas ou foneticistas, ou então um sistema de reescrita baseado nas regras de acentuação da língua materna, comuns em dicionários para aprendizes (Atkins; Rundell, 2008, p.206) –, que podem ser eventualmente incluídas conforme sua disponibilidade na base de dados.

2.3.5 Etimologia

A etimologia de uma palavra consiste em sua origem e desenvolvimento histórico. São informações que permitem construir-lhe uma espécie de "biografia". Etimologia também é o nome do tipo de pesquisa que permite (re)construir essa biografia, como define Durkin (2009, p.2):

> Etimologia pode ser definida como a aplicação, no âmbito de cada palavra, dos métodos e *insights* de diferentes áreas da linguística histórica para produzir uma

[45] Por exemplo: *dĭco* 'dedicar' e *dīco* 'dizer'; *occĭdo* 'cair' e *occīdo* 'matar'. Vale notar, porém, que, à exceção de versões adaptadas para aprendizes, as edições de referência dos textos latinos não marcam tipograficamente essa distinção, o que impede que a informação sobre pronúncia tenha utilidade na resolução de ambiguidades quanto à lematização.

descrição coerente da história da palavra.[46]

O autor inclui, como objetos dessa investigação, tanto questões intralinguísticas (como mudanças de forma e de significado) quanto relações interlinguísticas (como empréstimos linguísticos e cognatos). Além disso, aponta como fator determinante da pesquisa a disponibilidade de informação diacrônica sobre uma língua, variável conforme o nível de documentação histórica de cada período: alguns são bem documentados, outros exigem a reconstrução de estágios hipotéticos.

Salmon-Alt (2006, p.3–4), entendendo etimologia em sentido estrito (diacrônico) como "origem e evolução de um lexema *antes* de este entrar na língua objeto de descrição" (grifo adicionado), observa que esse tipo de informação é normalmente apresentada em dicionários na forma de um ou mais étimos associados a uma classe etimológica (herança, empréstimo, criação linguística).[47] No entanto, como nota Welker (2004, p.116), muitos dicionários incluem no lugar destinado à etimologia na microestrutura (caracterizado por uma mesma posição e apresentação gráfica)[48] não apenas informações históricas, relativas a diferentes períodos da diacronia, mas também informações de natureza morfológica acerca dos elementos formadores do item lexical (i.e. bases lexicais e afixos); nesse caso, considera-se que palavras primitivas "dão origem" às derivadas e compostas e, portanto, são anteriores a elas, ainda que possam coexistir em uma dada sincronia da língua.

No dicionário da Língua Portuguesa de Houaiss (1986), por exemplo, encontram-se, sob o rótulo da etimologia, as três classes de informação acima mencionadas: para o verbo *descontar*, os elementos formadores "*des-* + *contar*"; para *futebol*, a origem por empréstimo linguístico ("ing. *football*: id."); e para *dicção*, a origem por herança latina ("lat. *dictio,ōnis*: palavra, discurso, maneira de dizer"). Na mesma obra, aliás conhecida pela abundância de informações dessa natureza, essa retrospectiva pode chegar em alguns casos à sua máxima profundidade histórica, combinando métodos e classes de informação para retroceder até o estágio mais antigo documentado; a etimologia da palavra *ocidente*, por exemplo, remonta ao "lat. *occĭdens,entis* 'ocidente, poente', substv. de *occĭdens,entis*, part.pres. de *occidĕre* 'pôr-se (falando-se dos astros), cair', de *ob-* + *cadĕre* 'cair'".

Dicionários de latim compartilham desse cenário de diversidade informacional relacionada à etimologia. Há, porém, uma diferença importante com relação às línguas modernas no que tange ao aspecto hereditário das palavras: o latim não tem um ascendente linguístico docu-

[46] No original: "etymology can be defined as the application, at the level of an individual word, of methods and insights drawn from many different areas of historical linguistics, in order to produce a coherent account of that word's history."

[47] No original: "As a part of diachronic information, etymology properly speaking is concerned with the origin and evolution of the lexeme before it entered into the target language. It is generally presented as a set of one or more etymons, associated with an etymological class (inheritance, loan word, word generation).".

[48] Em estudos de lexicografia descritiva, não é incomum definir um tipo de informação por seu lugar na microestrutura, com base na caracterização gráfica, como se vê em Durkin (1999, p.1): "In the present article the word 'etymology' will be taken to refer, very broadly and simply, to all material presented between square brackets at the head of an entry in the printed text."

mentado. Logo, a exposição histórica do vocabulário latino se divide em duas abordagens (Ernout; Meillet, 2001): (1) a observação de seu desenvolvimento (histórico) desde os documentos mais antigos até a época Romana, com base no estudo filológico dos textos; (2) a consideração de seu desenvolvimento pré-histórico, isto é, os prováveis estágios entre o indo-europeu comum e os primeiros testemunhos escritos dedutíveis pelo método comparativo. No primeiro caso, trata-se de fornecer informação sobre as bases lexicais (formas primitivas) e os diversos itens lexicais (igualmente atestados) em que esta comparece como elemento de composição. No segundo, trata-se de apontar quais itens lexicais de outras línguas explicam a presença de determinada forma no léxico latino. Nas palavras de Ernout e Meillet (2001, p.vii, tradução própria):

> Em uma língua como o latim, deve-se considerar, de um lado, as palavras indo-europeias ou feitas de elementos indo-europeus e, de outro, os empréstimos linguísticos. A posição adotada foi a de que uma etimologia de base indo-europeia só seria utilizada se a associação com outras línguas da família fosse considerada certa ou, no mínimo, muito provável. [49]

Não por acaso, o provimento de informação etimológica pelos dicionários de latim se baseia normalmente em três estratégias:

a) apontar os elementos formadores das lexias complexas – e.g. *memorabilis* é feito de "MEMORO + BILIS" segundo Glare (1968, p.1096);

b) apontar a origem contemporânea dos empréstimos linguísticos, especialmente frequentes no caso de termos oriundos do grego antigo – e.g. o termo *metaphora* tem descrita sua proveniência como "Gk. μεταφορά" em Glare (1968, p.1105);

c) apontar palavras cognatas em outras línguas clássicas, como forma de indicar a origem indo-europeia comum – e.g. o verbo *sum* é descrito como "da m. família q. o gr. εἰμί, o sancr. *asmi*, o hebr. יש *iéch*" por Saraiva (1927, p.1156).

Quanto à disponibilidade da informação etimológica nos dicionários, Atkins e Rundell (2008, p.208) afirmam que ela é encontrada com mais frequência nos grandes dicionários monolíngues e raramente em dicionários bilíngues ou para aprendizes, mas advertem que o surgimento dos dicionários eletrônicos tende a alterar essa situação. Welker (2004, p.117) dá razão a essa prerrogativa dos grandes dicionários de informar sobre etimologia quando afirma que esta "interessa apenas a determinados grupos de usuários, não ao consulente comum", com base na constatação, resultante de pesquisas de uso de dicionários, de que esse tipo de informação é o menos procurado. Por outro lado, estudos na área do ensino-aprendizagem de línguas estrangeiras mostram que o uso de técnicas ativas de aprendizagem de vocabulário baseadas em informação etimológica tem impacto positivo aprendizagem de vocabulário, aumentando a capacidade do aprendiz tanto de fazer inferências de significado pelo conhecimento das partes que formam a

[49] No original: "Dans une langue comme le latin, il faut envisager, d'une part, des mots indo-européens ou faits avec des éléments indo-européens, de l'autre, des mots empruntés. On a estimé qu'une étymologie indo-européenne n'était utile que si le rapprochement proposé avec d'autres langues de la famille était ou certain ou du moins très probable."

palavra quanto de reter novos itens lexicais pelo agrupamento por famílias (Ilson, 1983; Pierson, 1989; Nation, 2013).

Considerando a alta probabilidade de que o usuário do **eLatim** seja um aprendiz de Língua Latina, torna-se especialmente importante fazer constar a informação sobre a origem e a formação das palavras latinas. Além do potencial didático mencionado acima, acrescenta-se a expectativa de que o usuário de dicionários de uma língua histórica tenha um interesse especial em informações históricas sobre o léxico, principalmente quando este constitui a principal matriz lexical de sua língua materna, como é o caso do falante de português.

Atkins e Rundell (2008, p.208) sugerem aprimorar a compreensibilidade da informação por meio de duas estratégias complementares. A primeira consiste em distinguir os três tipos de informação etimológica em seções específicas, destinadas respectivamente à formação da palavra, à origem de um empréstimo e aos cognatos em outras línguas clássicas. A segunda consiste em evitar todo tipo de obstáculo à compreensão do usuário não especialista, como é o caso das inúmeras abreviaturas e outras convenções tipográficas. Esse tipo de aprimoramento depende de recursos ainda não totalmente disponíveis, mas que fazem parte do horizonte de possibilidades de desenvolvimento do **eLatim**.[50]

2.3.6 Definição e equivalentes

A busca pelo significado de uma unidade lexical constitui, sem dúvida, a maior motivação para uma consulta ao dicionário, fazendo da definição um "segmento informativo central em qualquer dicionário de natureza semasiológica" (Bugueño Miranda, 2009, p.243–4). Na verdade, pode-se dizer, com Mel'čuk e Polguère (2018), que a descrição do significado das palavras é a atividade metalinguística mais fundamental, precedendo de fato à própria lexicografia.[51]

Definir o(s) significado(s) de um signo linguístico não é trivial. O próprio termo 'definição' admite um emprego mais amplo ou mais estrito, dependendo da perspectiva adotada. Uma primeira distinção a ser feita separa a definição propriamente lexicográfica – própria de dicionários, que descreve o significado de uma perspectiva linguística, mais preocupada com o signo do que com o referente real[52] – das chamadas definições enciclopédicas (que fornecem não simplesmente uma definição, mas um resumo às vezes extenso de conhecimentos acerca do referente) e terminológicas (que se ocupam de determinar o conjunto de caracteres que fazem parte da compreensão de um conceito) (Welker, 2004).[53]

---

[50] Uma fonte de informação acessível sobre a formação das palavras latinas é Litta, Passarotti e Mambrini (2019).

[51] Nas palavras do autor: "Word definitions are firmly associated with dictionaries, but the practice of defining words – the most essential metalinguistic activity – is probably as old as linguistic communication itself and, in all certainty, is far more ancient than lexicography" (Mel'čuk; Polguère, 2018, p.418).

[52] Segundo Welker (2004, p.118), citando Imbs, a definição lexicográfica quer "apenas traduzir o que, a respeito de um dado 'objeto', a palavra sugere à mente num dado ambiente histórico".

[53] O que não significa que estes tipos não possam ocorrer em dicionários gerais, quando da presença de nomes próprios de pessoas e lugares ou termos técnicos na nomenclatura, respectivamente (Atkins; Rundell, 2008, p.198).

Segundo Mel'čuk e Polguère (2018, p.421), a definição encontra seu fundamento no princípio da "equivalência semântica entre duas entidades linguísticas", tradicionalmente chamadas *definiens* e *definiendum* (em latim, 'o que define [algo]' e 'o que está sendo definido', respectivamente), que pode ser representado pela equação: *definiendum* ≡ *definiens*. No contexto do dicionário, o que está sendo definido, do lado esquerdo da fórmula, é o item lexical representado pelo lema; do outro lado, o que define toma a forma de uma paráfrase linguística, quer dizer, uma expressão considerada sinônima do lema. Sendo o *definiens* o elemento principal da equação, é comum referir-se a ele como sendo a própria definição. Logo, quando se fala em tipologia da definição, trata-se geralmente de uma classificação de tipos de paráfrase com base na observação de suas propriedades formais (semânticas e sintáticas).

O tipo mais prestigiado é, sem dúvida, a chamada definição lógica (ou definição analítica, também chamada de aristotélica), com estrutura altamente regrada e uma longa tradição normativa que remonta de fato a Aristóteles e à lógica grega. Diversos trabalhos em metalexicografia a consideram sinônimo de definição lexicográfica e buscam determinar suas propriedades mais adequadas, geralmente consistindo em uma fórmula baseada em um hiperônimo (*genus*) acompanhado de certas características específicas (*differentiae*).[54] Em contrapartida, definições baseadas em sinônimos ou antônimos são consideradas "pseudo-definições" ou "falsa paráfrases" por não apresentarem a consistência e a estrutura sintática da primeira (Welker, 2004).

Em uma abordagem descritiva mais abrangente das paráfrases explanatórias,[55] Bugueño Miranda (2009, p.248) explica essa diferença com base na "perspectiva assumida pelo dicionário em relação ao ato comunicativo da consulta". Nesse caso, o fator determinante é o propósito que leva o usuário a consultar o dicionário, quer dizer, se ele busca conhecimento sobre a significação de um item lexical que lhe serve de ponto de partida (perspectiva semasiológica) ou se ele quer conhecer quais significantes são capazes de expressar determinada ideia ou conceito (perspectiva onomasiológica). A paráfrase explanatória analítica serve ao primeiro caso, consistindo em uma "reescrita (ou glosa) que explicita, por meio de uma proposição, o mesmo conteúdo da

[54] Por exemplo, Mel'čuk e Polguère (2018, p.420): "The lexicographic definition of a lexical unit – from now on, simply *definition* – is a paraphrase of this lexical unit (1) done in the same language as the lexical unit defined, (2) based exclusively on the knowledge of the language itself, and (3) formulated in accordance with six principles […] three of them concern the informational content of a definition, and three its form." O autor faz uma tipologia dos componentes semânticos da definição, distinguindo inicialmente entre o componente central (paráfrase mínima do lema) e componentes periféricos em quantidade ilimitada, cada um "necessário para expressar um fragmento do significado" Mel'čuk e Polguère (2018, p.435), e identificando em seguida uma série de marcadores semânticos e sintáticos que ajudam a definir os itens lexicais de diferentes classes.

[55] Além da discussão apresentada a seguir, Bugueño Miranda (2009) menciona outros tipos de paráfrase segundo o conteúdo em que se baseia o *definiens*, entre os quais: paráfrase meronímica, baseada na relação parte/todo (e.g. *braço* 'membro do corpo'); paráfrase antonímica, baseada na relação de oposição semântica (e.g. *insalubre* 'o que não é saudável'); paráfrase serial, que descreve o lema como um elemento numa série (e.g. *quarta-feira* 'dia entre a terça-feira e a quinta-feira'); paráfrase mista, em que se relacionam sistemas semióticos (e.g. *y* 'segunda variável numa equação matemática'); paráfrase operacional, que indica o significado pelo emprego situacional da unidade (e.g. *finesse* 'fazer algo com *finesse* é agir com estilo e habilidade'); paráfrase etimológica, baseada na relação de derivação (e.g. *bovino* 'relativo a boi'); e paráfrase decomposicional, que descreve o significado por decomposição (e.g. *guarda-roupas* 'móvel para guardar roupas'); note-se que os dois últimos tipos se baseiam em características morfossemânticas.

unidade léxica" (Bugueño Miranda, 2009, p.249); já a paráfrase explanatória sinonímica, por consistir em uma reescrita baseada na substituição de um significante por outro, tem elementos de ambos: de um lado, toma um item lexical como ponto de partida; de outro, pressupõe "um *tertium comparationis* implícito perante o qual o signo-lema e o sinônimo são duas designações para um mesmo conteúdo", que permanece oculto (Bugueño Miranda, 2009, p.250–1).

A discussão é importante porque os equivalentes de tradução, forma mais comum de expressão de significados em dicionários bilíngues, entram no segundo grupo de paráfrases, como expressões de "sinonímia entre elementos de duas línguas distintas" (Molina Garcia, 2006, p.73).[56] De fato, esse tipo de dicionário têm uma dupla função a desempenhar (Szende, 1996, p.118): (1) informar sobre os significados e os empregos das palavras na língua de partida; (2) sugerir formas de traduzir cada um desses significados e empregos na língua de chegada. Logo, pode-se dizer que nem definições analíticas nem equivalentes de tradução, por si mesmos, são suficientes para cumprir tais objetivos, na medida em que ambas consistem em informações parciais que exigem do usuário um esforço de dedução:[57] os equivalentes exigem que ele deduza a forma do conteúdo expresso pelo item na língua estrangeira; as definições, que ele encontre em sua língua materna a forma da expressão desse conteúdo.

Em se tratando de dicionário bilíngue, todavia, parece haver consenso de que a equivalência lexical é fator determinante para sua concepção (Molina Garcia, 2006). A questão que se coloca incide sobre a capacidade dos equivalentes de cumprirem sua função informativa na consulta, isto é, de auxiliar o usuário a resolver seu problema de compreensão. Nesse sentido, há diversas críticas ao modo como os dicionários tradicionais lidam com o problema do anisomorfismo, isto é, a frequente não equivalência de recorte cognitivo entre línguas diferentes (Szende, 1996).[58]

Hanks (2022, p.480) lamenta que dicionários tendam a ser "vítimas da falácia cientificista de tentar definir palavras isoladamente", normalmente apresentando simples listas numeradas de equivalentes (ou mesmo definições) que podem encorajar a crença de que a relação entre palavras e significados é uma simples questão de "*checklist* de condições necessárias e suficientes" (Hanks, 2022, p.478, tradução própria), ao passo que falham em especificar o que distingue uma acepção da outra. Reconhecendo o caráter impreciso desse "acúmulo de quase-sinônimos", Molina Garcia (2006, p.47–48) enumera quatro possibilidades de escolha por parte do usuário diante desse conjunto indiscriminado de equivalentes, todas sujeitas a equívocos: (a) escolher a primeira das opções; (b) escolher qualquer uma; (c) escolher a opção que lhe soa mais familiar (que já tenha usado anteriormente, por exemplo); e (d) escolher aquela que faça sentido

[56] Também (Cimiano et al., 2011) define *translation* como "interlingual synonymy" em sua ontologia de termos lexicográficos (cf. <https://lexinfo.net/ontology/3.0/lexinfo.owl#translation>).

[57] Vale notar que os chamados dicionários semibilíngues ou bilingualizados foram desenvolvidos em torno da combinação de definições analíticas na língua de partida e equivalentes de tradução na língua de chegada, muitas vezes com objetivos pedagógicos. Para um panorama sobre o tema, cf. Duran e Xatara (2005).

[58] Sobre a relação entre a não-equivalência e a hipótese do relativismo linguístico, cf. Biderman (1998, esp. p.96–102). Para uma revisão da literatura sobre a impossibilidade da equivalência, cf. Molina Garcia (2006, esp. p.74–6).

na sua língua materna.[59]. Estudos em lexicografia pedagógica também reconhecem que simples listas de equivalências tornam difícil a escolha do usuário e defendem a importância de se fornecer ao aprendiz de línguas informações explícitas de natureza sintagmática (valência sintática, restrições selecionais, colocações e expressões idiomáticas) e contrastiva (exemplos de uso contextualizados e acompanhados de tradução) (Bergenholtz; Tarp; Duvå, 1995; Welker, 2005; Zavaglia, 2006; Meliss; Sánchez Hernández, 2015; Molina Garcia, 2006).

Argumentos semelhantes são encontrados em estudos específicos sobre dicionários bilíngues latim-português. Lima (1995), por exemplo, identifica nos dicionários tradicionais de latim inadequações no modo de registrar tanto a *significação* (ideia ou conceito que a palavra pode virtualmente representar) quanto o *valor*, isto é, o sentido adquirido pela palavra na "relação que tem com outras palavras com as quais contrasta no interior do sintagma e no discurso" (p.105). O problema principal, segundo o autor, é a falta de clareza da informação, que se encontra normalmente diluída na forma de exemplos ou codificada em abreviaturas gramaticais indicativas de padrões de construção sintática, exigindo em ambos os casos um esforço de inferência por parte do usuário que pode não ser eficaz no caso de aprendizes.

(Wharton, 2011), por sua vez, analisa os modos de distinguir e ordenar significados no âmbito do *Oxford Latin Dictionary* (OLD), apontado como "o mais atualizado e bem aceito dicionário latim-inglês" da contemporaneidade (p.255). Partindo do princípio de que é praticamente impossível chegar uma descrição precisa e completa do sentido de qualquer palavra[60] – e que mesmo os sentidos mais básicos apresentam propriedades semânticas que raramente podem ser capturadas por meio de uma tradução termo-a-termo ou uma série de sinônimos na língua de chegada –, o autor reconhece como qualidades do OLD (p.257): (1) informar sobre o desenvolvimento histórico dos significados das palavras; (2) categorizar os significados de acordo com o contexto em que aparecem, destacando eventuais nuances e sutilezas semânticas; e (3) oferecer sugestões convenientes de tradução vinculadas a cada um desses usos. Todavia, percebe problemas na forma de ordenar significados, na medida em que não deixa claro qual deles é o mais antigo ou o mais básico, de modo que pode "induzir usuários a erro quanto ao entendimento de como o falante latino apreendia a palavra" (p.259).[61] O autor propõe, como forma de produzir

[59] A alta probabilidade de equívoco nessa operação de escolha mencionada por Molina Garcia (2006) pode ser usada para contestar a afirmação, feita por diversos autores, de que dicionários voltados para produção em língua estrangeira precisam de mais informações contextuais do que dicionários para recepção. A título de exemplificação, veja-se o comentário de Duran e Xatara (2007, p.317, grifos nossos): "Na situação de decodificação, o usuário busca o significado e tem o contexto para contribuir com essa busca (pode usar de inferência). Nesse caso, portanto, a criticada lista de equivalentes que os dicionários bilíngues apresentam poderia atender razoavelmente a necessidade do usuário pois ele poderia *testar cada um dos equivalentes* e decidir *com alguma margem de segurança* qual deles é adequado ao contexto em questão. Por outro lado, na situação de codificação, equivalentes só podem ser utilizados com segurança na produção do usuário se o dicionário informa em que *situação de uso* eles ocorrem." É justamente esse tipo de "testagem" de equivalentes (em detrimento de uma escolha linguística consciente) que induz o usuário a erro.

[60] Nas palavras do autor: "…we can't achieve a completely accurate description of the meanings of any word, even in the best of circumstances…every attempt at lexical description is at best an approximate description of things that have an essentially protean nature" (Wharton, 2011, p.256).

[61] No original: "the OLD's arrangement of senses almost certainly misleads modern users who want to understand how ancient speakers typically apprehended the word".

essa compreensão, uma abordagem que distingue entre significados básicos (menos dependentes de pistas contextuais e, portanto, mais frequentes e estáveis) e significados *ad-hoc* (em inglês, *nonce-meanings*, produzidos de forma criativa pelo falante e, portanto, dotados de função estética e caráter temporário). Além disso, um arranjo estruturado em *clusters* de significados que compartilham propriedades comuns permitiria representar o fato de que muitas vezes "não há um limite denotacional claro entre significados" (p.264), evitando que sejam interpretados como discretos e mutuamente exclusivos.

Percebe-se, em suma, que a crítica aponta sobretudo para a dificuldade de captar, apenas por meio de equivalentes, a complexidade semântica que o uso contextualizado de qualquer palavra implica. De fato, na medida em que um equivalente é tradução de um lexema em termos semânticos e também pragmáticos, fica evidente a importância de se fornecer informações adicionais, de modo que atuem como elementos diferenciadores para que o usuário possa escolher o significado (e, consequentemente, o equivalente) mais adequado (Welker, 2004). Tais informações são apresentadas nas seções a seguir, divididas em informação paradigmática, informação sintagmática e informação sobre uso, conforme o tipo de relação entre as unidades dentro do sistema linguístico que descrevem.

2.3.7 Informação paradigmática

No contexto das dicotomias saussureanas (língua e fala, forma e substância, sintagma e paradigma, sincronia e diacronia), relações paradigmáticas são "aquelas existentes entre uma unidade particular em um dado sintagma e outras unidades que lhe são substituíveis nesse sintagma" (Lyons, 1980, p.197). O uso da substituição como critério implica no entendimento da estrutura do sistema linguístico como dependente dos princípios complementares de seleção e combinação e se relaciona com a própria noção de informação; no caso da informação semântica, "a possibilidade de selecionar uma unidade e não outra […] é uma pré-condição da transmissão de mensagens diferentes dentro do sistema em questão" (Lyons, 1980, p.198). O tipo de relação semântica encontrada entre lexemas de um mesmo paradigma permite estabelecer uma tipologia das relações paradigmáticas em termos de similaridade e contraste.

No espectro da similaridade, lexemas são considerados sinônimos quando a seleção de um lexema e não de outro não exerce efeito sobre a mensagem transmitida, ainda que, eventualmente, acarrete alteração de seu significado social ou expressivo (e.g. *comida/alimento*, *rápido/veloz*, *acontecer/ocorrer*). Vale dizer que o contexto é determinante para a sinonímia: o adjetivo *forte* pode ser substituído por *torrencial* em "chuva forte", mas não em "pessoa forte", por exemplo; além disso, é possível haver equivalência também entre segmentos desiguais, como "choveu sem parar" e "choveu ininterruptamente" (Ježek; Sprugnoli, 2023, p.45–46). Lexemas hipônimos são aqueles com um significado mais específico (ou subordinado) em relação a um lexema de significado mais geral (ou superordenado) (Lyons, 1980, p.235); nesse caso, a similaridade reside no fato de compartilharem contextos e atributos, como no caso das palavras *gato* e *cachorro*, que podem ser consideradas hipônimos de um mesmo hiperônimo (*animal de esti-*

mação) porque (1) aparecem junto aos mesmos verbos (*acariciar*, *adotar*, *castrar*) e (2) podem ser qualificadas de modo semelhante (e.g. *filhote*, *abandonado*, *de raça*). Uma terceira possibilidade é a relação de meronímia entre dois lexemas, dos quais se considera merônimo aquele que significa uma parte do todo significado pelo outro (e.g. *braço/corpo*, *roda/bicicleta*, etc.).

No espectro do contraste, as relações podem ser divididas, em um primeiro momento, em binárias e não binárias. No primeiro grupo estão as relações de oposição, tradicionalmente classificadas como antônimos. Lyons (1980, p.220), distinguindo, entre os opostos, aqueles que são graduáveis (isto é, que admitem gradação e comparação, e.g. *quente/frio*) e aqueles em que a gradação não é possível (e.g. *vivo/morto*), prefere utilizar o termo antônimo apenas para os primeiros, na medida em que "manifestam a propriedade de polaridade mais abertamente" (Lyons, 1980, p.226), chamando de complementares os segundos. O autor elenca ainda entre os tipos de oposição lexical: os conversos (e.g. *pai/filho*) e os opostos direcionais (e.g. *subir/descer*, *direita/esquerda*). No segundo grupo estão lexemas que denotam elementos ordenados serialmente ou ciclicamente (e.g. *domingo/segunda/...*, *janeiro/fevereiro/...*, *primavera/verão/...*), numa relação que pode ser descrita como sendo de incompatibilidade (Lyons, 1980, p.233).

A aplicação dessas relações à totalidade do léxico permite estabelecer conjuntos de lexemas e apontar-lhes traços distintivos.

Pela relação de similaridade se definem os campos lexicais, isto é, conjuntos de lexemas "que recobre uma área conceptual e, por meio das relações de sentido entre eles, a estrutura" (Lyons, 1980, p.207). Um exemplo conhecido é a área conceitual da COR, que é uma faixa contínua do espectro luminoso que, ao ser lexicalizado pelas línguas, é recortado em partes por lexemas (e.g. em português, *azul*, *amarelo*, *vermelho*, este último subespecificado em *escarlate*, *carmim*, etc.); importa notar que esse processo resulta em recortes desiguais em quantidade, dimensão e profundidade, observáveis não só na comparação entre línguas diversas, mas também entre duas sincronias de uma mesma língua, a despeito de uma percepção física comum representada pelo arco-íris.

As relações de oposição podem ser aplicadas para analisar o significado de cada lexema em termos de componentes ou traços semânticos. Concebida como uma extensão da teoria dos campos semânticos (Lyons, 1980, p.262), a análise componencial se baseia na detecção (ou atribuição) de valores positivos ou negativos a conceitos atômicos, conforme estejam presentes ou ausentes no conceito molecular representado pelo lexema. Por exemplo, os conceitos moleculares de *homem* e *mulher* são similares por terem em comum os traços +HUMANO +ADULTO, mas se distinguem como complementares quanto à polaridade +MACHO/+FÊMEA.

Esse método permite agregar uma série de subdivisões internas, de base semântica, à divisão tradicional em classes de palavras. Unidades lexicais[62] associadas a lemas substantivos podem ser classificadas, primeiramente, quanto a ±ANIMACIDADE, conforme designem seres ani-

[62] É importante perceber que as classificações semânticas se referem à unidade lexical, enquanto associação unívoca de lema e significado. O fenômeno da polissemia se expressa também na possibilidade de haver diversas combinações de traços possíveis para um mesmo lema.

mados ou inanimados; seres animados podem ser divididos pelo traço ±HUMANO em pessoas e animais, seres inanimados pelo traço ±DEFINIDO em concretos ou abstratos (Márquez Cruz, 2017). Unidades lexicais associadas a lemas verbais são categorizadas segundo o tipo de evento que expressam, com base três traços: ±DINAMISMO (alteração de estado), ±DURAÇÃO (arco temporal) e ±TELICIDADE (conclusão pressuposta) (Ježek; Sprugnoli, 2023, p.54); assim, verbos de estado indicam um evento que possui apenas duração, sem implicar mudança ou conclusão (e.g. *habere* 'possuir (um objeto)', *sedere* 'estar sentado'); verbos de processo indefinido possuem dinamismo e duração, mas não indicam quando terminam (e.g. *natare* 'nadar', *nauigare* 'navegar'); verbos de processo definido se caracterizam por dinamismo, duração e uma conclusão pressuposta (e.g. *exhaurire* 'esvaziar', *parare* 'preparar'); por fim, verbos instantâneos (ou pontuais) expressam o momento em que uma mudança se dá, possuindo portanto dinamismo e telicidade, mas não duração no tempo (e.g. *uenire* 'chegar (a um lugar)', *inuenire* 'encontrar (alguma coisa)').

Sabe-se que dicionários gerais de cunho semasiológico costumam dar informações sobre sinônimos, indicando eventualmente os contextos em que lexemas intercambiáveis carregam diferentes conotações ou valores emotivos; as demais relações, porém, raramente são explicitadas ou mesmo passíveis de serem inferidas das definições, e "nenhum dicionário distingue sistematicamente os diferentes níveis de oposição lexical que se encontram em uma língua" (Lyons, 1980, p.242).

Assim como se disse com relação aos componentes da formação da palavra presentes na informação etimológica, não se deve esperar que o conhecimento das relações paradigmáticas ou dos traços semânticos de uma palavra seja suficiente para dar conta da complexidade de seu significado. Mas é evidente sua capacidade de situar o lexema dentro de um campo lexical afim, ou mesmo dentro do sistema de signos da língua, com vantagens tanto para a interpretação quanto para a retenção de vocabulário, na medida em que auxilia o usuário a memorizar e organizar as unidades lexicais em seu léxico mental da L2 com base nas semelhanças e diferenças entre elas.

No caso dos traços semânticos em particular, um aspecto de sua interdependência com a teoria dos campos importa para a lexicografia bilíngue pela relação com a equivalência: a distinção coseriana entre semas e classemas. Lyons (1980, p.262) explica que semas são traços distintivos mínimos de significado, que operam em um único campo lexical e servem para estruturar esse campo (e.g. os elementos que distinguem *cadeira* e *poltrona* dentro do campo 'móveis de sentar'), ao passo que classemas são componentes de significado de caráter mais geral, presentes em lexemas de diferentes campos lexicais, com tendência para serem não apenas lexicalizados, mas gramaticalizados (e.g. ANIMADO/INANIMADO, CAUSAR/TER). Nesse contexto, afirma o autor (Lyons, 1980, p.263):

> As distinções e equivalências entre lexemas que são explicados em termos de classemas […] são de tal modo que (a) são facilmente identificáveis de uma língua para a outra e parecem depender menos da língua ou da cultura do que

> as distinções e equivalências em termos de semas; (b) são sintaticamente pertinentes e podem ser gramaticalizadas;[63] (c) não se restringem a um número reduzido de lexemas, mas se encontram largamente distribuídas por todo o vocabulário.

Como se vê, esta e outras formas de categorização semântica podem servir como guias para a distinção das acepções (expressas pelos equivalentes de tradução) relacionadas a um item polissêmico. Logo, espera-se de um dicionário bilíngue o provimento de um mínimo de informação paradigmática, constituída por: vinculação a um campo semântico; explicitação de classemas fundamentais (nominais e verbais); lexemas relacionados por similaridade (sinônimos); lexemas relacionados por contraste (antônimos).

2.3.8 Informação sintagmática

A presença de informação sintagmática em dicionários deriva da percepção de que a determinação do sentido das palavras passa pela observação do modo como elas se combinam. A ideia de que o sentido reside no nível da oração – e não no nível das palavras, onde o sentido é apenas parcial – é recorrente nos estudos da linguagem natural desde pelo menos o *Sofista* de Platão.[64] Na linguística moderna, o entendimento do sentido como produto de uma combinatória comparece em duas correntes de análise, com convergência de resultados, segundo McEnery e Hardie (2012): a abordagem funcionalista da gramática e os estudos de *corpora* da escola neo-firthiana.

Do lado do funcionalismo, a perspectiva integrativa do estudo da forma, do significado e do uso busca revelar, pelo estudo das sequências linguísticas, os significados codificados por elas (Neves, 1997). Em sentido amplo, importam, para os propósitos deste trabalho, duas formulações teóricas:

a) o exame das relações sintático-semânticas da estrutura enquanto ativação da função ideacional de Halliday e Matthiessen (2004), isto é, a língua servindo à codificação da experiência através de uma gramática sistêmica em que, entre as relações paradigmáticas no nível abstrato e profundo e as unidades sintagmáticas no nível da realização na superfície, é a escolha que produz significado.[65]

b) a teoria de componentes integrados de Simon Dik (1978), que contempla a estrutura da sentença desde a representação semântica até a forma da superfície, distinguindo os diferentes níveis de organização da frase, da semântica à pragmática, passando pela sintaxe; aqui importa a ideia da centralidade do verbo como predicador principal na

[63] O uso operacional dos traços pelos modelos integrados sintático-semânticos será tratado na próxima seção.

[64] O assunto será retomado no Capítulo 2.

[65] Vale notar que uma das maiores dificuldades reportadas no aprendizado do latim é o sistema de flexão nominal de caso, para cuja compreensão se considera extremamente esclarecedor saber que "toda língua, em estrutura profunda, tem casos semânticos abstratos (agente, paciente, beneficiário), que em cada uma delas ganha substância concreta de forma variada" (Vasconcellos, 2015, p.56).

estruturação da frase, que comparece na teoria das valências de Tesnière (1959),[66] na gramática de casos de Fillmore (2003) e na classificação semântica de Chafe (1979) e Levin (1993).

A integração das teorias funcionalistas de análise linguística para fins de lexicografia encontra aplicação nos trabalhos de Borba (Borba, 1992; Borba, 1996; Borba, 2003).[67] Sua proposta consiste na criação de dicionários de língua com base em uma atitude reflexiva (não intuitiva) perante a língua, que parte da observação do comportamento dos itens em um *corpus* organizado de realizações efetivas e autênticas para então realizar uma análise que vai do esquema superficial para o esquema profundo, ou seja, da sintaxe para a semântica. O objetivo é descrever o léxico por meio de regras de combinação, estabelecendo tanto a valência quanto as relações sintático-semânticas entre o predicado e os argumentos. O dicionário assim construído ultrapassa a função de simples repositório de palavras, tornando-se um "guia de uso, […] um instrumento pedagógico de primeira linha" (Borba, 2003, p.16). De fato, informações sobre valência sintático-semântica, sobretudo quanto à forma dos actantes e às restrições selecionais, são tidas como altamente relevantes para os usuários (Welker, 2005).

Para tanto, a análise e descrição do léxico deve se realizar no contexto de uma teoria gramatical integrada em termos sintático-semânticos. Partindo do princípio de que "falar é predicar" (Borba, 1996, p.13), a análise sintática parte da referência para determinar o modo como o valor semântico se constrói a partir dos conceitos expressos pelas palavras no interior do predicado, entendido como estrutura mínima do ato de comunicação e também como função, isto é, uma "relação entre duas grandezas que se condicionam mutuamente" (Borba, 2003, p.12), cujas variáveis são os argumentos. É um modelo centrado no léxico, que tenta descrevê-lo por meio de regras de combinação. O método de análise consiste em (1) identificar a unidade que preenche a função de núcleo do predicado; (2) identificar o número de *casas vazias* (argumentos) que comporta a unidade e, então, (3) determinar as relações de dependência entre as unidades que (co)ocorrem em um contexto, isto é, sua valência (e.g. "X depende de Y" ou "X pressupõe Y"). Nesse conceito o autor inclui três dimensões (Borba, 2003, p.20-21): o número de argumentos que um predicado pode ter (valência quantitativa); as propriedades morfossintáticas dos itens que preenchem os argumentos (valência qualitativa); as características categoriais ou traços (e.g. ±ANIMADO, ±HUMANO, ±CONTÁVEL), bem como as funções temáticas ou papéis (agente, beneficiário, experimentador) e as restrições selecionais do item que funciona como predicado,

[66] A pesquisa bibliográfica localizou um trabalho dedicado a aplicar a teoria de Tesnière ao estudo do latim (Bussarello, 1981), mas o acesso ao material não foi possível.

[67] Entre outras propostas de análise sintático-semântica, pode-se mencionar a análise de padrões de Hanks (*Corpus Pattern Analysis* ou CPA), para a qual um padrão consiste em uma valência baseada nas categorias da teoria de Halliday (i.e. a definição de quantidades e tipos de papéis sintáticos associados a cada significado de uma palavra) e uma seleção de colocados típicos em cada papel sintático (Hanks, 2022, p.486); e os cenários de Fillmore (*Frame Semantics*), para quem "the proper way to describe a word is to identify the grammatical constructions in which it participates and to characterize all of the obligatory and optional types of companions (complements, modifiers, adjuncts, etc.) which the word can have in such constructions, in so far as the occurrence of such accompanying elements is dependent in some way on the meaning of the word being described" (Atkins; Rundell, 2008, p.144).

que determina a co-ocorrência ou exclusão de características do item na função de argumento (valência semântica).

Do ponto de vista lexical, qualquer classe de palavras pode funcionar como predicador, desde que esteja relacionada com um argumento (este sempre representado por um nome) (Borba, 2003, p.47). No âmbito da oração, a classe de palavras que tipicamente ocupa a posição de núcleo do predicado são os verbos, cujo significado determina o "esqueleto semântico" da estrutura (Ježek; Sprugnoli, 2023, p.50). Em termos gerais, verbos denotam eventos (e.g. *chover*, *nascer*, *alugar*, *emprestar*) ou estados (e.g. *morar*, *possuir*). Com exceção de um evento determinado pela natureza como 'chover', nenhum desses significados constitui uma predicação completa por si, exigindo um ou mais argumentos que forneçam informações sobre os participantes do evento ou estado a que se referem. Em "meu filho nasceu ontem", por exemplo, o verbo *nascer* pressupõe pelo menos um argumento expressando quem ou o que nasceu (*filho*), que ocupa a função sintática de sujeito e deve possuir traço semântico de um ser vivo (+ANIMADO) no papel temático de experimentador; já em "Morei sete meses na Itália", o verbo *morar* implica em dois argumentos, um designando quem mora (sujeito +HUMANO no papel temático de tema) e outro designando o lugar (complemento locativo).[68]

A conexão sintático-semântica se verifica com mais clareza, todavia, quando se considera o fenômeno da alternância argumental (Ježek; Sprugnoli, 2023, p.51). A análise de um verbo como *emprestar* identifica ao menos duas interpretações que se refletem em duas estruturas argumentais distintas:

- quando significa 'tomar emprestado', pressupõe (1) um sujeito humano no papel de beneficiário, (2) um complemento objeto (e.g. *livro*) e (3) uma origem humana ou não (e.g. *colega* ou *biblioteca*);
- quando significa 'ceder por empréstimo', pressupõe (1) um sujeito que pode ser humano ou inanimado (*um colega* ou *a biblioteca*); o objeto e o beneficiário como complemento.

A estrutura argumental de nível profundo especifica os dois níveis da conexão sintático-semântica. De um lado, o conjunto de traços ou as classes semânticas com que as palavras podem ou costumam combinar-se é o que constitui as chamadas preferências selecionais. Por esse termo se designa "a tendência de uma palavra para selecionar ou restringir semanticamente quais outras palavras podem ocorrer numa relação sintática direta com ela" (Roberts; Egg, 2014, p.511). Por exemplo, o verbo *beber*, em seu uso literal, implica uma restrição sobre o sujeito (traço +ANIMADO, classe Animal) e sobre o objeto (traço +LÍQUIDO, classe Bebida): "bebi um suco de laranja" (mas não "bebi uma laranja"); porém, processos metafóricos podem relativizar tais restrições, já que podemos dizer, por analogia, que "o carro bebe muito", no sentido de

[68] Vale notar que, nos dois casos, a informação sobre o momento (*ontem*) ou a duração do evento (*sete meses*) não são indispensáveis para a completude da predicação, não fazendo parte, portanto, da estrutura argumental; não obstante, esse caráter acessório dos adjuntos ou satélites, como são chamados, não se aplica ao nível discursivo, onde podem desempenhar papéis decisivos do ponto de vista da comunicação.

que consome muito combustível. De outro, a estrutura argumental também especifica a subcategorização, isto é, a forma que cada argumento deve assumir quando alcança a superfície da linguagem. Retomando o caso de *emprestar* significando 'pegar emprestado', por exemplo, o argumento que representa a origem se expressa por um sintagma preposicionado na forma *de* + Nome; quando significa 'ceder por empréstimo', o beneficiário se expressa por um sintagma preposicionado na forma *para* + Nome.[69]

Os princípios da teoria funcionalista são aplicados de forma sistemática à descrição da gramática latina por Pinkster (2015) e Baños Baños (2009). Este último também propõe uma análise dos *marcos predicativos*, termo com que designa em conjunto os diversos aspectos do predicado e os elementos que nele co-ocorrem (Baños Baños, 2003) e que inclui informações acerca da forma fonética do predicado, de seu conteúdo semântico, de sua classe de palavras, de sua valência quantitativa, bem como do conteúdo lexical típico de cada argumento e da função semântica que tal argumento desempenha.[70]

O critério da tipicidade lexical, aliás, está ligado a uma das principais contribuições da Linguística de *corpus*: o princípio idiomático formulado por Sinclair (1991). Segundo ele, na realização da linguagem, o usuário tem a sua disposição um grande número de frases mais ou menos pré-fabricadas que constituem escolhas únicas, mesmo que teoricamente possam ser analisadas. Essa observação se fundamenta na alta frequência de casos em que argumentos deixam de ser "casas vazias" (isto é, preenchíveis por livre escolha do usuário) para terem suas possibilidades massivamente reduzidas senão preenchidas de forma fixa. Diante disso, o autor postula a existência de um princípio de produção de enunciados que levaria o falante a selecionar combinações "pré-fabricadas" de palavras como unidades cujos elementos (palavras) perdem suas propriedades individuais (ou "identidades semânticas") a ponto de, em certos casos, subverterem as próprias regras gerais de combinação (Sinclair, 1991, p.110). Assim, o princípio idiomático teria no mínimo a mesma importância da gramática para explicar como o sentido se constrói no texto e indicar o conhecimento que os usuários de uma língua têm para interpretar cada segmento de um enunciado.

Segundo Sinclair (1991), o princípio idiomático está ligado a uma tendência geral das palavras mais frequentes de terem sentidos menos independentes e, por conseguinte, mais difíceis de identificar e explicar. Essa tendência pode ser vista como uma espécie de delexicalização progressiva, isto é, de "redução da contribuição distintiva da palavra para o significado" (Sinclair, 1991, p.113), de modo que se deva falar em usos de uma palavra, mais do que de significados. Para o autor, o fenômeno que melhor ilustra o princípio idiomático é a colocação, entendida como ocorrência estatisticamente significativa de duas ou mais palavras num texto a uma curta

[69] Tentativas de fazer o mapeamento completo das possibilidades valenciais do português podem ser encontradas nas obras de Borba (1996) e Perini (2008). A aplicação do modelo na descrição "tentativamente exaustiva" de 6 mil verbos do português encontra-se em Borba (1992). Em tempo: a análise das preferências selecionais participa de tarefas computacionais como desambiguação de sentido (*word sense disambiguation*) e resolução de pronomes (*pronoun resolution*), entre outras aplicações (Zapirain et al., 2013).

[70] Um exemplo de aplicação desse modelo para delimitar os marcos predicativos do lema verbal latino *uenio* encontra-se em Cabrillana (1997).

distância (no máximo, quatro palavras) uma da outra, que pode ter tanto um caráter inusitado ou ser de frequente repetição (Sinclair, 1991, p.170). Importa perceber que essa definição reconhece as colocações exclusivamente do ponto de vista lexical,[71] isto é, como co-ocorrências de palavras observadas por meio de linhas de concordância, em que a palavra central é chamada de nó e as palavras à direita e à esquerda são os colocados. O próprio autor sugere, não obstante, que a associação de padrões lexicais com padrões de escolha gramatical pode conduzir a um tesouro de expressões idiomáticas e frases feitas. Nesse âmbito se posicionam pesquisas sobre preferências semânticas (Stubbs, 2001) e sobre padrões gramaticais de complementação de palavras nucleares, tendo em vista a co-ocorrência entre estruturas sintáticas e itens lexicais particulares (Hunston; Francis, 2015).

A aplicação integrada da estrutura argumental e do princípio idiomático permite descrever com mais precisão e detalhes o fenômeno das lexias complexas. Borba (2003) estabelece três propriedades que definem um item desse tipo: caráter recorrente fixo, estrutura sintática típica e valor semântico específico; em seguida, apresenta uma classificação em termos de "graus de estruturação e automatização das construções sintáticas". Assim, em oposição ao sintagma livre, em que os elementos se distribuem de modo independente e guardam autonomia sintático-semântica, o autor identifica três tipos de sintagmas especiais: sintagma semi-fixo, sintagma fixo e oração fixa (frase feita). Um sintagma semi-fixo se verifica quando há dependência unilateral entre termos co-ocorrentes (isto é, quando um dos termos só ocorre em combinação com o outro, e.g. o complemento de *arregalar* é necessariamente "os olhos", *potável* só se aplica a *água*). O sintagma fixo se dá quando os elementos formam um bloco único e imutável e escapam à dinâmica construcional, sendo substituíveis apenas em conjunto por uma palavra equivalente; podem ter um núcleo verbal (e.g. "o cocheiro amarrou a cara e saiu furioso", isto é, aborreceu-se) ou um núcleo nominal (e.g. "falou em alto e bom som", isto é, em voz alta); suas características básicas são "a fixidez dos constituintes associada a um valor semântico-pragmático particular, resultante da combinatória específica estabelecida" (Borba, 2003, p.34);[72] A oração fixa ou frase feita nomeia construções como provérbios, adágios, ditados, máximas e sentenças, os quais possuem estrutura interna congelada, sintaticamente autônoma, podendo ser usada como enunciado independente; do ponto de vista semântico, tais construções expressam um valor de verdade geral ligado a um aspecto ou à totalidade da situação de enunciação na medida em que, "pelo seu caráter lexicalizado e, então, fixo e convencional, reduzem globalmente a situação a uma

[71] A definição proposta por Ježek e Sprugnoli (2023, p.40) pretende incorporar critérios linguísticos: "una collocazione è una combinazione di parole soggeta a una restrizione lessicale, per cui la scelta di una specifica parola (il **collocato** per esprimere un determinato significato, è condizionata da una seconda parola (che chiamaremo qui **base**) alla quale questo significato è riferito".. Uma revisão da literatura sobre o conceito de colocação encontra-se em Beneduzi (2008).

[72] O autor inclui nesta categoria os nomes compostos, por serem blocos formados pela justaposição de formas livres que não permitem determinação parcial, isto é, a especificação e a modificação é sempre global (e.g. "o pé-de-cabra" > *"o pé-da-cabra", "muita alma penada" > *"alma muito penada". Além disso, observa que "a incidência de verbos particulares na formação de sintagmas fixos liga-se à sua produtividade nos textos: quanto maior a frequência de um verbo e maior sua variação contextual, maior a possibilidade de cristalização de algumas de suas estruturas" (id., p.30); como exemplo, menciona os 110 sintagmas fixos com o verbo *dar*, 35 com o verbo *ir*, entre outras registradas em seu *Dicionário de Usos do Português do Brasil*.

instância de uma categoria conhecida" (Borba, 2003, p.35).

Esses tipos de sintagmas correspondem ao que, em linguística computacional, se conhece como expressões multipalavras (em inglês, *multiword expressions* ou MWE). Segundo Ramisch (2012, p.25–27), três propriedades caracterizam as MWE: alto grau de arbitrariedade e institucionalização; variabilidade limitada; e heterogeneidade. A primeira se refere ao fato de que muitas expressões correspondem a modos convencionais de dizer as coisas, sem qualquer explicação aparente do ponto de vista sintático ou semântico, mas em conexão direta com as tendências colocacionais dos itens.[73] A variabilidade limitada se verifica em dois níveis, sintático e semântico. No nível sintático, trata-se da 'não-modificabilidade' da expressão, isto é, o fato de que as regras sintáticas convencionais não se aplicam às MWE, fato que se traduz em duas sub-propriedades: a extragramaticalidade, que explica o caráter imprevisível e a estranheza que expressões como "dar para trás" ou "é um Deus nos acuda" podem causar em quem conhece apenas o léxico básico e as regras gramaticais (aprendizes de segunda língua, por exemplo); e o grau de lexicalização referente ao conhecimento de que um conjunto de palavras pertence ao léxico em bloco, devendo ser representado como uma unidade lexical autônoma. No nível semântico, a variabilidade limitada engloba sub-propriedades como:

- caráter não-composicional, isto é, o sentido do todo não pode ser totalmente inferido da soma de sentidos das partes, em uma escala que vai da expressão quase transparente (e.g. "parada de ônibus") à expressão totalmente opaca (e.g. "amarrar o burro na sombra");
- caráter não-substituível, relacionado com o anterior, já que não é possível substituir um dos componentes de uma expressão por um sinônimo ou equivalente (e.g. *"amarrar o jumento na sombra");
- impossibilidade de tradução palavra-por-palavra, como resultado das duas propriedades anteriores, podendo ser usado como teste para a detecção de MWE;
- idiomaticidade ou especificidade de domínio, na medida em que as MWE costumam pertencer a um contexto de uso específico, seja de caráter técnico-científico, seja de caráter regional/dialetal ou, ainda, ligado a gêneros literários.

Por fim, a última propriedade é a heterogeneidade das expressões multipalavra, que resulta da soma das anteriores e torna difícil encontrar uma definição e classificação abrangentes.

Considerando as diversas propostas de classificação e hierarquização encontradas na literatura, Ramisch (2012, p.30–31) constrói uma tipologia simplificada com duas classificações: a primeira se baseia na função morfossintática das expressões no enunciado;[74] a segunda se baseia na dificuldade que oferecem para sua detecção por métodos computacionais.

[73] O autor aponta ser esse um dos motivos, aliás, para que as MWE sejam consideradas difíceis para aprendizes de segunda língua, uma vez que eles conhecem o léxico e as regras de combinação, mas muitas vezes não possuem conhecimento sobre o uso da língua.

[74] Uma classificação semelhante se encontra em Borba (2003, 28–30).

A classificação de base morfossintática divide as expressões multipalavras de acordo com a classe da palavra que atua como núcleo do sintagma. Assim, denominam-se expressões nominais as combinações de palavras que atuam como um substantivo no enunciado; são subtipos os compostos nominais, isto é, sequências formadas por um núcleo nominal associado a outros elementos, normalmente adjetivos ou locuções adjetivas (e.g. "roleta russa", "sinal de trânsito"); os nomes próprios, que denotam entidades nomeadas específicas como lugares, pessoas, instituições (e.g. "São Paulo", "Chico Buarque", "Bossa Nova", respectivamente); e termos técnicos usados em domínios especializados do conhecimento para denotar conceitos específicos (e.g. "buraco negro", "espaço-tempo"). Chamam-se expressões verbais as expressões que atuam como predicado no enunciado, geralmente constituídas por um verbo atuando como núcleo ao qual se associam outros elementos; o principal subtipo são as construções com verbo suporte (em inglês, *light verb constructions*), nas quais o complemento carrega o maior peso semântico da expressão, reservando ao verbo o papel de suporte de categorias gramaticais como tempo, modo, número e pessoa (e.g. "tomar banho", "ter sangue quente", "ficar a ver navios"). Por fim, têm-se expressões adverbiais e adjetivais quando estas atuam no enunciado como advérbios (e.g. "mais ou menos", "sem mais nem menos") e adjetivos (e.g. "de cabeça para baixo"), respectivamente.[75]

A classificação relativa ao método computacional utilizado para identificar MWE define três tipos que são ortogonais aos tipos definidos acima. São eles:

- expressões fixas, passíveis de serem detectadas por uma abordagem baseada em sequências de palavras e espaços (*word-with-spaces approach*); normalmente elas atuam como palavras funcionais (e.g. "em suma", "a fim de", "com respeito a"), contêm palavras estrangeiras ("ad hoc", "dejà vu") ou escapam às regras gramaticais convencionais;

- expressões idiomáticas, que podem ser de qualquer dos três tipos morfossintáticos e se caracterizam por possuir um sentido inteiramente não composicional, isto é, que se afasta completamente da soma dos significados literais de seus termos (ainda que possam ocorrer em sentido literal, cf. e.g. a possível ambiguidade da expressão "amarrar o burro na sombra"); por tais características, são extremamente difíceis de detectar automaticamente sem o auxílio de recursos semânticos;

- colocações verdadeiras, que correspondem às expressões institucionalizadas que constituem sintagmas completamente composicionais tanto do ponto de vista sintático quanto do ponto de vista semântico, mas ocorrem com frequência maior do que o esperado.

A importância das MWE para o vocabulário das línguas faz delas uma parte essencial dos dicionários, especialmente em dicionários bilíngues e para aprendizes, uma vez que seus usuários

[75] Cf. também Gross (1989) sobre as equivalências de cada tipo de MWE a determinada classe de palavras do ponto de vista de sua função no enunciado. A proposta de análise baseada no modelo da léxico-gramática deste autor (Gross, 1984) foi aplicada às expressões cristalizadas do português por Vale (2001).

podem não ter capacidade de reconhecê-las como unidades de significado e, consequentemente, ter dificuldades com sua compreensão (Atkins; Rundell, 2008).

Na prática lexicográfica, a informação sintagmática se apresenta em diferentes tipos e formas de apresentação. Welker (2004, p.137) reporta as quatro classes de informação sintagmática que, segundo Hausmann, comparecem nos dicionários: (1) especificações sobre construção (valência); (2) combinações lexicais binárias (colocações); (3) combinações lexicais fixas (expressões idiomáticas); (4) exemplificações (frases ou partes de frases). Por sua vez, Jacinto García (2015, p.154) mapeia as formas como esse conhecimento, considerado imprescindível para o domínio da língua, é veiculado pelos dicionários. O autor elenca cinco possibilidades: infralemas; exemplos lexicográficos; indicações de cotexto; indicações de contorno de definição; indicações de regime verbal.

Infralemas são uma forma comum de dicionários incluírem expressões idiomáticas, no sentido de "unidades pluriverbales que funcionalmente constituyen una sola unidad léxica o una oración completa" (Jacinto García, 2015, p.156).[76] Dado que a descrição de tais unidades pode envolver informações específicas a respeito de sua definição, tradução e, eventualmente, de suas propriedades gramaticais e pragmáticas, a posição de infralema permite organizar essas informações sem romper com o padrão de lematizar, como entradas principais, apenas unidades formadas por uma só palavra. No caso das colocações, o estatuto de infralema é menos utilizado, já que a colocação normalmente não implica em mudanças de significado nem de categoria gramatical.

O conceito de exemplo lexicográfico tem sido objeto de extenso debate no campo da metalexicografia. Para Jacinto García (2015, p.159), trata-se, antes de tudo, de uma categoria formal cujo conteúdo linguístico depende em grande medida do que o dicionarista concebeu como exemplo. Ainda assim, pode-se estabelecer uma "condição mínima, suficiente e necessária" para caracterizar o exemplo lexicográfico: o signo lemático deve aparecer em uso em um determinado cotexto (isto é, em combinação com outras palavras). Quanto à forma, o exemplo pode estar reduzido a um sintagma (unidade mínima de cotexto) ou constituir enunciados completos (orações com autonomia sintática empregadas em contexto reconhecível). Quanto à função, duas se distinguem como fundamentais (Jacinto García, 2015, p.148): (1) uma função linguístico-comunicativa, relacionada com a produção, a recepção ou a tradução de mensagens; (2) uma função extralinguística-cognitiva, encontrada em repertórios cujo objetivo é a transmissão de informações sobre determinados temas (incluindo temas linguísticos, como etimologia das palavras, por exemplo). A apresentação de informação sintagmática por meio de exemplos lexicográficos atende à primeira função, servindo para "ensinar ao usuário de que modo se usa a

[76] Cabe lembrar a distinção entre infralema e sublema feita por Hausmann e Werner (Jacinto García, 2015, p.155): embora ambos constituam entradas secundárias da microestrutura, infralemas representam palavras complexas que incluem o item lexical da entrada principal, ao passo que sublemas representam palavras complexas que não o incluem, como no caso de palavras derivadas apresentadas como subentradas de uma palavra primitiva.

expressão léxica representada pelo lema ou infralema: que preposições rege, com que palavras costuma se combinar, a que estilo pertence, etc." (Jacinto García, 2015, p.160).[77] Além disso, para cada classe de palavras há um tipo de exemplo que melhor atende aos propósitos informativos: para substantivos, os exemplos podem ensinar sobre possíveis combinações típicas (colocações); para adjetivos, exemplos indicam não só as colocações nominais, mas também as preposições que eventualmente regem (e.g. em português, "apto para", "desejoso de", "insatisfeito com"); advérbios podem aparecer na forma de colocações ("perdidamente apaixonado") ou locuções ("por cima de"); quanto aos exemplos de verbos, devem mostrar os argumentos que selecionam, de modo que ilustrem não apenas seu significado, mas também suas propriedades sintagmáticas.

A indicação de cotexto é um tipo de informação sintagmática que se refere a certa restrição léxica que incide sobre a combinação argumental, com consequências sobre o significado do signo lemático (Jacinto García, 2015, p.162). Seu grau de condensação é maior que o do exemplo, já que apenas a base da colocação é indicada através de um componente não integrado na paráfrase semântica, ao passo que a indicação do colocado pode se dar através de itens lexicais ou de categorias semânticas. O primeiro tipo poderia ser ilustrado por um verbete que apresentasse opções de tradução para o português do verbo latino *dare* ('dar') conforme o item lexical que lhe serve de argumento (e.g. *dare operam* 'prestar atenção'; + *dare poenam* 'sofrer castigo', etc.). O segundo tipo se refere ao que Hausmann e Werner denominam categorizadores de cotexto (Jacinto García, 2015, p.163), que normalmente correspondem a uma classe léxica representada por um hiperônimo, isto é, por uma palavra com um significado genérico o suficiente para englobar um conjunto de palavras concretas. Nesse sentido, seria possível distinguir duas acepções do verbo latino *uenire* utilizando os seguintes categorizadores: quando acompanhado de *in* + LUGAR, *uenire* significa 'chegar (em algum lugar)'; quando acompanhado de *in* + SENTIMENTO, significa 'provocar (tal sentimento)'. Na medida em que indicam ideias gerais por meio de metalinguagem, os categorizadores são mais bem compreendidos quando acompanhados de exemplos lexicográficos (e.g. *uenire in portum* 'chegar ao porto' vs. *uenire in odium* 'provocar o ódio'). Além disso, uma vez que certas palavras podem adquirir valor distinto conforme o argumento que as acompanha (caso típico dos verbos), indicações de cotexto podem ser usadas como marcadores distintivos das acepções de uma palavra polissêmica.

O contorno da definição é uma variante da indicação de cotexto, cuja particularidade é ser integrada à paráfrase semântica, em vez de enunciada de forma isolada. Jacinto García (2015) distingue dois modos de expressar o contorno da definição em um dicionário: como contorno condensado e como contorno não condensado (fórmula restritiva). O primeiro tipo consiste em indicar, entre colchetes ou parênteses ou por meio de tipografia especial, o tipo de argumento de uma palavra de modo integrado à sua definição; por exemplo, Ferreira (1966, p.130) utiliza

[77] Há pesquisas que sugerem que o significado e o uso sintático do léxico é compreendido e assimilado melhor através de enunciados naturais do que através de instruções metalinguísticas, fazendo com que exemplos sejam particularmente recomendados para dicionários didáticos (Jacinto García, 2015, p.160).

informação entre parêntese para precisar o significado do verbo latino *censeo* como "emitir uma opinião (numa assembleia)", "fazer o recenseamento (dos eleitores, dos contribuintes)" ou "declarar (os bens)". O segundo tipo consiste em fornecer uma explanação sobre um ou mais argumentos antes da definição ou tradução; por exemplo, seria possível distinguir duas acepções do verbo *peto* utilizando um par de definições explicativas como 'Dirigindo-se a uma pessoa, pedir algo. Dirigindo-se a um lugar, invadir'; também se incluem neste grupo as fórmulas iniciadas por "diz-se de…" ou "aplica-se a…" para indicar com que tipo de palavras certos modificadores costumam ser empregados.

As indicações de regime verbal, por fim, são a forma mais compacta de indicar os componentes da estrutura argumental de verbos. Elas se baseiam num processo que Rey-Debove denominou neutralização, que consiste na "simplificação de um exemplo autêntico, reduzindo-o a sua mera estrutura argumental" (Jacinto García, 2015, p.165). Geralmente o verbo assume uma forma não marcada (normalmente o infinitivo ou a terceira pessoa) e vem acompanhado de argumentos representados por categorias semânticas abstratas (e.g. PESSOA, ANIMAL, LÍQUIDO) ou então por indicações das propriedades morfológicas dos argumentos, especialmente em dicionários de línguas flexivas como o latim (e.g. indicações de caso morfológico como "cum accus.", "cum dat."). Uma forma de neutralização eficaz nesse sentido é a pronominalização do argumento, capaz de indicar não apenas a morfologia de caso, mas também o tipo semântico do argumento; por exemplo, um padrão de complementação como ""aliquem / aliqua re"" traz implícita a informação morfológica ACUSATIVO / ABLATIVO bem como a informação semântica +HUMANO / -ANIMADO, respectivamente.

As cinco formas de apresentar a informação sintagmática, vistas acima, possuem diferentes graus de lexicalização, em ordem decrescente: de um lado, combinações lexicais formatadas como infralemas, exemplos ou indicações de contexto lexical; de outro, as representações abstratas de cotexto por hiperônimos e de estruturas argumentais por categorias sintáticas. Todas podem servir à distinção de acepções e empregos especiais, relacionados tanto à variação de cotexto lexical quanto à variação de estrutura argumental. Dicionários baseados em modelos sintáticos-semânticos podem conjugar diferentes estratégias de apresentação da informação, desde a caracterização sintática e semântica dos argumentos até a abonação da análise por meio de exemplos extraídos de *corpus*, incluindo a indicação de expressões idiomáticas. Veja-se, por exemplo, Borba (1992, s.v.):

> DESCRUZAR – I. Indica ação-processo com sujeito **agente** e com complemento expresso por nome **concreto não animado** no plural. Significa *separar, desmanchando o cruzamento, fazer deixar de estar cruzado*: *Padre Bulhões descruzou as pernas (MAP, 30)* […] II. **Expressões: descruzar os braços** = pôr-se a trabalhar, agir: *Era preciso agir, descruzar os braços (DI, 56)*.

Dicionários tradicionais de latim podem incluir informação sobre as propriedades sintagmáticas dos lemas, mas geralmente falham em estabelecer uma associação clara delas com cada unidade lexical (isto é, cada acepção). Em alguns casos, tais propriedades são veiculadas por

meio de exemplos fornecidos em bloco na seção destinada à fraseologia, sem qualquer indicação metalinguística nem distinção entre sintagmas livres, colocações, expressões idiomáticas ou frases feitas; em outros, é fornecida uma lista completa e indistinta dos regimes associados a certo lema verbal na seção destinada à observações diversas (Lima, 1995).

Segundo (Welker, 2004, p.137), a estratégia de deixar a cargo do usuário a identificação e extração da informação sintagmática parece desaconselhada, já que o usuário nem sempre tem a competência de inferir a que casos as construções registradas se aplicam. Além disso, ao nosso ver, tal habilidade será necessariamente mobilizada para a compreensão do texto real, em que a palavra problemática foi encontrada. Portanto, parece contraditório que um dicionário, cujo propósito é oferecer solução para um problema de compreensão, o faça criando um novo problema de compreensão ao utilizar formas implícitas de indicar as propriedades sintático-semânticas associadas aos significados.

Estudos consideram a dificuldade de aprendizes iniciantes obterem informação diretamente de inferências, especialmente quando se trata de extrair, dos exemplos e abonações fornecidos pelo dicionário tradicional, informações a respeito das possibilidades de combinação das palavras (em especial dos verbos) e sua relação significativa, uma vez que essa operação requer um domínio do idioma que o aprendiz ainda não desenvolveu (Laufer, 1992; Zavaglia; Silva, 2006). Nesse sentido, Laufer (1992) observa que a compreensão de palavras novas é mais efetiva quando exemplos ilustrativos são acompanhados de descrição linguística. Para McGee (2012), é preciso introduzir o estudante no aprendizado por observação de dados antes de se empregar uma abordagem plenamente dedutiva baseada em concordâncias extraídas de *corpus* (Johns, 1991), e sugere o uso de dicionários de colocação como uma ponte entre as duas abordagens.

O percurso de consulta formulado por Bogaards (1996) prevê uma operação de comparação de contextos para se determinar qual acepção resolve do modo mais acertado o problema de compreensão gerado pelo item lexical. No caso do latim, a interpretação necessita apoiar-se antes de tudo em determinantes linguísticos. A informação mais básica a respeito do contexto original que o usuário pode utilizar na consulta ao dicionário é a análise morfológica dos elementos que compõem o enunciado: afinal, a própria lematização só é possível pela identificação de categorias morfológicas como caso, gênero, número de substantivos e adjetivos, ou modo, tempo, número e pessoa dos verbos. No entanto, sabe-se que as regras gerais de análise morfológica fornecidas pela gramática tendem a ser insuficientes para a (re)elaboração do sentido dos enunciados. Professores de latim estão habituados com certas expressões da dificuldade por parte dos alunos, do tipo "fiz a análise de todas as palavras, mas não consigo 'montar' a frase em português". Por isso, torna-se especialmente importante para o usuário contar com uma análise linguística que lhe forneça o máximo de informações gramaticais sobre formas e estruturas em sua relação com a significação, uma vez que, "como conjunto de regras que rege o encadeamento das formas de conteúdo na sucessão do discurso, a sintaxe tem um caráter conceptual, o que significa que cada combinatória de formas produz um determinado sentido" (Fiorin, 1992, p.22). Nas palavras de Baños Baños (2003, p.9), "la descripción del componente léxico de una

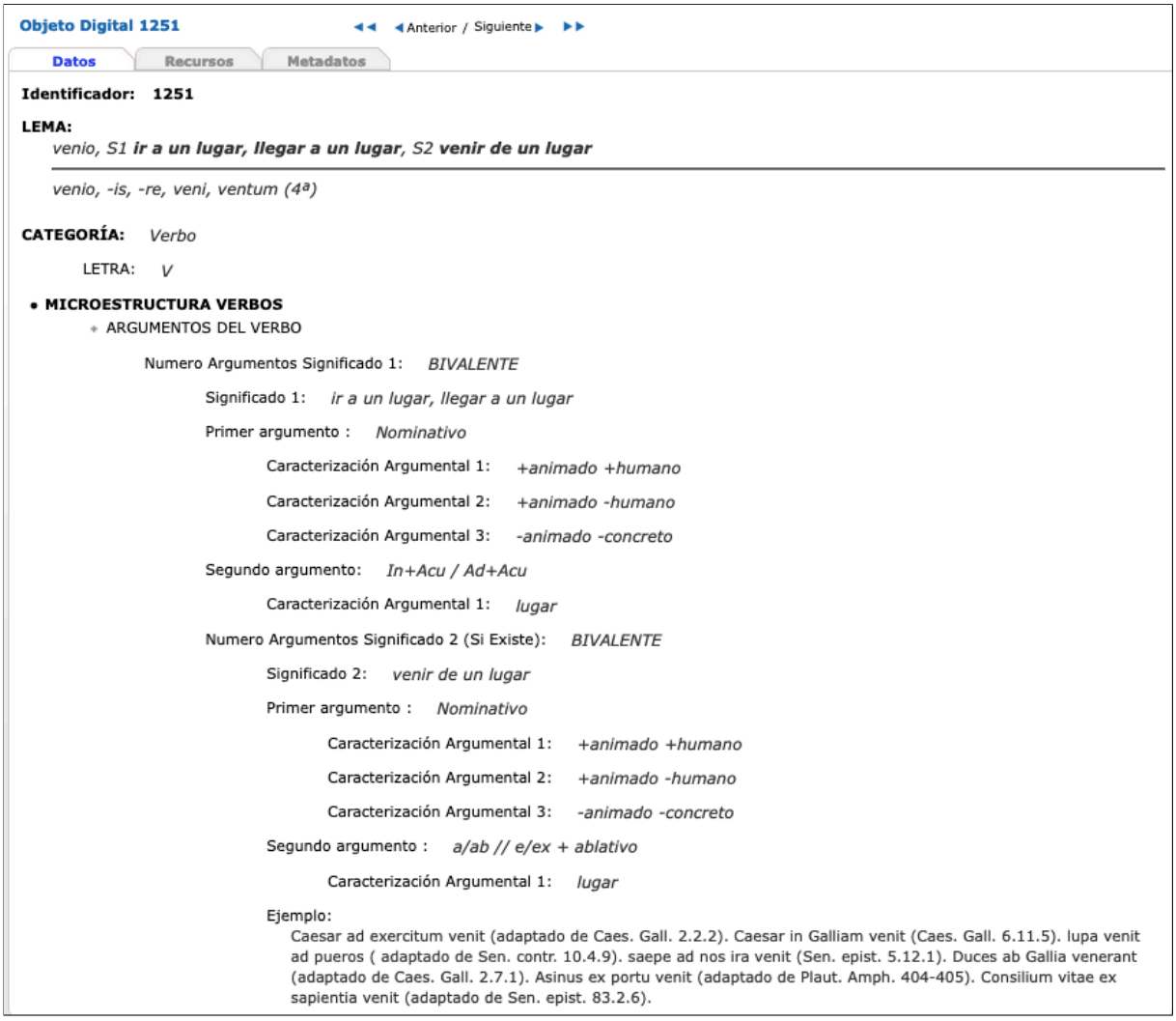
lengua deberá incorporar, junto con las indicaciones relativas al significado de cada término, las características propias de los complementos que pueden y, a veces, deben ir asociados sintácticamente a tal término". O autor inclui nessas características o caráter idiossincrático de certas estruturas de complementação (ou regime), em que as relações sintagmáticas entre itens lexicais não podem ser derivadas de regras gramaticais superiores e que está estreitamente associado ao conteúdo semântico dos mesmos itens (Baños Baños, 2003).

No âmbito da lexicografia do latim, esse método é pouco explorado. Uma experiência singular nesse sentido é o dicionário de verbos de Márquez Cruz e Fernández-Pampillón (2021), que classifica os significados de acordo com a valência dos lemas tanto do ponto de vista quantitativo (monovalente, bivalente e trivalente), quanto do ponto de vista qualitativo, conforme os traços semânticos dos substantivos que preenchem os argumentos: pessoas (+ANIMADO +HUMANO); plantas e animais (+ANIMADO –HUMANO); entidades concretas, perceptíveis pelos sentidos (–ANIMADO +DEFINIDO); entidades abstratas (–ANIMADO –DEFINIDO); e, por fim, lugar. Embora os autores utilizem recursos computacionais para automatizar o processo de seleção da nomenclatura, o trabalho de descrição dos significados, seleção de exemplos e elaboração dos verbetes é feito de modo "predominantemente manual" (Márquez Cruz, 2017, p.366), o que torna sua abrangência bastante limitada no que diz respeito à quantidade de significados associados a cada estrutura argumental. Um exemplo evidente é o verbete relativo ao lema verbal *uenio*, que descreve apenas dois tipos de argumentos preposicionados relativos a LUGAR (destino e origem) (ver Figura 3), contra os seis marcos predicativos identificados por Cabrillana (1997). Além disso, a forma de apresentação das informações sobre valência, argumentos, significados e exemplos é homogênea do ponto de vista tipográfico, o que pode gerar dificuldades de identificação de informações por parte do usuário.

Uma proposta inovadora de caracterização sintático-semântica de significados poderia tomar como base a variedade de exemplos lexicográficos encontrados nos dicionários tradicionais de latim. Hanks (2013a, p.428) chama atenção para a utilidade que a massa de evidências fornecidas pelos grandes dicionários históricos na seção dedicada à fraseologia pode ter nas tentativas de descrever o significado das palavras em contexto. No caso dos dicionários bilíngues, há a vantagem adicional de que grande parte das frases citadas como exemplo lexicográfico estão acompanhadas de tradução, oferecendo a um só tempo uma amostra de combinações típicas e especiais e o significado particular a cada construção. Com isso, são capazes de suprir ao mesmo tempo a necessidade de informação contrastiva, que constitui uma das maiores dificuldades com relação à construção de dicionários bilíngues exclusivamente a partir de fontes primárias (*corpus*).

Por fim, no que se refere à identificação e classificação das expressões multipalavras, há alguma informação disponível em gramáticas do latim de base funcionalista (Baños Baños, 2009; Pinkster, 2015) e em trabalhos específicos sobre o léxico de certos campos conceituais (e.g. o mapeamento do vocabulário sexual por Adams (1990), que inclui certo número de expressões) ou voltados para o estudo das metáforas, cuja função estruturante de certos domínios permite

Figura 3 – Captura de tela com o verbete *uenio* no *Diccionario Didáctico Digital de Latín* (Márquez Cruz; Fernández-Pampillón, 2021).



Fonte: <http://repositorios.fdi.ucm.es/DiccionarioDidacticoLatin/view/cm_view_virtual_object.php?idov=1251>.

explicar os padrões recorrentes em um conjunto de expressões tradicionalmente classificadas como idiomáticas (e.g. o estudo de Short (2011) a respeito das expressões relativas à atividade mental sistematicamente estruturadas em latim em torno de verbos indicativos de movimento físico no espaço). Porém, um repertório abrangente especializado nesse tipo de item ainda está por ser feito.

2.3.9 Informações sobre uso

Um último grupo de informações diz respeito a variações e especificidades observadas no uso de certos itens lexicais.

Variações podem ter caráter espacial (quando o uso varia de uma região para outra) ou social (quando se relacionam aos "diferentes registros utilizados pelas pessoas nas diferentes situações da vida social" (Borba, 2003, p.315)). Segundo este autor, dicionários costumam fornecer esse tipo de informação por meio de uma série de rótulos especiais, ainda que de forma

irregular, chamados marcas de uso.

A informação sobre o caráter marcado de um determinado lexema, isto é, "que foge, sob certos aspectos, ao uso normal da língua de uma comunidade linguística" (Welker, 2004, p.130), conta com diversas tipologias, que normalmente correspondem à diversidade de marcas encontradas nos dicionários. Uma das mais abrangentes é a de Hausmann, que distingue onze tipos de marcas (Welker, 2004, p.131):

- diacrônicas (por exemplo, *antiquado, envelhecido, neologismo*)
- diatópicas (aplicadas a acepções restritas a certas regiões ou países)
- diaintegrativas (usadas para assinalar estrangeirismos)
- diamediais (diferenciam entre as linguagens oral e escrita)
- diastráticas (por exemplo, *chulo, familiar, coloquial, elevado*)
- diafásicas (diferenciam entre as linguagens formal e informal)
- diatextuais (assinalam que o lexema – ou acepção – é restrito a determinado gênero textual; por exemplo, *poético, literário, jornalístico*)
- diatécnicas (informam que a acepção pertence a uma linguagem técnica, a um tecnoleto)
- diafrequentes (em geral: *raro, muito raro*)
- diaevaluativas (mostram que o falante, ao usar o lexema, revela certa atitude; por exemplo, *pejorativo, eufemismo*)
- dianormativas (indicam que o uso de certa acepção – ou lexema – é errado pelas normas da língua padrão).

Atkins e Rundell (2008, p.230) computam entre as marcas de uso (*linguistic labels*) caracterizações de significado (e.g. literal, figurado, por extensão), na medida em que se apresentam nos dicionários na mesma posição tipográfica dos demais indicadores de significado (cujo caráter inclusivo pode abranger até mesmo informações sintáticas como colocados). Welker (2004, p.136) chama a atenção para essa confusão, sugerindo que aquelas caracterizações devam ser entendidas mais propriamente como operadores semânticos ou indicadores de transferência semântica.

Além das variações, alguns itens lexicais só cumprem sua função comunicativa na situação de discurso (Borba, 2007) e exigem, portanto, uma explicação do significado quanto às suas propriedades pragmáticas. Estão nesse grupo as interjeições, as expressões de saudação ou imprecação, as exclamações, as blasfêmias, os insultos, bem como as frases feitas (provérbios e máximas), que compartilham a característica comum de constituírem reações vocais do interlocutor a algo ocorrido ou relatado, geralmente traduzindo sentimentos ou emoções como surpresa, alegria, pesar, desejo, etc.

Dicionários tradicionais de latim fornecem esporadicamente esses tipos de informação, em geral restringindo-se as indicações sobre variação a eventuais marcas de tipo diacrônico, diatextual ou diatécnico, bem como indicações pragmáticas do valor das interjeições, provavelmente porque a relação de equivalência costuma ser pouco esclarecedora e, portanto, seu significado, ou melhor, seu efeito pragmático, precisa ser explicitado.

Diante disso, ganha importância a possibilidade de se extrair dados quantitativos sobre o uso das palavras diretamente dos *corpora* de textos. O cálculo da frequência, associado a informações contextuais como autor, título, gênero discursivo ou registro eventualmente fornecidas pelos metadados do *corpus*, permitiria emoldurar cada uso da palavra a um contexto sócio-discursivo bem como calcular sua maior ou menor dispersão (Atkins; Rundell, 2008, p.150).

2.3.10 Ordenação das acepções

Welker (2004) faz um levantamento dos critérios elencados por diferentes lexicógrafos para a ordenação das acepções. Entre os mais citados estão: o critério histórico ou cronológico, que organiza as acepções conforme as mudanças semânticas da palavra ao longo do tempo; o critério da frequência de uso, que depende de estatísticas específicas sobre a ocorrência das palavras em cada acepção determinada; o critério lógico, em que "parte-se de um significado original e organizam-se as acepções conforme se trate de especialização do significado (marcado no dicionário por *especialmente*), expansão do significado (*por extensão*), sentido figurado ou uso analógico" (p.128); e o critério distribucional ou sintático, que se baseia na distribuição sintática do lema em cada uma de suas acepções. Diante de diferentes possibilidades, o autor conclui que não existe solução ideal, seja porque faltam conhecimentos seguros para qualquer das formas de distinção, quer porque nem sempre os interesses dos usuários coincidem com os dos lexicógrafos. Em outras palavras, todos os critérios são teoricamente justificáveis, mas a questão é saber qual é mais vantajoso para guiar o usuário até a acepção compatível com seu problema de compreensão.

O critério de frequência é defendido por Wharton (2011), pelo fato de que significados mais frequentes são menos dependentes de contexto e, portanto, teriam um acesso mais imediato na mente do falante. Esse tipo de ordenação é normalmente utilizado em dicionários para aprendizes, tendo em vista o menor número de acepções que o usuário terá de percorrer até achar a que lhe interessa (Welker, 2004).

O critério sintático é considerado particularmente efetivo para os verbos, como utilizado por Borba (1992). Ainda que alguns autores reforcem o lugar privilegiado da semântica em lexicografia, com base na expectativa de que a motivação básica para a consulta ao dicionário seja a compreensão do significado[78], o critério para a discriminação do significado não nos parece uma questão secundária. De fato, o significado é o objetivo justamente porque é desconhecido; um usuário sem base para decidir qual o significado corresponde ao contexto que ele está tentando compreender tende a tomar decisões aleatórias ou intuitivas, ambas igualmente propensas a equívocos.

Baños Baños (2003, p.10) observa, em dicionários de línguas clássicas como latim e grego, que "las informaciones sobre las posibilidades de complementación […] no aparecen organiza-

[78] Por exemplo, (Biderman, 2001, p.50–51) afirma que "em se tratando de lexicografia, a semântica deve comandar e a sintaxe vir em segundo lugar mesmo no caso do verbo, […] porque a primeira coisa que um consulente quer saber é qual o significado da palavra".

das según criterios sistemáticos o asentados en una teoría gramatical sólida, sino en función de las diferentes acepciones de significado que se reconocen, generalmente de forma intuitiva, en cada término". Nesse sentido, a distinção de significados por critérios sintáticos seria especialmente apropriada em dicionários de latim para guiar o usuário pelas diversas acepções, na medida em que a informação morfossintática é praticamente o único dado relativamente seguro que o usuário traz da análise do contexto para o qual busca explicação ao acessar o dicionário.

2.4 CONCLUSÃO DO CAPÍTULO

Neste capítulo, buscou-se embasamento para um projeto de dicionário bilíngue latim-português a partir de fundamentos linguísticos e metalexicográficos.

A definição do latim como língua histórica condiciona a situação típica de uso desse tipo de dicionário como sendo a recepção escrita e, portanto, sua função principal de auxiliar na compreensão das realizações linguísticas autênticas e, eventualmente, em sua tradução.

O foco na recepção escrita permitiu definir, com base em critérios gráficos de segmentação, o tipo de item lexical que deve compor a nomenclatura: apenas lexias simples, isto é, palavras; informações sobre afixos, abreviaturas e expressões são incluídas junto à descrição fornecida pelos verbetes.

O conteúdo e a forma de estruturar a descrição lexicográfica foram discutidos buscando relacionar os elementos típicos que compõem a microestrutura dos dicionários em geral à sua relevância específica para guiar os usuários de dicionários de latim. As principais decisões nesse sentido foram:

- a manutenção do critério tradicional de lematização;
- a necessidade de informação explícita sobre classes gramaticais, formas flexionadas e formas variantes;
- a obrigatoriedade de informação contrastiva, expressa na forma de equivalentes, que podem servir tanto como índice de significado para a compreensão quanto como sugestões concretas para a tradução; e
- a importância de informações paradigmáticas e, principalmente, sintagmáticas para auxiliar na determinação e diferenciação de acepções, considerando que a informação prévia extraída do texto pelo usuário costuma ser as propriedades morfossintáticas dos elementos do enunciado.

Também foi discutida a viabilidade de se incluir todas essas informações em um único dicionário. De um lado, a capacidade quase ilimitada de armazenamento oferecida pelo formato eletrônico permite a inclusão de qualquer informação considerada relevante. De outro, a necessidade de informação estruturada, legível por máquina, é pre-condição para o processamento

automatizado e consistente da base de dados em verbetes destinados à visualização humana. Logo, a inclusão de cada tipo de informação no dicionário depende, acima de tudo, da disponibilidade dessa informação nas fontes utilizadas para a composição da base de dados.

Diante da impossibilidade de se utilizar métodos subjetivos e intuitivos, dada a inexistência de falantes nativos, e da limitação dos *corpora* atuais quanto a anotações de caráter semântico e, especialmente, contrastivo, argumentou-se pela utilização de fontes secundárias de informação (isto é, outros dicionários). O próximo capítulo procura identificar, a partir de um levantamento histórico-bibliográfico da lexicografia latina, possíveis candidatos a integrar uma base de dados voltada para a produção de um dicionário eletrônico latim-português.

3 PANORAMA HISTÓRICO DA LEXICOGRAFIA DO LATIM

O estudo da história dos dicionários pode ser feito utilizando-se diversas abordagens. Hartmann (2001) enumera sete perspectivas possíveis:

- abordagem antropológica-universal, voltada para o desenvolvimento tecnológico ligado às obras de referência;
- abordagem diacrônico-linguística, voltada para o desenvolvimento das línguas e o registro dessas mudanças nas obras de referência;
- abordagem histórico-cultural, voltada para a associação entre princípios lexicográficos e tradições intelectuais;
- abordagem específica por gênero, voltada para o desenvolvimento de diferentes tipos de dicionários e suas finalidades específicas;
- abordagem genealógica, voltada para as relações de parentesco entre obras de referência;
- abordagem pessoal-biográfica, voltada para as formas de trabalho e as contribuições de cada lexicógrafo para o campo;
- abordagem historiográfica, voltada para o desenvolvimento de teorias (meta)lexicográficas ao longo do tempo.

O caráter panorâmico do histórico da lexicografia do latim apresentado neste capítulo não permite contemplar, para cada repertório, todas essas perspectivas. Procura-se aqui oferecer uma visão de conjunto do desenvolvimento contínuo da lexicografia através dos vinte séculos em que ela foi praticada com respeito ao latim. Logo, as perspectivas acima servem principalmente de guia de observação.

Hanks (2013b) chama a atenção para o impacto que o desenvolvimento de novas tecnologias de suporte para escrita teve sobre a lexicografia. Logo, o desenvolvimento tecnológico serve de critério para dividir o *continuum* histórico em três fases:

1. Lexicografia manuscrita, que abarca um período de quinze séculos, dos primeiros repertórios produzidos na Antiguidade latina até o final da Idade Média;
2. Lexicografia impressa, que abrange um período que vai do final do século XV até o final do século XX; e
3. Lexicografia computacional, que tem início na segunda metade do século XX e se estende até a contemporaneidade.

Dentro de cada fase, os repertórios lexicais têm seu conteúdo descrito considerando, na medida do possível, as condições sócio-culturais de sua produção, o posicionamento e os vínculos de cada compilador em relação a uma tradição intelectual e/ou pedagógica (eventualmente expressos em prefácio), bem como os indícios de uma provável dependência genealógica de trabalhos anteriores, identificados pela comparação de verbetes.

Evidentemente, recebem maior destaque os dicionários impressos latim-português, por seu potencial de servir como fonte de informação para uma base de dados bilíngue. Todavia, um retorno aos primeiros tempos parece importante pelo fato de que o latim clássico foi a língua da alfabetização e do letramento por mais de 1500 anos, sem interrupção, nos territórios onde os vernáculos europeus iriam se desenvolver. Logo, muitas características observadas nos dicionários bilíngues da tradição impressa são heranças dessa tradição.

3.1 REPERTÓRIOS MANUSCRITOS

Repertórios lexicais são utilizados como instrumentos pedagógicos desde tempo imemoriais (Boisson; Kirtchuk; Béjoint, 1991). Se as práticas orais de ensino de vocabulário por meio de listas lexicais temáticas, testemunhadas por etnólogos em sociedades ágrafas bilíngues, permitem pressupor que o uso de léxicos estruturados para a aquisição de segunda língua antecede a própria invenção da escrita, são os repertórios lexicais produzidos por quase todas as culturas que adotaram a escrita que constituem testemunhos materiais da disseminação desse instrumento-guia das práticas de alfabetização e letramento. No primeiro caso, é provável que repertórios lexicais tenham sido produzidos em Roma desde o desenvolvimento do alfabeto latino, por volta do século VII a.C. (Britannica, 2024), porém sem deixar vestígios. No segundo, a origem é bem conhecida: a cultura grega helenística, responsável pelo letramento da sociedade romana. Esta, inspirada pelo arcabouço teórico fornecido pela filosofia, desenvolveu não apenas diversas técnicas de estudo da língua, mas também um sistema pedagógico que demarcou simultaneamente a formação intelectual grega e romana, sem solução de continuidade até a época de Isidoro de Sevilha, quando tem início um novo padrão cultural que vai predominar na Idade Média (Moreschini; Norelli, 2000). As formas de descrever e registrar o léxico apresentadas a seguir fazem parte dessa formação.

Antes de observá-las, deve-se saber que os testemunhos de repertórios lexicais manuscritos produzidos na Antiguidade e na Idade Média são extremamente diversos, tanto no conteúdo quanto na forma de apresentação, o que faz da elaboração de um panorama uma tarefa complexa. Adverte-se, portanto, que o critério cronológico utilizado na apresentação a seguir não reflete necessariamente uma evolução linear. No caso do período da Antiguidade, em particular, a existência de certo tipo de repertório em determinado período quase nunca marca o momento de sua invenção; sua sobrevivência mascara uma enorme quantidade de textos que não conhecemos e pode ser resultado tanto da valorização de seu conteúdo pelas gerações posteriores – razão pela qual ele permaneceu sendo republicado (entenda-se, copiado à mão) – quanto de sua

rejeição, quando seu manuscrito foi utilizado como suporte material para outros textos e o acaso o transformou em um palimpsesto.

3.1.1 Antiguidade latina

A lexicografia em Roma se desenvolve no contexto de uma cultura letrada de base helenística (Cavallo; Chartier, 2004), que a sociedade romana adota, não sem conflitos, a partir do século II a.C.[1]

Essa cultura se fundava no cultivo da linguagem e dos valores transmitidos pelo patrimônio literário deixado pelos grandes autores gregos da época clássica, preservado na forma de livros manuscritos nas grandes bibliotecas e transmitido através das escolas (Holtz, 1981). Nas primeiras, a discrepância entre cópias de um mesmo texto leva ao desenvolvimento de técnicas de estudo filológico, voltadas para a determinação do que seria a versão correta, autêntica, tendo como subprodutos a produção de inventários e catálogos não apenas das obras disponíveis mas também dos recursos linguísticos empregados pelos autores: de um lado, as regras gerais de uso da língua são repertoriadas nas 'artes gramaticais' (Baratin; Desbordes, 1984); de outro, palavras que se distanciavam do uso corrente (em grego, *glossas*), por motivos diacrônicos ou dialetais, levam à produção de repertórios lexicais (Holtz, 1996). As escolas, por sua vez, oferecem uma "formação literária e cultural […] através da leitura e exegese de um programa específico de autores considerados clássicos" (Moreschini; Norelli, 2000, p.28). O gramático, até então um alfabetizador, passa a ser o professor responsável pelo letramento, utilizando-se das técnicas desenvolvidas pelos filólogos para ensinar gramática, entendida como "a perícia no que o mais das vezes está dito nos poetas e também nos prosadores" (Santos, 2007, p.154). Entre as atividades relacionadas ao aprendizado gramatical está "a exposição, em língua corrente, das palavras estranhas (*glossas*) e das histórias" (entenda-se, a mitologia) (Santos, 2007, p.154), ao lado da cópia manuscrita livre de erros, da leitura técnica do texto (em voz alta, em estilo declamatório) e da análise da linguagem com base nos princípios da analogia e da etimologia. Nessa escola está o alicerce de uma cultura letrada, propensa à imitação e emulação dos textos considerados clássicos, favorecendo a formação de cânones (Moreschini; Norelli, 2000, p.28).

Um marco da consolidação desse processo é a institucionalização da escola de retórica em Roma, na segunda metade do século I d.C., que teve Quintiliano como seu primeiro professor oficial (Pereira, 2005). Sua obra *Institutio oratoria* trata da formação do orador e pretende abarcar todas as etapas necessárias para a educação do cidadão segundo o ideal romano do *uir bonus dicendi peritus*, isto é, "um homem bom e hábil na arte de discursar". Nesse caminho, os estudos

[1] Momigliano (1991) destaca a excepcionalidade do caso romano, única das sociedades em que o partido favorável à adoção dessa "alternativa cultural" de origem grega venceu a disputa política. Cabe destacar também a excepcionalidade da posição romana nesse processo, um povo militarmente vencedor que se reinventou culturalmente com base em formas culturais do vencido. Essa aparente contradição foi imortalizada nos versos de Horácio (Horatius Flaccus, 1959, Ep. 2.1.155–6, tradução própria): "uma vez cativa, a Grécia cativou o feroz conquistador e introduziu as artes no Lácio rural" (no original: *Graecia capta ferum victorem cepit et artes intulit agresti Latio*).

de gramática desempenham uma dupla função: "a arte de falar corretamente e a explicação dos poetas" (Pereira, 2005, p.85). A primeira envolve outras qualidades (ou virtudes) do discurso além da correção (na fala e na escrita), como a precisão no uso do vocabulário e a elegância de estilo. A segunda fornece modelos de uso da linguagem (com base na noção de *auctoritas*, 'autoridade'), bem como os elementos de uma cultura literária compartilhada pela sociedade letrada. Os diversos repertórios lexicais produzidos nesse contexto parecem responder a essas funções.

O mais antigo repertório lexical latino preservado é o *De significatu verborum* ('Sobre o significado das palavras'), de Sexto Pompeu Festo (séc. II d.C.). Apesar de fragmentário, permite que se tenha uma visão ampla das preocupações da lexicografia latina desses tempos pioneiros, uma vez que menciona nominalmente como fontes os repertórios produzidos por três gerações de filólogos anteriores, como Lúcio Élio Estilão (atuante por volta de 100 a.C.), Marco Terêncio Varrão (116–27 a.C.) e Marco Vérrio Flaco (60 a.C.–20 d.C.). A obra consiste em uma compilação eclética de vocabulário, organizado alfabeticamente, acompanhada de comentários descritivos variados sobre o significado de palavras incomuns, pertencentes a domínios especializados do conhecimento ou empregadas de modo atípico, tanto no nível do significante quanto do significado, incluindo formas antiquadas, termos técnicos, empréstimos linguísticos, entre outros. Lhommé (2011, p.33–35) identifica alguns padrões de conteúdo explicativo associados a cada tipo de lema, alguns destinados a se fixarem como fórmulas metalinguísticas usadas na lexicografia manuscrita:[2]

a) definições são fornecidas eventualmente para palavras do vocabulário geral e frequentemente para termos técnicos;[3]

b) sinônimos ou explicações (normalmente introduzidas pela preposição latina *pro*, 'no lugar/sentido de') acompanham palavras incomuns ou empregadas de modo atípico por autores antigos;[4]

c) estrangeirismos são explicados por um equivalente ou uma definição do significado em latim (eventualmente introduzida pela locução *id est*, 'isto é');[5]

d) significados podem ser justificados com base em etimologia (introduzida por uma conjunção explicativa como *quod, quia, uelut*, 'porque' ou 'como que', ou por uma locução que designa origem, como *ab eo quod*, 'pelo fato de que');[6]

---

2 Os exemplos fornecidos em nota foram extraídos da edição de Lindsay (1913); todas as traduções são próprias.

3 Por exemplo: "METONYMIA est tropos, cum ab eo quod continet, significatur id, quod continetur" ('Metonímia é o tropo pelo qual o que é contido é indicado pelo que contém'). "OCCASIO, opportunitas temporis casu quodam provenientis est" ('Ocasião é a conveniência do momento fruto de certo acaso').

4 Por exemplo: "OBSTINATO, obfirmato, perseveranti, ut tenere possit" ('Obstinado: firme, perseverante, a ponto de poder suportar.'). "MAGISTRARE pro regere" ('*Magistrare* no sentido de dirigir'). "ME pro mihi dicebant antiqui" ('*Me* [acusativo] os antigos diziam no lugar de *mihi* [dativo]').

5 Por exemplo: "METAPHORAM quam graeci vocant, nos tralationem, id est domo mutuatum verbum" ('Metáfora os gregos chamam ao que nós chamamos *tralatio* ('transferência'), isto é, uma palavra emprestada para outra casa'). "NECTAR graece significat deorum potionem" ('Néctar em grego significa poção dos deuses').

6 Por exemplo: "MAMERCUS praenomen Oscum est ab eo, quod hi Martem Mamertem appellant" ('Mamerco é um prenome Osco, com origem no fato de que eles chamam Marte de Mamerte'). "OGYGIA MOENIA […] significat Thebas, quia eam urbem Ogygus condidisse traditur" ('Muralha ogígia […] significa Tebas, porque se conta

e) palavras atípicas do ponto de vista formal recebem um comentário relativo ao aspecto desviante (ortográfico ou gramatical);[7]

f) palavras caídas em desuso são marcadas por advérbios como *olim* ('antes') e *nunc* ('agora');[8]

g) palavras empregadas em sentido figurado são contrastados com o sentido próprio marcado pelo advérbio *proprie* ('propriamente');[9]

h) palavras de uso popular são marcadas pelo advérbio *vulgo* ('popularmente');[10]

i) palavras quase sinônimas recebem uma explicação que ressalta as diferenças específicas (*differentiae*);[11] e

j) palavras designando cidades, festas, instituições, a explicação se dá por enumeração de espécies.

Tais explicações são acompanhadas quase sempre de abonações constituídas por citações de autores clássicos ou de filólogos anteriores.

Nos séculos seguintes, os repertórios tendem a se tornar menos ecléticos, com cada tópico organizado em livros ou seções específicas. O maior exemplo dessa organização é a *Compendiosa doctrina* de Nônio Marcelo, a grande obra lexicográfica legada pela Antiguidade latina. Composta no século IV d.C., é uma coletânea que trata dos diversos tópicos da lexicologia antiga de modo separado, em vinte livros, todos fartos em citações de autores usadas como abonação. O texto está preservado integralmente (Lindsay, 1903), o que permite uma sumarização precisa de seu conteúdo. Cinco livros tratam de questões semânticas, a saber: do significado próprio das palavras (livro 1), das palavras empregadas de modo inusitado por autores antigos (livro 2), das palavras com múltiplos significados (livro 4), das diferenças entre palavras semelhantes na forma (quase-homônimos) ou no significado (quase-sinônimos) (livro 5); e do uso metafórico de palavras atestado nos autores antigos (livro 6). Outros sete livros tratam de questões gramaticais, normalmente ligadas à variação encontrada nos autores: substantivos que variam em gênero (livro 3) ou em paradigma flexional (livro 8), verbos que apresentam variação de paradigma (livros

que essa cidade foi fundada por Ogigo'). "OPPORTUNE dicitur ab eo, quod navigantibus maxime utiles optatique sunt portus" ('Oportunamente se diz pelo fato de que, para os navegadores, a coisa mais útil e desejada são os portos').

7 Por exemplo: "POLET, pollet: quia nondum geminabant antiqui consonantis" ('*Polet, pollet* [tem poder]: porque os antigos ainda não duplicavam consoantes'). "SPECUS feminino genere pronuntiabant antiqui, ut metus et nepos" ('*Specus* [espelho] os antigos proferiam no gênero feminino, assim como *metus* [medo] e *nepos* [neto]')

8 Por exemplo: "OPS antiqui dicebant, quem nunc opulentum" ('*Ops* os antigos chamavam quem hoje [chamamos] de opulento').

9 Por exemplo: "TORRERI a torro deductum, proprie significat siccare atque arefacere; sed usurpatum est iam pro eo, quod sit igne urere" ('*Torreri* é derivado de *torrus* [tição] e significa propriamente secar; mas é empregado agora no sentido de queimar com fogo').

10 Por exemplo: "RIVUS vulgo appellatur tenuis fluor aquae" ('*Rivus* [rio] é como o povo chama um tênue fluxo de água').

11 Por exemplo: "Inter OBSIDIONALEM et CIVICAM hoc interesse: quod altera singularis salutis signum est; altera diversorum civium servatorum" ('Entre [as coroas] obsidional e cívica há esta diferença: a primeira é signo de uma salvação particular; a segunda, de diversos cidadãos terem sido preservados').

7 e 10), uso irregular de casos e números (livro 9), formação irregular de advérbios (livro 11) e outras questões discutidas pelos gramáticos (livro 12). Por fim, os livro de 13 a 20 trazem repertórios temáticos com nomes de embarcações, vestimentas, recipientes, calçados, cores, alimentos e bebidas, armas e relações de parentesco, respectivamente. Apenas os livros 2-4 têm os lemas agrupados em ordem alfabética pela primeira letra. Mais importante é a menção constante à autoridade antiga como uma espécie de permissão para o uso: "TRANSGRESSVS ['passagem'] se diz propriamente em referência ao passo [*gradu*] dos pés; mas a autoridade da antiguidade permite um uso mais livre da expressão…" (Lindsay, 1903, p.726),[12] o que indica que alguns desses repertórios podiam ser utilizados como parâmetro tanto para atividades de compreensão quanto para atividades de produção discursiva (Zetzel, 2018).

A finalidade produtiva é mais evidente nos repertórios lexicais motivados por questões de correção gramatical, que passam a circular como apêndices dos manuais de gramática.[13] Entre os séculos IV e VI, encontram-se repertórios voltados para a correção fonético-ortográfica (pronúncia e escrita) e sintática (concordância e regência). O *Appendix Probi*, conhecido por testemunhar as formas do latim vulgar na seção sobre ortografia – e.g. "uernaculus, non uernaclus; articulus, non articlus" (Stok, 1997, p.197) –, também contém seções destinadas à combinação correta de palavras (em latim, *locutiones*), organizadas pelo caso sintático do complemento:

> locução com o caso genitivo…: *similis illius* ['semelhante dele'], *auidus honoris* ['ávido por um cargo'], *neglegens amicorum* ['negligente com os amigos'] […]
> locução com o caso dativo: *similis Mercurio* [semelhante a Mercúrio], *fido uirtuti* ['confiante na coragem'], *suadeo ciui* ['convenço ao cidadão'] (Stok, 1997, p.196)

O recurso ao emprego de pronomes como representantes genéricos da forma do complemento, observada em *similis illius*, acima, é empregado de forma sistemática em um repertório do século IV especializado em padrões de combinação sintática: os *Exempla elocutionum* ('Exemplos de frases'), de Arusiano Méssio. A forma de lematizar as expressões, posicionando a palavra regente no início da frase, aliada ao agrupamento por ordem alfabética, resulta em um compêndio de mil frases contendo diversas possibilidades de combinação de uma mesma palavra abonadas pelos grandes autores. Como exemplo, vejam-se as primeiras linhas do texto, dedicadas ao verbo *abundare* ('existir em abundância'):

> ABUNDANS ILLIUS REI. Virgilius in buco. 'Nivei quam lactis abundans'.
> ABUNDANS ILLA RE. Cicero pro Cornelio libro II 'Quis tam abundans copiis?'
> ABUNDAT ILLA RE. Cicero pro Cluentio 'Mulier abundat audacia'.

[12] No original: "TRANSGRESSVS a pedum gradu proprie dicitur; sed auctoritas vetustatis liberiorem loquendi usum dedit.

[13] Segundo um gramático grego anônimo (Uhlig, 1965, t.3, p.170, tradução própria): "Os trabalhos [gramaticais] consistem em produzir tecnicamente um discurso metrificado ou prosaico. O objetivo é não errar, nem em relação a uma única palavra, nem em relação a mais de uma." No original grego: Ἔργα τὸ τὸν ἔμμετρον καὶ τὸν πεζὸν λόγον τεχνᾶσθαι. Τέλος τὸ μὴ ἁμαρτάνειν μήτε περὶ μίαν λέξιν μήτε περὶ πλείονας.

> ABUNDE EST HUIUS REI. Virg. Aen. VII ‘Terrorum et fraudis abunde est’. – (Di Stefano, 2011, p.5)[14]

Um repertório como esse demonstra o quanto o estudo dos autores clássicos como Cícero e Virgílio servia de fundamento não só para o uso correto e autorizado, mas também como modelo de uso elegante da língua latina. Diversos outros repertórios parecem servir a essa mesma função, observando o valor estético tanto da palavra individualmente quanto das combinações.[15] Um conjunto de repertórios dessa natureza compõe o quinto livro do manual de gramática de Flávio Carísio, professor em Constantinopla no século IV. A elegância nas palavras individuais é contemplada por uma seção sobre “sinônimos de Cícero dispostos em ordem alfabética” (*synonyma Ciceronis ordine litterarum conposita*), estruturada como séries de sinônimos iniciada por um lema aparentemente de significado geral seguido por palavras semanticamente similares, por exemplo:

> AMOR. ardor. aestus. alacritas. cupido. desiderium.
> AMICITIA. societas. commercium. hospitium. pax. foedus. ius iurandum. fides. religio. indutiae. leges. condiciones. pactiones. exceptiones. conuentiones. transactiones.
> AMOR. adfectus. affectio. caritas. pietas.
> AB EXORDIO. ab initio. a primordio. ab origine. a principio. ab infantia. ab incunabulis. a rudimentis.
> (Barwick, 1964, p.412)

A elegância nas combinações é tratada em duas seções, uma contendo “frases diversas com o mesmo significado” (*glossulae multifariae idem significantes*), outra “sobre a latinidade” (*de latinitate*). A primeira repete a estrutura da seção sobre sinônimos, apresentando séries de fraseados equivalentes em sentido (aparentemente no nível pragmático, área de atuação da retórica).[16] A segunda é uma lista de combinações organizadas alfabeticamente, aparentemente sem relação entre si; itens como *ardet inuidia* (‘arde de inveja’), *bono animo est* (‘está de boa vontade’), *crimen capitale* (‘acusação capital’), *fragilis memoria* (‘memória fraca’) indicam que se trata de uma coleção de combinações típicas do Latim (daí o título da seção, “sobre a latinidade”).

---

[14] Tradução própria: “ABUNDANTE DE ALGO [genitivo]. Virgílio nas bucólicas ‘quão abundante de níveo leite”’. | ABUNDANTE EM ALGO [ablativo]. Cícero a favor de Cornélio, livro 2 ‘Quem tão abundante em tropas?’. | ABUNDA EM ALGO [ablativo]. Cícero a favor de Cluêncio ‘A mulher abunda em audácia’. | HAVER UMA ABUNDÂNCIA DE ALGO [genitivo]. Virgílio Eneida 7 ‘há uma abundância de terrores e engano”’.

[15] A consciência dessa distinção é atestada na seção sobre retórica da obra de Marciano Capela: “O método de falar com elegância tem duas partes: uma, pela qual o brilho se manifesta nas palavras individuais; outra, que faz com que o valor da elocução seja garantido pela beleza da combinação delas” (Kopp, 1836, p.509, tradução própria). No original: “Huius rei [= ornate dicere] duplex ratio est: una, qua in singulis uerbis lumen appareat; altera, ut dignitas eloquendi copulationis ipsius decori seruetur”).

[16] Por exemplo: “LIVRA-ME DE PREOCUPAÇÃO. Dá-me tranquilidade. Faz com que eu esteja tranquilo. Tira-me o medo. Alivia o temor. Faz com que eu não tema. Diminui minha preocupação. Faz com que eu não tenha medo […] Alivia-me a ansiedade” (Barwick, 1964, p.409, tradução própria). No original: “LIBERA ME SOLLICITUDINE. securitatem mihi praesta. fac ut securus sim. aufer mihi metum. timorem eleua. fac ne formidem. curam mihi minue. fac ut caream metu […] anxietatem mihi releua.”

O caráter idiomático dessas combinações fica evidente no fato de que muitas delas se encontram em um repertório bilíngue grego-latim anexo à gramática de Dositeu. Pares de expressões como "εὐθυμεῖ bono animo est","καταμαρτυρηθεὶς ἄνθρωπος detestabilis homo" ou "κροτητικὴ αἴτησις plausibilis petitio" (Tolkiehn, 1913, p.89) são indicativos de que essas obras serviam a um público de aprendizes de línguas estrangeiras, em ambas as direções (Zetzel, 2018, p.111).

Nesse contexto, aliás, a comparação entre os modos de expressão grego e latino dava ensejo a diversos repertórios contrastivos de "particularidades" (*idiomata*, em grego), definidas por Carísio como "tudo aquilo que proferimos segundo o nosso costume e não segundo os gregos" (Barwick, 1964, p.379, tradução própria).[17] Na base desse conceito estava a percepção do latim como uma língua dependente da língua grega (senão até como uma espécie de dialeto grego) (Zetzel, 2018, p.115)[18], de modo que diferenças gramaticais observadas em formulações equivalentes são consideradas como particularidades do latim em relação ao sistema do grego.

Os principais tipos de *idiomata* se definem em torno de duas categorias gramaticais: gênero e caso. No primeiro grupo, podem ser nominais (e.g. "furor μανία", equivalentes de 'loucura', masculino em latim, feminino em grego) ou verbais[19] (e.g. "medeor θεραπεύω", equivalentes de 'curar', depoente em latim, ativo em grego). No segundo grupo, são mapeadas diferenças quanto ao caso do complemento de verbos, nomes, adjetivos e particípios equivalentes, por exemplo: "pudet me amoris ἐρυθριῶ τὸν ἔρωτα ['o amor me deixa envergonhado']" (Keil, 1857, p.566), construído em latim com genitivo e, em grego, com acusativo. Além desses repertórios, propriamente gramaticais, também há registros de glossários bilíngues relativos a termos técnicos e eruditos, bem como a itens do vocabulário cotidiano e familiar (Kramer, 1996).

O apanhado incompleto de repertórios produzidos na Antiguidade latina, apresentado acima, deve ser suficiente para dar dimensão da diversidade de abordagens e critérios de seleção de fatos linguísticos ligados ao léxico que interessavam às escolas de gramática e retórica. Seu conteúdo servirá de matéria-prima para a composição de uma segunda leva de repertórios manuscritos, desta vez ambientada em um contexto intelectual diferente, caracterizado sobretudo pela substituição dos textos utilizados como fontes de autoridade na língua: os clássicos da literatura latina antiga saem de cena; em seu lugar, entra o cânone de textos cristãos. A próxima seção procura trata dessa nova fase da lexicografia latina manuscrita.

---

17 No original: "idiomata […] sunt omnia quae pro nostro more efferimus et non secundum Graecos".

18 Essa relação é expressa por Carísio nos seguintes termos (Barwick, 1964, p.380, tradução própria): "Tendo em vista que a Língua Latina parece derivar do grego, as coisas que são ditas fora do costume dos gregos, quer por licença dos antigos, quer por um caráter próprio do latim, são chamadas *idiomata*". (No original: "cum ab omni sermone Graeco Latina lingua pendere uideatur, quaedam inueniuntur uel licentia ab antiquis uel proprietate linguae Latinae dicta praeter consuetudinem Graecorum, quae idiomata appellantur.") A menção aos antigos se explica porque, em alguns casos, o costume dos antigos foi modificado por influência grega, como atesta o gramático Diomedes (Keil, 1857, p.320): "há *idiomata* que os antigos enunciaram usando o caso acusativo, mas que nós enunciamos usando o dativo, conforme os gregos" (no original: "[sunt] idiomata quae ueteres quidem per accusatiuum extulerunt, nos autem per datiuum secundum Graecos efferimus").

19 Lembre-se de que, na gramática antiga, também os verbos eram classificados em gêneros (ativo, passivo, neutro, depoente, comum).

3.1.2 Idade Média europeia

Estabelecer um limite preciso à passagem contínua do tempo é uma dificuldade conhecida do discurso histórico: "a história não se detém, as datas são pontos de suspensão…" (Moreschini; Norelli, 2000, p.11). Tradicionalmente, a data que marca o fim da Antiguidade e o início da Idade Média é a de um evento de caráter político: a queda do último imperador romano, em 476 d.C. Com a destituição do poder constituído, o território se divide em reinos, cada um com sua própria mescla de cultura latinizada e não latinizada, a exemplo dos reinos Visigótico, na península Ibérica, e Franco, no noroeste da Europa. A conversão desses reinos para o catolicismo viabiliza a manutenção do latim como língua do letramento nos mesmos moldes da Antiguidade (i.e. baseado nas técnicas de estudo filológico transmitidas pela gramática), porém com a progressiva substituição do repertório textual greco-latino pelo repertório judaico-cristão, especialmente o conjunto de textos canônicos sagrados conhecido por *Biblia* ('livros', em grego).

Um personagem-chave dessa transição é Isidoro, bispo de Sevilha no início do século VII, durante o reinado dos visigodos. Nascido em uma família hispano-romana e educado nas artes liberais provavelmente no mosteiro onde seu irmão era abade,[20] Isidoro foi responsável pelo primeiro esforço de preservação da cultura letrada na Europa após o fim do Império Romano, tendo atuado politicamente, como presidente do Conselho de Toledo em 633 d.C., para aprovar um decreto que estabelecia centros educacionais em todas as catedrais da Espanha. Sua grande obra, as *Etymologiae* ('Etimologias'), pode ser vista como reflexo desse esforço:[21] um compêndio enciclopédico do conhecimento herdado da Antiguidade que "preserva o antigo e prenuncia o novo" (Moreschini; Norelli, 2000, p.11), constituindo-se simultaneamente em "resumo, conclusão e transição para novas formas de interpretação" (Zetzel, 2018, p.117) que serão as referências de toda a cultura letrada a partir de então.

Em certo sentido, Isidoro "usa listas de palavras para organizar o mundo" (Zetzel, 2018, p.117); em outrtas palavras, as *Etimologias* são como um "imenso glossário nominal organizado não por ordem alfabética, mas etimológica e temática" (Holtz, 1996, p.20). Para além de uma distribuição meramente temática, Barney et al. (2006, p.20) identifica também uma organização pedagógica na divisão da obra em vinte livros. Os três primeiros são dedicados às artes liberais, em ordem de importância: gramática (livro 1); dialética e retórica (livro 2); matemáticas, isto é, aritmética, geometria, música e astronomia (livro 3). Em seguida, descreve as ciências da medicina, da jurisprudência e da teologia (livros 4 a 6), para então apresentar o "grande encadeamento dos seres", que começa com o mundo divino (livro 7), passa pelo mundo humano e animal (livros 8 a 12), para terminar com a ordem do universo e do mundo material (livros 13 a 20). Nesse grande sistema ontológico, o estudo da gramática ocupa a primeira posição não só por sua posição análoga no sistema pedagógico (como propedêutica para a retórica e a dialé-

[20] Leandro, irmão mais velho de Isidoro, era amigo pessoal do papa Gregório I e foi predecessor de Isidoro como bispo de Sevilha, tendo participação direta na conversão da família real visigótica do arianismo ao catolicismo, no final do século VI (Barney et al., 2006).

[21] Segundo Barney et al. (2006, p.18), Isidoro escreveu suas obras visando a classe dirigente do reino visigótico, a qual poderia patrocinar a difusão da educação liberal.

tica), mas também porque, no contexto intelectual católico, em que os erros de compreensão das escrituras são fontes de heresias, a leitura e explicação correta de textos escritos era percebida como habilidade fundamental.[22]

O livro sobre a gramática se estrutura como os manuais herdados da Antiguidade tardia: das letras às classes de palavras, passando pelas sílabas, noções de métrica, acentuação e sinais de pontuação usados na crítica textual; em seguida, das palavras aos tipos de discurso, passando pelos vícios e virtudes no uso da linguagem. Entre esses dois blocos, Isidoro acrescenta cinco capítulos que representam os princípios e a aplicação de seu método investigativo (Codoñer, 1996b). De um lado, três capítulos teóricos que tratam, respectivamente, da forma correta da palavra escrita (ortografia), de sua forma gramatical (analogia) e da relação entre forma e significado (etimologia). Do outro, dois capítulos práticos, que tratam das formas de descrever o significado das palavras, isto é, por meio de glosas e de diferenças.[23]

A etimologia, em especial, é a técnica que permite inferir, mediante o conhecimento da origem dos vocábulos, "a verdadeira força [significativa] de um verbo ou de um nome, […] cujo conhecimento frequentemente tem um uso necessário em sua interpretação" (Lindsay, 1911, liber I, §39, tradução própria).[24] Essa relação entre nome e realidade pode se dar a partir de um modo de ser (e.g. *rex* 'rei' é assim chamado porque deve agir *recte* 'corretamente'), de uma origem material (e.g. *homo* 'pessoa' tem origem em *humus* 'barro') ou de uma ideia contrária (e.g. *lutum* 'lama' vem de *lavare* 'lavar'); também há etimologias que se explicam por derivação (e.g. o adjetivo *prudens* 'prudente' deriva do substantivo *prudentia* 'prudência'), por um empréstimo linguístico (principalmente do grego) ou, por fim, devido a simples semelhança fonética. A técnica é aplicada ostensivamente no livro 10 das *Etimologias*, intitulado *De vocabulis* ('Sobre vocábulos'), que consiste em uma série alfabética de lemas adjetivos relativos a atributos humanos (boa parte com valor moral) explicados com base em um dos tipos de relação etimológica.

As glosas, por sua vez, são um tipo de explicação do significado baseado na relação de equivalência, no sentido de que uma palavra desconhecida é explicada por meio de uma (e não mais que uma) palavra conhecida; Isidoro exemplifica essa técnica com pares que seguem a fórmula "*x* é *y*": "contiscere est tacere" ('fazer silêncio é ficar quieto'), "terminus est finis" ('término é fim'), "populatus est vastatus" ('assolado é devastado') (Barney et al., 2006, p.55). Essa prática representa, para Holtz (1996), um alargamento da acepção do termo *glosa*, que deixa de designar a palavra rara encontrada nos autores antigos ou nos poetas e passa a abranger qualquer palavra da língua que requer uma explicação de significado, independentemente do

[22] Nesse sentido, Isidoro replica o entendimento, já manifesto por Agostinho, Jerônimo, Boécio e Cassiodoro, de que o domínio das técnicas pagãs é benéfico para o intelectual cristão, com as artes liberais sendo consideradas a propedêutica ideal para o estudo da Bíblia e da teologia (Barney et al., 2006, p.16).

[23] Glosas e diferenças são dois dentre quinze tipos de 'definição' que Isidoro extrai do tratado a esse respeito escrito por Mário Vitorino no século IV d.C. (Barney et al., 2006, p.22).

[24] No original: "Etymologia est origo vocabulorum, cum vis verbi vel nominis per interpretationem colligitur […] cuius cognitio saepe usum necessarium habet in interpretatione sua."

motivo. Assim, "a *glossa* se torna equivalente de *léxis*" (Holtz, 1996, p.19).[25]

As diferenças, por fim, são um tipo de definição baseada no contraste entre palavras com significados similares, cujas características específicas é preciso distinguir e definir. Dado o caráter enciclopédico das *Etimologias*, Isidoro apenas exemplifica o tópico com a diferença entre dois tipos de monarcas, o rei e o tirano: "o rei é moderado, o tirano é cruel" (Barney et al., 2006, p.55). Seu uso da técnica pode ser observado em uma obra sobre as *Diferenças*, publicada alguns anos antes, cuja divisão em dois livros é significativa: o primeiro trata de "diferenças entre palavras", o segundo de "diferenças entre coisas". O primeiro consiste em um repertório semelhante aos encontrados na Antiguidade, com o incremento de estar disposto em ordem alfabética; são 610 pares ou séries de palavras semelhantes pelo significado ou pela forma (especialmente formas prefixadas da mesma raiz), apresentadas segundo a fórmula "entre *x* et *y*" e diferenciadas por particularidades como:

> Entre PROFUNDO e ELEVADO, SUBLIME e ÁRDUO: *profundo* é relativo à parte mais alta ou mais baixa; *elevado*, apenas à mais alta; *sublime* é elevado com uma certa dignidade; *árduo* é elevado com uma certa dificuldade. (Arevalo, 1802, p.3)[26]

O segundo não é um repertório propriamente dito, mas um protótipo de enciclopédia, em que uma distinção entre conceitos dá origem a uma apresentação sucinta mas completa do tema tratado em cada um dos quarenta capítulos, sempre de interesse doutrinário. O livro começa com a distinção entre *Deus* e *Dominus* e entre a *Trindade* e a *Unidade* e termina com a diferença entre os oito principais vícios morais, em parte coincidentes com o que hoje se conhece como pecados capitais. Na diferença entre *eloquência* e *sabedoria* se encontra o princípio de divisão da obra: "Entre a sabedoria e a eloquência [eles] distinguem assim: […] a *eloquência* é a ciência das palavras, já a *sabedoria* é o conhecimento das coisas e o entendimento das causas" (Arevalo, 1802, p.109);[27]; em seguida, a equivalência entre o termo latino *sapientia* e o grego *philosophia* dá ensejo à explicação enciclopédica sobre as partes da filosofia antiga – quais sejam, física, lógica e ética – e as respectivas subdivisões, com destaque para a doutrina dos vícios e virtudes, pertinentes ao âmbito da última.

Isidoro cita uma grande variedade de autores e obras antigas como suas fontes, que incluem tanto compêndios técnicos quanto textos literários e filosóficos cristão e pagãos; Virgílio, por exemplo, é quase tão citado quanto a Bíblia (Barney et al., 2006, p.15). A boa recepção de sua obra tornou ele próprio um *auctor*, fonte de conhecimento para os séculos futuros (Holtz, 1996), principal obra de referência para toda matéria ligada às artes e às ciências (Shaw, 1997),

[25] (Holtz, 1996) destaca que, a partir desse momento, o foco passa a recair sobre o par formado pela palavra a explicar e a explicação, numa evolução de significado que estaria a meio caminho do significado atual de glosa nas línguas modernas, restrito ao segundo elemento do par.

[26] Tradução própria do original: "Inter ALTUM et EXCELSUM, SUBLIME et ARDUUM: altum, ex superiore parte et inferiori est; excelsum, ex superiori tantum; sublime excelsum est cum aliqua dignitate; arduum excelsum est cum aliqua difficultate."

[27] Tradução própria do original: "Inter SAPIENTIAM et ELOQUENTIAM ita distinguunt: […] eloquentia scientia est verborum; sapientia autem cognitio rerum et intellectus causarum".

exercendo influência direta em toda a tradição de léxicos e enciclopédias produzidos no contexto europeu até a invenção da imprensa (Barney et al., 2006, p.25). As disciplinas da derivação e da etimologia, em particular, terão papel fundamental na aprendizagem da gramática e no conhecimento da língua (Hamesse, 1996, p.xi).

Um segundo movimento político de instituição, ou melhor, de padronização da cultura letrada ocorre um século mais tarde, durante o reinado de Carlos Magno entre os francos. De fato, o período carolíngio é caracterizado por um esforço de universalização das convenções da escrita em meio a um contexto linguístico complexo, formado por falantes de diferentes dialetos românicos (a oeste do Reno) ou germânicos (a leste) (McKitterick, 1989). O uso do latim como língua administrativa e religiosa, portanto do letramento, era uma herança romana que desempenhava um papel importante na preservação da unidade linguística do reino. Por essa razão, o *status* do latim como meio primário de comunicação escrita foi não só mantido, mas reformado com base na restauração de formas clássicas e na imposição de uma grafia e de uma gramática uniformes.

Dois tipos de obras lexicográficas estão diretamente ligadas a esse processo. De um lado está o *De Orthographia* ('Sobre a grafia correta'), de Alcuíno de York, que se insere na tradição de obras que descrevem lemas selecionados por motivos formais, seja para apresentar formas corretas de escrita (do tipo "*x*, non *y*"), seja para distinguir lemas com formas escritas ou sonoridades semelhantes (nesse caso, absorvendo parte do conteúdo das diferenças de palavras). De outro, repertórios de glossas, voltadas para a descrição do significado com base na sinonímia ou na etimologia, quer coletados das anotações marginais dos manuscritos, quer reproduzidos das numerosas coleções de glosas anteriores; dentre as muitas compilações produzidas nas diversas bibliotecas existentes, destacam-se o resumo do *De verborum significatione* de Festo, feito por Paulo Diácono, e a maior delas, intitulada *Liber glossarum*, que reúne não apenas glosas explicativas sinonímicas, mas também inclui algumas breves definições e descrições enciclopédicas (Codoñer, 2016).

A formalização e correção da escrita do latim segundo os modelos clássicos teve como efeito colateral, por assim dizer, uma maior disparidade entre as formas escritas e faladas da língua nativa dos francos, até então ainda percebida como sendo o latim. Em outras palavras, é provável que esse processo tenha contribuído para alterar a percepção (e, consequentemente, o ensino) do latim de língua materna para língua estrangeira,[28] e implica consequentemente em novas atitudes e novos instrumentos didáticos. Os repertórios lexicais, parte integrante desse contexto, também passam por modificações, consideradas prototípicas dos dicionários como conhecemos.

A primeira mudança de paradigma ocorre no século XI, quando Papias, um clérigo da Lombardia, publica seu *Elementarium doctrinae erudimentum* ('Introdução básica à doutrina'). Num período caracterizado pelo aumento do número de pessoas que aprendiam a ler e escrever

[28] Para uma discussão aprofundada sobre o *status* percebido do latim como língua nativa (e não estrangeira) no período carolíngio, ver McKitterick (1989, pp.7–22).

em latim, acompanhando a difusão desse aprendizado para outras esferas da sociedade além da religiosa (comércio, política, vida social) (Shaw, 1997), ao mesmo tempo que o latim se torna definitivamente uma segunda língua com a consolidação das línguas vernáculas (Lazard, 2002), Papias parece perceber a necessidade de uma obra de referência lexical mais abrangente e de fácil consulta.

No prefácio, Papias insere sua obra na tradição dos glossários, mas se esforça por se destacar dela quando discute a natureza de sua "arte":

> Essa arte, outrora chamada *Glossarium*, porque consistia simplesmente na interpretação de palavras individuais, agora poderia se chamar mais apropriadamente de *elementarium doctrinae erudimentum*, devido ao acréscimo de definições, anotações de acordo com as regras [da gramática], bem como sentenças e muitas outras coisas do gênero (Daly; Daly, 1964, p.230, tradução própria).[29]

O próprio compilador explica de que anotações se trata: um asterisco para marcar as palavras que carecem de autoridade ou de conformidade com os princípios gramaticais (principalmente de derivação), informação sobre gênero, declinação e conjugação por meio de abreviaturas e formas flexionadas, indicação de sílaba longa ou breve quando necessário. Além disso, ressalta que a ordem alfabética abrange até a terceira letra da palavra e que inclui abonação das informações fornecidas pela indicação dos autores onde elas se encontram de forma abreviada.

Embora essa informação nova não seja inserida de forma sistemática, fazendo com que muitas entradas sejam idênticas às encontradas na obra de Isidoro ou no *Liber glossarum* (Shaw, 1997),[30] a obra de Papias é considerada um protótipo dos dicionários modernos por diversos motivos, quais sejam: a presença de indicações gramaticais para auxiliar o usuário (Codoñer Merino, 2002); o uso de sentenças "não apenas para definir o domínio semântico da palavras, mas também exemplificar seu emprego" (De Angelis, 1998, p.122); e, acima de tudo, a unificação em um único repertório de materiais antes distribuídos por repertórios distintos, isto é, ortográficos, de glosas, de diferenças e de etimologias.

Uma inovação em particular teve grande repercussão no futuro da lexicografia: a criação de verbetes dedicados à enumeração das possibilidades derivativas das palavras. Aqui, o princípio gramatical da derivação, que em Isidoro constava entre os tipos de explicação etimológica do significado de determinada palavra, é utilizado por Papias para concentrar sob uma única entrada uma série de palavras derivadas cujo significado pode ser deduzido a partir do conhecimento do termo primitivo. Tome-se como exemplo o início da entrada para o lema verbal *ago*:

---

[29] No original (grifo adicionado): "…[Haec ars] etsi olim, quia ad verbum et simpliciter unius alicuius dictionis retinebat interpretationem, *glossarium* vocaretur, iam vero, diffinitionibus et secundum regulas notationibus, sententiis quoque et multis id genus superadditis, altius atque aptius *elementarium doctrine erudimentum* nominari poterit."

[30] Vale notar que também esse esforço de concentração não é sistemático e contrasta com a multiplicação de entradas para descrever diferentes aspectos do significado de um mesmo lema, típica dos repertórios anteriores. O lema *ager* ('campo'), por exemplo, é contemplado com uma série de dez entradas, cf. De Angelis (1978, p.143–5)

> AGO *-gis*: daí se deriva *agens* e *actus -ta -tum*; de *agentis*, *-tis* para *-dus*, se faz o particípio *agendus -da -dum*, e o nome em *-dum* significa 'coisa digna de ser feita'. Também do particípio *actus* [se deriva] *actus*, nome participial da quarta declinação, do qual [se deriva o adjetivo] *actualis actuale* e o advérbio *actualiter*; daí *actuarius -ria -rium* e *actualitas, actor, actrix, actio*. De *ago* ainda [se deriva o adjetivo] *agilis agile agilior agillimus*, do qual o advérbio *agiliter* e a forma *agilitas* […] (De Angelis, 1978, p.152, tradução própria)[31]

O verbete prossegue apresentando a rede de derivados e compostos de *ago* por meio de uma textualização baseada em conjunções como *unde* 'de onde' e *inde* 'daí' e locuções prepositivas como *a quo* 'do qual', sem ordem aparente e sem pretensão de exaustividade.[32]

Hunt (1991, p.371) relaciona a emergência de dicionários como repertórios lexicais distintos dos livros de glossas à incorporação, na construção de verbetes, do princípio derivacional já presente na gramática antiga e, até então, provavelmente utilizado como método oral de ensino do léxico. Nesse sentido, a obra de Papias marca o primeiro estágio de um processo que irá se consolidar no século seguinte, com a *disciplina derivationis* assumindo lugar central nas obras de Osberno de Gloucester e Hugutio de Pisa.

Osberno foi monge e professor de gramática no mosteiro de Gloucester, no sul da Inglaterra, em meados do século XII (Busdraghi et al., 1996, p.vii). Em seu *Liber derivationum* ('Livro das derivações'), ele utiliza os mesmos métodos lexicográficos vistos em Papias – a derivação e a glosa –, porém com a diferença de que não unifica os dois tipos de informação em um mesmo repertório, mas os apresenta separadamente em repertórios distintos, organizados por letra do alfabeto. Assim, para cada letra, uma primeira parte (chamada *derivationes*) concentra informações sobre a forma, partindo da que é considerada primitiva (tradicionalmente, a primeira pessoa singular do indicativo para verbos e o nominativo singular para nomes), acompanhada de sua classificação gramatical, para apresentar em seguida algumas séries de palavras derivadas e compostas. A segunda parte (chamada *repetitiones* 'repetições') concentra informações sobre o significado de um certo número de itens, geralmente utilizando palavras e expressões sinônimas, às vezes breves definições, ocasionalmente acompanhadas de citações de autores clássicos e medievais (Shaw, 1997).[33] A título de comparação, veja-se o início da entrada para o verbo *ago*:

> *Ago is egi actum ctu* verbo ativo, donde *actor ris*, e *actus a um*, e *actio onis*, donde o diminutivo *actiuncula le* e *actus ui*, e o verbo frequentativo *actito as*, donde os deverbais, e *activus va vum*, e *actualis actuale*, donde o adverbium

[31] No original: "ago -gis: inde derivatur agens actus -ta -tum; ab agentis -tis in -dus fit agendus -da -dum participium, et nomen -dum significat rem dignam agi. ab actus quoque participio actus participiale nomen quartae declinationis, a quo hic et haec actualis et hoc actuale et actualiter adverbium; inde actuarius -ria -rium et actualitas actor actrix actio. ab ago etiam hic et haec agilis et hoc agile agilior et agillimus, a quo agiliter adverbium et agilitas forma […]"

[32] Após citar mais de 130 palavras, o autor conclui com "e outras que algum estudioso possa encontrar" (De Angelis, 1978, p.154, tradução própria)

[33] O mesmo esquema comparece em três manuscritos anônimos encontrados em Munique e datados do final do século XII, descritos por Lunardini (2009).

> *actualiter*, e o advérbio *actutum*, isto é, *cito* [‘rápido’] que encontrarás em Terêncio. (Busdraghi et al., 1996, p.30, tradução própria)[34]

O estilo é evidentemente muito semelhante àquele encontrado em Papias, porém há dois pontos importantes de se observar. O primeiro é a incorporação de três formas ao início do verbete – o perfeito *egi* e os supinos *actum, actu* –, criando uma espécie de lema composto que tende a se consolidar como formato tradicional de enunciação dos verbos; o mesmo vale para as outras classes de palavras, como se verifica nas citações de lemas no interior do verbete: substantivos são citados pelas formas (eventualmente abreviadas) de nominativo e genitivo singular (e.g. *actio -nis*); adjetivos, pelo nominativo singular de cada gênero (e.g. *activus va vum* e *actualis actuale*). O segundo é a inserção sistemática de informação gramatical classificatória para lemas verbais (e.g. “verbo ativo”), aparentemente inaugurando a prática lexicográfica de apresentar a classe de palavras logo após a enunciação do lema. Importa notar que nenhum desses modos de apresentação é utilizado no bloco voltado para as informações semânticas (*repetitiones*); aqui uma única forma serve de lema (nominativo singular para nomes, infinitivo para verbos), seguida imediatamente pela explicação do significado.

Para Hunt (1991, p.382), essa configuração faz do *liber derivationum* uma obra tanto para estudo e memorização do léxico quanto para consulta de palavras consideradas difíceis. Não por acaso, foi uma das poucas obras britânicas a ter uma circulação significativa na Europa, entrando para a corrente de obras didáticas principalmente após sua utilização por Hugutio de Pisa.

Hugutio foi um especialista em direito canônico que atuou como bispo de Ferrara no final do século XII. Sua obra *Derivationes* (ou *Magnae derivationes*, i.e. ‘Grandes derivações’) é estruturada inteiramente segundo o método da derivação, representando uma mudança crucial no modo de organizar o material lexical (Hunt, 1991, p.382). De fato, Hugutio retoma ostensivamente o material de Osberno, porém sem preservar a estrutura bipartite baseada na distinção entre derivação e glosa; ao contrário, procura unificar ambos os repertórios em um único instrumento (Shaw, 1997). Com isso, a informação semântica e as citações de autores, bem como outras informações suplementares sobre ortografia, pronúncia e gramática, são incorporadas aos verbetes, muitas vezes de forma inorgânica, resultando em uma obra pouco prática de ser usada para consulta (Hunt, 1991, p.382). Provas dessa dificuldade são os lemas inscritos pelos copistas à margem dos manuscritos, para viabilizar sua localização mais rápida. Ainda assim, foi rápida e amplamente copiado, sendo citado como fonte por lexicógrafos até pelo menos o século XV, bem como por escritores do nível de Dante, Petrarca e Bocaccio (Hunt, 1991, p.383).

Quase um século depois, mais precisamente em março de 1286, João de Gênova, o Balbo (em latim, Johannes Januensis de Balbis) publica a *Summa grammaticalis quae vocatur Catholicon*. Como o título indica, a obra reúne um compêndio de gramática e um amplo repertório

[34] No original: “Ago is egi actum ctu verbum activum, unde hic actor ris, et actus a um, et hec actio onis, unde hec actiuncula le diminutivum et hic actus ui, et actito as verbum frequentativum, unde verbalia, et activus va vum, et hic et hec actualis et hoc actuale, unde actualiter adverbium, et actutum adverbium .i. cito quod in Terentio reperies.”

lexical, o qual consiste essencialmente em uma versão revisada do trabalho de Hugutio.[35] Seu esforço de resolver as dificuldades de localização das palavras resultou na primeira obra lexicográfica a apresentar os lemas em ordem alfabética estrita (Sharpe, 1996, p.97). Já o fato de articular em uma única obra os dois instrumentos pedagógicos principais (gramática e repertório lexical) talvez seja a razão para que o *Catholicon* tenha sido uma das primeiras obras impressas da história, com a edição por meio de tipos móveis publicada em Mainz em 1460, possivelmente pelas mãos do próprio Guttenberg (Daly; Daly, 1964, p.237).

Repertórios lexicais de conteúdo complexo como os de Papias, Osberno, Hugutio e João de Gênova certamente exigiam um conhecimento razoável de latim para sua utilização, pressupondo um usuário já em estágio avançado de aprendizado. Por isso, cabe mencionar uma segunda tradição que se estabelece no contexto da Universidade de Paris a partir do século XII e é constituída de repertórios lexicais organizados por assunto e estruturados como texto em prosa, voltados para o ensino de palavras comuns do cotidiano para aprendizes iniciantes (Sharpe, 1996). Os pioneiros dessa segunda tradição são Adam do Petit-Pont, Alexandre Neckam e João de Garlândia.

Adam do Petit-Pont (ou Adam de Balsham) escreve, na primeira metade do século XII, um repertório lexicográfico em forma de carta intitulado *De utensilibus* ('Sobre os utensílios'), em que simula uma espécie de visita guiada a uma fazenda para descrever e nomear as coisas que compõem essa realidade. O plano vai do exterior para o interior, começando com os tipos de terrenos e pastagens, passando pelos animais e chegando aos edifícios, que enseja a enumeração dos diversos objetos típicos de cada cômodo (e.g. as armas expostas na torre, os instrumentos de escritório, os objetos da capela) e das construções adjacentes (e.g. os equipamentos e veículos do celeiro, dos estábulos, da vinícola), para então retornar à casa e descrever a cozinha, a sala de tecelagem, roupas e ornamentos, móveis, entre outros objetos. Segundo Hunt (1991, p.171), a influência de Isidoro é mais do que evidente.

Alexandre Neckam escreve seu *De nominibus utensilium* ('Sobre os nomes dos utensílios') na segunda metade do século XII. A obra é muito semelhante à anterior, porém restringe a matéria aos utensílios domésticos e equipamentos: começa na cozinha; passa ao armário e à dispensa, onde descreve apetrechos de viagem; em seguida, passa à mobília dos quartos, então retorna à preparação de alimentos e bebidas; depois descreve o modo de construção de castelos e as armas de defesa, daí desce à descrição do celeiro e dos pássaros, da tecelagem; saindo da vida rural, enumera os tipos de navios e termina com materiais ligados à escrita e à joalheria.

O terceiro autor dessa tradição de repertórios elementares é João de Garlândia, que compôs por volta de 1220 uma obra intitulada *Dictionarius*, no que se acredita ser a primeira ocorrência dessa palavra (Considine, 2022, p.6), talvez cunhada pelo próprio autor, que a explica no primeiro parágrafo da obra:

[35] No sumário que o autor fornece ao final da obra, o repertório é descrito como "a parte desta obra que trata da etimologia" (Balbus, 1460, s.p.), o que demonstra a persistência da conexão entre derivação e etimologia herdada de Isidoro.

> Este livrinho se chama *Dicionário* devido às palavras [*dictiones*] mais necessárias que qualquer estudante deve guardar, não numa caixa feita de madeira, mas no armarinho do coração, para que possa alcançar mais facilmente a construção da oração e a enunciação. Que ele saiba primeiro nomear as coisas comuns. Portanto, optei por iniciar pelas partes do corpo humano, desenvolvendo um repositório das coisas existentes (Hunt, 1991, p.195, tradução própria)[36]

Trata-se, portanto, de um repertório "de coisas" (*rerum*), o que sugere uma lista enciclopédica de nomes concretos, elaborado com o objetivo de promover confiança e fluência através do uso de palavras relativas à vida cotidiana (Hunt, 1991, p.195), sem qualquer menção ao conhecimento gramatical clássico demonstrado nas outras obras pedagógicas do autor.[37] As palavras são agrupadas por temas, iniciando pelas partes do corpo, como se viu, passando pelos instrumentos utilizados no comércio, armas e utensílios domésticos, itens clericais e obras litúrgicas, vestimentas e objetos usados na igreja, ocupações das mulheres; seguem-se listas de pássaros, peixes, animais domesticados, fauna e flora, árvores frutíferas, então os tipos de navios e de instrumentos musicais que os ricos têm em casa, para terminar com os monstros que habitam o Inferno.

Compilações temáticas nominais (*nominalia*) para uso escolar são encontradas em profusão nos séculos XIII e XIV, muitos deles iniciando pelas partes do corpo e incluindo nomes de plantas e animais, objetos domésticos e religiosos, armas e jogos, entre outros domínios (Lendinara, 2023).[38] Sua contraparte eram os *verbalia*, listas de verbos organizados alfabeticamente ou por conjugação, acompanhados de uma ou mais formas flexionadas. Em alguns casos, esses inventários recebiam glosas em língua vernácula, transformando-se assim em glossários bilíngues. Foi o que ocorreu ao único representante conhecido da lexicografia latina manuscrita em Portugal: o manuscrito alcobacense ALC-286.

3.1.3 O manuscrito alcobacense ALC-286

Após o intervalo imposto pelo domínio do árabe em quase toda a península Ibérica, o ensino de latim retorna ao território português com a fundação de escolas catedrais em Coimbra e Braga no final do século XI e de escolas monásticas nos mosteiros de Coimbra e Alcobaça em 1132 e 1153, respectivamente (Fernandes, 2012, p.325). O latim permanecia como língua franca do espaço europeu, sendo utilizada no letramento, na administração, na justiça, na liturgia e no comércio. Quase na mesma época, o vernáculo começava a ter *status* literário com os primeiros

[36] Dictionarius dicitur libellus iste a dictionibus magis necessariis, quas tenetur quilibet scolaris, non tantum in scrinio de lignis facto, sed in cordis armariolo retinere, ut ad faciliorem constructionem orationis et enunciationem possit pervenire. Primo sciat vulgaria nominare. Placet igitur a menbris humani corporis inchoare, rerum promptuarium evolvendo.

[37] Além do *Dictionarius*, João de Garlândia escreveu outras obras lexicográficas voltadas para o ensino de latim, incluindo uma versão em verso das derivações (*Unus omnium*), um livro de sinônimos (*Synonima*) e outro dedicado aos homônimos, intitulado *Equivoca* ('[palavras] ambíguas').

[38] Uma possível aplicação real dessa prática pode estar representada por alguns manuscritos do século XIV contendo "inventários domésticos […] cujo objetivo era realmente listar os equipamentos, as roupas e todo tipo de bens contidos na casa e na fazenda" (Lendinara, 2023, p.297, tradução própria do original: "Tracts on household accounts […] are de facto intended to list house and farm implements, clothing and all sort of goods in detail").

registros trovadorescos, cuja época de ouro, no século seguinte, coincide com o incremento do uso do português em documentos públicos e privados, culminando com sua oficialização como língua do reino por Dom Dinis I, em 1297.

Testemunhos materiais de textos escolares desse período são escassos. Boa parte se encontra no Fundo Alcobaça da Biblioteca Nacional de Lisboa (Amos, 1988–1990), que inclui cópias dos grandes repertórios lexicográficos apresentados acima, a saber: as *Etimologias* de Isidoro, datada de meados do século XII; o *Elementarium* de Papias (intitulado *Vocabularium latinum*), em três volumes, com data na folha de rosto indicando o ano de 1200;[39] o *Liber derivationum* de Hugutio de Pisa, de meados do século XIII; e as *Derivationes* de Osberno de Gloucester, do final do mesmo século.[40] De repertórios produzidos por compiladores portugueses, há notícias de apenas dois: um vocabulário bilíngue de nomes, organizado alfabeticamente, escrito por Frei Martinho de Arraiolos por volta de 1170, do qual não há vestígios (Verdelho, 1995, p.194); e um repertório anônimo de formas verbais latinas, datado do início do século XIII, que se vê transformado em vocabulário quando, no início do século seguinte, são inseridos equivalentes de tradução para a Língua Portuguesa à margem do texto, na forma de glosas.

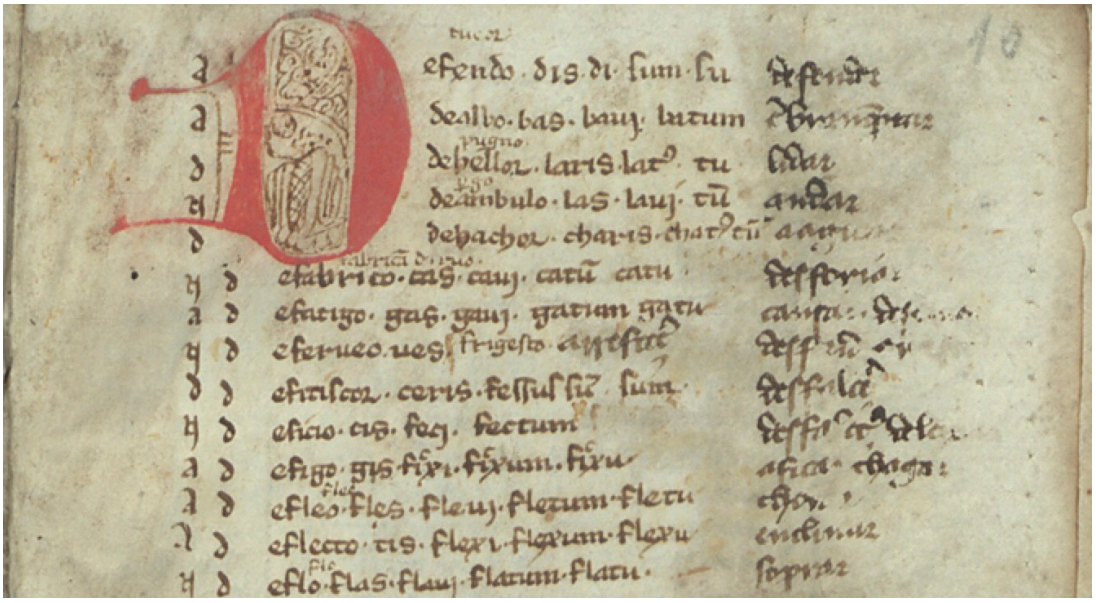
A Figura 4 reproduz o terço superior do folio 10r. A letra D capitular, em tinta vermelha, marca o início do conjunto de verbos com essa letra inicial. O espaço gráfico se organiza em três partes. À esquerda, duas colunas contendo uma letra: a primeira indica o 'gênero' do verbo, segundo a classificação da gramática antiga ('a' para verbo ativo, 'n' para neutro e 'd' para depoente); a segunda corresponde à inicial do lema, que orienta a consulta alfabética. Ao centro, um bloco composto por uma série relativamente homogênea de formas verbais, correspondentes às 1ª e 2ª pessoas do singular do presente do indicativo, à 1ª pessoa do singular do perfeito do indicativo e a duas formas de supino, no acusativo e no ablativo, respectivamente. À direita, uma coluna com os equivalentes de tradução portugueses, escritos em caligrafia e cor diversas, o que constitui indício de terem sido adicionados posteriormente. A transcrição parcial abaixo, feita com base na edição de Carter (1952), permite ver com mais clareza essa estrutura, que se repete pelos 2930 lemas que compõem o repertório:

a	D	efendo · dis · di · sum· su	*defender*
a	d	ealbo · bas · baui · batum	*embranqueitar*
d	d	ebellor · laris · latus · tu	*lidar*
…			
a	d	efleo · fles · fleui · fletum · fletu	*chorar*
a	d	eflecto · tis · flexi · flexum · flexu	*enclinar*
n	d	eflo · flas · flaui · flatum · flatu	*soprar*

[39] O terceiro volume tem anexados, ao final do texto de Papias, uma cópia do *Livro das traduções dos nomes hebraicos* de Jerônimo de Stridon e dos *Nomes das regiões e locais extraídos dos Atos dos Apóstolos*, de Beda. Ambos os tipos de itens lexicais (nomes hebraicos e topônimos) farão parte da nomenclatura dos dicionários impressos.

[40] Esse e outros itens possuem cópia digital disponível para acesso no site da Biblioteca Nacional de Portugal (<https://bndigital.bnportugal.gov.pt/>).

Figura 4 – Início da letra D do manuscrito ALC-286.



Fonte: <https://purl.pt/24439>.

Carter (1952, p.74) afirma que a compilação foi produzida para servir como obra de referência, provavelmente como "guia prático para o significado dos verbos latinos durante a leitura de manuscritos medievais"; as marcas de deterioração dos fólios tipicamente causadas pelo manuseio indicam que foi usado intensamente. Considerando apenas o conteúdo latino original, antes do acréscimo dos equivalentes, Verdelho (1995, p.99) supõe que o manuscrito tenha sido "elaborado, em princípio, exclusivamente como um dicionário gramatical", cujo "objetivo primário seria o ensino da flexão verbal". Para além dessa função mais evidente, a semelhança entre a estrutura das entradas do manuscrito 286 de Alcobaça e o início de cada verbete nas *Derivações* de Osberno, único dentre os grandes repertórios derivacionais que inclui as formas de supino e a classificação dos verbos por gênero (ativo, neutro, depoente), pode ser indício de que esse repertório tenha sido usado também para o ensino das formas derivadas.

No que se refere às glosas portuguesas, a edição permite perceber que elas foram adicionadas quase sistematicamente ao conjunto, com apenas 6% dos verbos desprovidos de equivalentes. Embora a imensa maioria seja constituída por um único equivalente (quase 85% dos casos), Verdelho (1995, p.203) identifica cinco formatos de explicação do significado em português:

a) ampliação paratáctica, com séries de equivalentes não necessariamente sinônimos, com muitos casos de polissemia ou alargamento da significação (e.g. *colo* "onrrar lavrar morar amar afeitar", *amplifico* "anchar acrescentar");

b) uso de compostos verbo-nominais, apoiados num verbo funcional ou auxiliar (e.g. *amico* "fazer amigo", *amicor* "ser amigo", *consulo* "dar / demandar conselho", *honorifico* "fazer onrra");

c) uso de compostos verbo-adverbiais (e.g. *amburo* "queimar derredor", *beo* "bem aventurar", *depetulor* "demandar desconvinhavilmente");

d) uso de locuções adverbiais articuladas por preposição ou conjunção (e.g. *actineo* "perteecer em linhagem", *agnio* "braadar conme anho", *hisco* "abrir com costringemento", *idiomatizo* "falar per linguagem";

e) recurso a frases casuais (e.g. *patrisso* "arremedar fazer feitos de padre", *scandalizo* "escarnecer dar de si mao exemplo".

Além de glosas portuguesas, o manuscrito também contém glosas em latim de igual diversidade, contemplando 27% dos verbos e podendo se referir tanto a aspectos semânticos quanto gramaticais. No primeiro caso, Verdelho (1995) menciona a existência de indicações sinonímicas (e.g. sobre a entrada "ambicio […] cubiçar" a glosa "cupio", lit. 'desejar'), paráfrases explicativas (e.g. sobre a entrada "doleo […] doer" a glosa "dolorem patior", lit. 'sinto dor'), indicações pragmáticas (e.g. sobre a entrada "uagio […] gemer" a glosa "proprie infantis clamantis est", lit. 'refere-se propriamente ao bebê que grita'), explicações etimológicas (e.g. a entrada para o lema *uerso* acompanhada da glosa "frequenter uertere", lit. 'verter muitas vezes', ou "deriuo […] derivar" glosado por "riuum a fonte traho", isto é, 'puxo o rio desde a fonte'). No segundo grupo estão indicações relativas à ortografia (e.g. "dedignor non dinor"), a certa particularidade flexional (e.g. "impersonale est" para o lema *oportet* ou "defectiuum" para o lema *inquio*), ao tipo de derivação (e.g. sobre *scisso* a glosa "inchoatiuum est scisco", lit. 'é o incoativo de *scisco*') ou, ainda, a marcas de uso como "non est in usu" ('não está em uso') ou "antiquum est" ('é antigo').

Em suma, técnicas como a glosa sinonímica em latim e em vernáculo, definições breves de cunho etimológico, a distinção de significados próprios e usos poéticos e antiquados, bem como o emprego da metalinguagem gramatical para descrever a forma e o uso das palavras latinas, são características persistentes no desenvolvimento da lexicografia latina em suporte manuscrito. O fato de o latim se tornar segunda língua é um fator-chave para explicar o desenvolvimento de um lema composto de formas básicas (para auxiliar na flexão e na derivação), bem como o surgimento de explicações em vernáculo (para tornar o significado mais acessível). Parafraseando Cormier (2010, p.133), pode-se dizer que, para o historiador da lexicografia, observar esse tipo de ancestral dos dicionários é um processo análogo à observação de sinais do *Big Bang* pelo astrofísico ou à observação de artefatos e indícios das primeiras comunidades humanas para o arqueólogo. De fato, é nesse contexto que surgirão os primeiros dicionários propriamente ditos, cuja criação é impulsionada pela tecnologia de tipos móveis introduzida na Europa por Gutemberg no século XV. A despeito da profunda mudança tecnológica, eles vão se construir seguindo de perto essa herança medieval, muitas vezes apenas reorganizando o espaço gráfico, sem alterações significativas nas técnicas de descrição ou no tipo de informação veiculada. A próxima seção traz os detalhes dessa passagem.

3.2 DICIONÁRIOS IMPRESSOS

É lugar-comum dizer que a invenção da imprensa marcou uma época, constituindo um dos símbolos de uma nova era, junto com a pólvora e a bússola (Burke, 2002). Esse sucesso costuma ser relacionado com seu enorme poder de difusão rápida de conhecimento em larga escala, muito superior ao processo lento e trabalhoso de produzir cópias manuscritas individualmente, em um contexto de aumento do número de homens e mulheres alfabetizados e, consequentemente, da demanda por livros. A relação com o letramento não é casual; as três obras provavelmente publicadas por João Gutemberg, o primeiro a utilizar tipos móveis na Europa, constituem os três elementos básicos da cultura letrada medieval: uma gramática elementar (*Ars Donati*, *c.* 1454), o conjunto de livros sagrados (a *Biblia*, em 1455) e um dicionário (o *Catholicon* de Balbo, em 1460). Como parte do universo letrado, a lexicografia europeia foi profundamente impactada por essa nova possibilidade de publicação e divulgação. Hanks (2012b) coloca a imprensa como um dos três fatores que, atuando em conjunto, participaram decisivamente do desenvolvimento da lexicografia impressa no século XVI, ao lado da redescoberta da literatura latina clássica e do pensamento questionador veiculado pela Reforma protestante.

3.2.1 Inventores de dicionário: Nebrija, Calepino, Estienne

De um ponto de vista cronológico, o surgimento da imprensa na Alemanha coincide com a disseminação na Itália do que ficou conhecido como Humanismo. Esse movimento de resgate do patrimônio literário e cultural da Antiguidade greco-latina, iniciado por Petrarca e Boccaccio no século anterior, está então em pleno processo de assimilação pelas classes dirigentes italianas como signos de prestígio, que incluem os valores morais e conhecimentos literários extraídos das leituras em seus processos educativos (Canfora, 1986).[41] A difusão destes pelas classes dirigentes de todo o continente europeu, no século seguinte, ampliou o horizonte intelectual e criou as bases para propostas de reforma das instituições, sendo especialmente marcante no âmbito do cristianismo.

Esse esforço de recriação do ambiente cultural também passava pela língua. O latim, usado desde há muito como língua franca e escolar, necessitava de uma limpeza da "ferrugem" medieval para voltar a ter o brilho dos tempos antigos.[42] Esse trabalho dependia da pesquisa e descrição do uso da língua e do estilo dos grandes autores para sua devida imitação, projeto este que condicionou a publicação de manuais e repertórios lexicais nos séculos seguintes. Nesse processo, desenvolve-se uma relação simbiótica entre o interesse acadêmico e a tecnologia: "na medida em que as afirmações sobre o passado podiam ser agrupadas e divulgadas na forma impressa, elas também podiam se constituir nas grandes compilações de conhecimento que eram as ferra-

[41] Segundo Considine (2008, p.26), o senso de pertencimento espacial em relação à Roma antiga induz a classe culta italiana a buscar modelos para construir sua própria cultura urbana, republicana, laica e ciente dos deveres do serviço público, em oposição à cultura cavaleiresca medieval.

[42] A referência ao rótulo de *rubiginosus* 'enferrujado' aplicado pelos humanistas ao latim medieval encontra-se em Sharpe (1996, p.97).

mentas filológicas fundamentais no século XVI" (Considine, 2008, p.55, tradução própria).[43]

Ao mesmo tempo, os vernáculos se consolidam como línguas de cultura e letramento através de sua gramatização, criando condições para a produção de grandes repertórios bilíngues latim-vernáculo e vice-versa, em contraste com os diversos ensaios realizados até então na forma de glossários. Duas experiências pioneiras nesse sentido ocorrem na Espanha do final do século XV, ambas realizadas por discípulos do humanismo italiano.

A primeira é o *Universal vocabulario en latín y en romance*, do cronista real[44] Alfonso de Palencia (Alfonsus Palentinus), publicado em 1490, cuja disposição em duas colunas comporta duas versões de um mesmo texto lexicográfico, constituído por cerca de 14 mil lemas latinos organizados alfabeticamente e acompanhados de glosas e derivações, na melhor tradição de Papias (Medina Guerra, 2008): à esquerda, a versão tradicional, em latim; à direita, a versão inovadora, com o mesmo texto traduzido para o espanhol (com exceção dos lemas, obviamente).

A segunda é publicada apenas dois anos depois por Elio Antonio de Nebrija (Aelius Antonius Nebrissensis) com o título *Lexicon hoc est dictionarium ex sermone latino in hispaniensem* ('Léxico, isto é, dicionário da língua latina em espanhol'), que demonstra a estima humanista pelos estudos de grego e parece remeter ao *Lexicon* greco-latino de Giovanni Crastone (Joannes Crastonus), publicado em 1476. No caso do *Lexicon* de Nebrija, trata-se de uma combinação simplificada e aparentemente original[45] de lemas compostos (constituídos por formas flexionadas) e glosas em vernáculo (normalmente um equivalente) marcadas sistematicamente pela preposição 'por' (que corresponde a *pro* em latim, esta já consagrada na lexicografia monolíngue como metalinguagem de marcação de equivalência). A compilação, distribuída em duas colunas por página, apresenta quase 29 mil entradas[46] impressas em tipografia única, das quais aproximadamente mil são palavras gregas transliteradas, que recebem equivalentes em latim marcados pelo verbo *interpretatur* ('é traduzido [por]'), e cerca de 1300 são lemas pluriverbais, na maior parte combinações fraseológicas acompanhadas da tradução para o espanhol; nesse caso, o equivalente simples pode ser tomado como índice de consciência do caráter idiomático de certas expressões, necessária tanto para a compreensão quanto para a produção linguística correta (e.g. a equivalência de "facio rem diuinam" em latim e *sacrificar* em espanhol). A microestrutura do *Lexicon* também inclui alguma informação gramatical em latim, muitas vezes de forma abreviada, em particular o gênero dos verbos (*actiuum [a.], neutrum [n.], deponens [d.]*), a classe

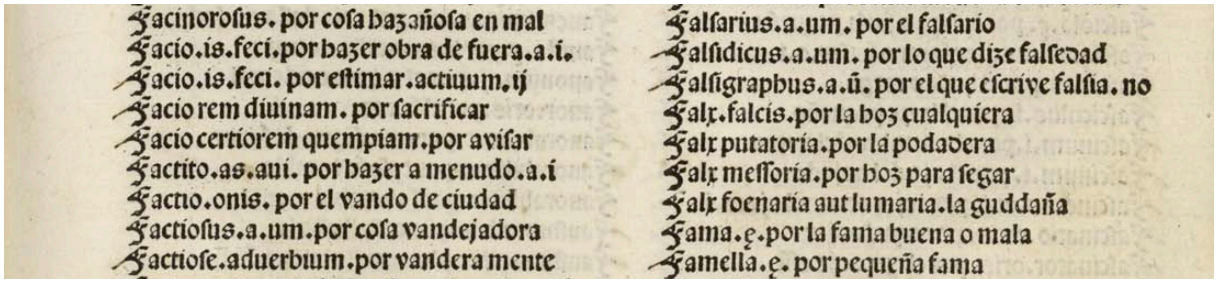
[43] No original: "Because statements about the past could be marshalled and communicated in print, they could be built up into the great learned compilations that are the fundamental philological texts of the sixteenth century. Technical possibilities and scholarly aims were in a symbiosis […]"

[44] A menção, no prefácio, de que a obra foi realizada "por mandado" da "muy excelente señora doña ysabel Reyna de castilla: y de leon: de aragon: y de siçilia […] al nombre de españa" (Medina Guerra, 2008, p.260, n.8) revela o interesse da corte tanto pelo letramento em latim quanto pela afirmação da língua nacional.

[45] Aparentemente, porque é bastante semelhante à estrutura observada no manuscrito de Alcobaça, feita da combinação entre um lema verbal composto e uma glosa em vernáculo. A prática de glosar materiais didáticos monolíngues não havia terminado, como mostram as glosas em espanhol inseridas ao lado dos exemplos gramaticais na primeira edição das *Introductiones latinae* de Nebrija, publicada em 1481 (disponível em <http://bdh-rd.bne.es/viewer.vm?id=0000174116>).

[46] As contagens apresentadas neste parágrafo se baseiam no texto digital da edição de 1492, disponível em <https://github.com/Diegeau/Nebrija_Lexico_1492_text_files> (acesso em 15.nov.2024).

Figura 5 – Excerto do *Lexicon* de Nebrija.


Facinoroſus. por coſa bazañoſa en mal
Facio.is.feci.por bazer obra de fuera.a.i.
Facio.is.feci. por eſtimar.actiuum.ij
Facio rem diuinam. por ſacrificar
Facio certiorem quempiam.por aviſar
Factito.as.aui. por bazer a menudo.a.i
Factio.onis. por el vando de ciudad
Factioſus.a.um.por coſa vandejadora
Factioſe.aduerbium.por vandera mente
Falſarius.a.um. por el falſario
Falſidicus.a.um. por lo que dize falſedad
Falſigraphus.a.ũ. por el que eſcrive falſia. no
Falx.falcis.por la boz cualquiera
Falx putatoria.por la podadera
Falx meſſoria.por boz para ſegar
Falx foenaria aut lumaria.la guddaña
Fama.ę.por la fama buena o mala
Famella.ę. por pequeña fama

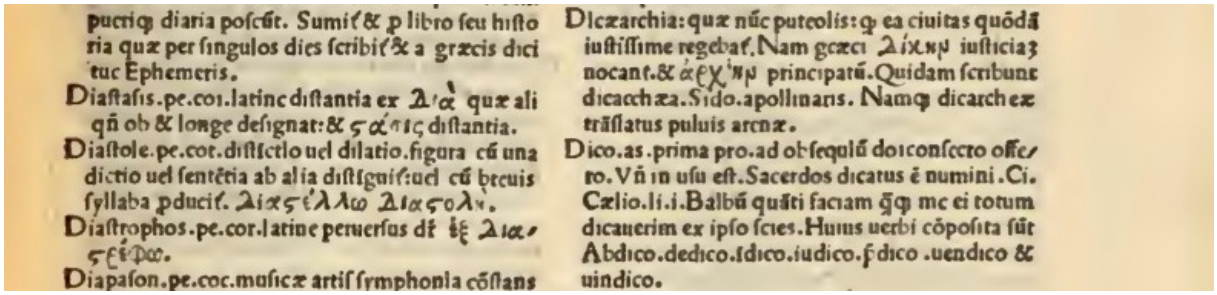
Fonte: Nebrija (1492, fol.62r).

das palavras gramaticais (*pronomen, aduerbium, praepositio, coniunctio, interiectio*) e o tipo de derivação que caracteriza o lema (*comparatiuum, superlatiuum, participium, praeteritum*). A Figura 5 reproduz um excerto constituído pelo verbo *facio* e seu entorno e permite observar algumas dessas características.

Ao contrário de Palência, que nunca teve sua obra reeditada, o enorme sucesso do *Lexicon* de Nebrija é demonstrado não só pela grande quantidade de edições ampliadas, publicadas desde o início do século XVI (Codoñer, 1996a), mas sobretudo por sua influência na lexicografia bilíngue europeia, para a qual se pode dizer que constitui o primeiro modelo (Medina Guerra, 2008).

A lexicografia monolíngue, por sua vez, também passou por um processo de renovação. Uma primeira tentativa é realizada por Johannes Reuchlin, que publica em 1478 um *Vocabularius breviloquus* já com algumas citações de autores clássicos mas ainda muito preso aos modelos medievais (em particular, a estruturação de longos verbetes com base na derivação) (Considine, 2008). Mas, assim como ocorrera com Palência, a obra de Reuchlin foi rapidamente ofuscada pela publicação, em 1502, do *Dictionarium* de Ambrogio Calepino (Ambrosius Calepinus). Seu principal diferencial foi incluir, junto à explicação do significado, numerosas citações de autores clássicos, que serviam como exemplo e abonação dos significados e, ao mesmo tempo, modelos para a renovação do uso do latim na época; além disso, a indicação frequente da quantidade silábica das palavras servia como guia para a pronúncia adequada nas conversações. Segundo Considine (2008, p.29), o Calepino (assim seu *Dicionarium* passou a ser conhecido) foi como que uma resposta à disseminação de textos em latim clássico, reeditados na forma impressa, que consistia em "documentar o vocabulário dessas obras e ao mesmo tempo excluir o que era próprio do mundo pós-clássico". A Figura 6 reproduz um excerto da primeira edição, suficiente para ilustrar a organização interna das informações no verbete. Observando-se, na coluna à direita, a entrada iniciada pelo lema "Dico.as", a primeira informação que se segue é a quantidade silábica, indicada pela expressão abreviada "prima pro[ducta]" ('primeira longa'); em seguida, a explicação do significado ("ad obsequium do, consecro, offero", i.e. 'dou em deferência, consagro, oferto') acompanhada por dois exemplos de uso: um anônimo ("Sacerdos dicatus est numini", 'o sacerdote se consagra à divindade'), outro de Cícero (indicado de forma abreviada "Ci.Caelio.li.i", isto é, 'primeiro livro do *Pro Caelio* de Cícero'); por fim, uma lista de

Figura 6 – Excerto do *Dictionarium* de Calepino.



puerisq; diaria poscit. Sumit & p libro seu historia quæ per singulos dies scribit & a græcis dicitur Ephemeris.
Diastasis.pe.cor.latine distantia ex διά quæ aliqñ ob & longe designat: & στάσις distantia.
Diastole.pe.cor.distinctio uel dilatio.figura cũ una dictio uel sentẽtia ab alia distinguit: uel cũ breuis syllaba producit. διαστέλλω διαστολή.
Diastrophos.pe.cor.latine peruersus dr ἐκ διαστρέφω.
Diapason.pe.coc.musicæ artis symphonia cõstans
Dicæarchia: quæ nũc puteolis: q; ea ciuitas quõdã iustissime regebat. Nam græci δίκαιον iusticiã nocant. & ἀρχὴν principatũ. Quidam scribunt dicaechæa. Sido.apollinaris. Namq; dicarchea trãslatus puluis arenæ.
Dico.as.prima pro.ad obsequiũ do: consecro offero. Vñ in usu est. Sacerdos dicatus ẽ numini. Cælio.li.i. Balbũ quãti faciam q; me ei totum dicauerim ex ipso scies. Huius uerbi cõposita sũt Abdico.dedico.ídico.iudico.ꝑdico.uendico & uindico.

Fonte: Calepinus (1502, s.p.).

verbos derivados de *dico* ("Abdico, dedico, indico […]"). Em caso de palavras polissêmicas, o esquema se repete para cada significado; com isso, alguns verbetes tornam-se bastante extensos, ocupando dezenas de linhas e contendo grande número de citações.

Tais características fizeram do Calepino o protótipo do dicionário humanista, contando com mais de duzentas edições entre 1502 e 1779, incluindo edições poliglotas com equivalentes para diversas línguas (Considine, 2008).[47] Todavia, nessa altura a filologia já estava desenvolvida o suficiente para notar certa falta de rigor em sua prática lexicográfica, que se manifestava em uma ordem alfabética não estrita, na abonação ausente ou mal referenciada, na distinção pouco clara dos significados. Ideias sobre as necessidades da lexicografia circulam entre os humanistas, especialmente entre aqueles preocupados com questões educacionais como Erasmo e Vives. Numa carta a Guilherme Budé (Guilielmus Budaeus) escrita em 1521 a respeito da composição de um dicionário de grego, Erasmo de Roterdã (Desiderius Erasmus Roterodamus) define que "um léxico verdadeiramente completo é feito não só de uma lista de palavras, mas da explicação das expressões (*idiomata*) e das figuras (tropos) da língua grega" (Considine, 2008, p.36).[48] Dez anos mais tarde, o educador espanhol Juan Luis Vives dedica algumas linhas de seu longo tratado sobre educação (*De tradendis disciplinis*) à crítica dos dicionários disponíveis e à definição do que seria adequado para o percurso formativo:

> Que desse conjunto [de autores] seja compilado o dicionário da Língua Latina; nesse sentido, nenhum é completo e preciso o suficiente. Que sejam dois: de um lado, apenas uma lista de palavras acompanhada de uma tradução breve; de outro, um [livro] mais abundante com citações de autores intercaladas; isso fará o leitor mais confiante não só para assimilar a explicação, mas também para saber como usará [a língua], coisa que talvez ficasse inacessível sem um exemplo. Será útil oferecer também aos mais jovens versões duplas em cada

[47] Segundo Acero Durántez (2008, p.19), a primeira edição trilíngue incorpora o grego e o italiano em 1550, chegando a onze na edição de 1588, com o acréscimo do francês, alemão, hebraico, flamenco, inglês, húngaro e polonês, respectivamente. Sobre repertórios poliglotas no século XVI, cf. também Stein (2017).

[48] Embora Budé não tenha publicado semelhante léxico, sua obra filológica é exemplo de uma nova atitude perante os textos e o conhecimento, que associa o significado das palavras com o estudo das realidades culturais da vida cotidiana no mundo antigo, manifestando assim não apenas a valorização do conhecimento enciclopédico mas também um senso de anacronismo (Considine, 2008, p.32). Já Erasmo é autor de uma compilação de ditos e provérbios publicada primeiro em Paris, em 1500 (a *Collectanea Adagiorum*, com cerca de 800 entradas), depois em Veneza, na oficina de Aldus Manutius em 1508, sob o título *Adagiorum chiliades tres* ('Três milhares de adágios'); a edição póstuma realizada por Paulus Manutius em 1608 conta com pouco mais de 4500 entradas (cf. edição digital em <https://github.com/nevenjovanovic/camena-neolatinlit>).

> uma das línguas vernáculas, uma em que palavras latinas se traduzam nos vernáculos e outra o contrário, com palavras vernáculas traduzidas em latim. Em nossa língua isso foi feito por Antônio de Nebrija, em uma obra não muito exata, útil mais para os iniciantes do que para os mais avançados (Vives, 1551, p.291, tradução própria).[49]

Em outra passagem, Vives estima o trabalho de Calepino como o de "um bom compilador de trabalhos anteriores, mas pouco apto a completar o que faltava a eles" (Vives, 1551, p.297, tradução própria).[50]

Do comentário de Vives se podem extrair duas diretrizes para a lexicografia humanista. A primeira é a preocupação pedagógica, que se verifica na recomendação de dois tipos de dicionários conforme o nível de aprendizado: para iniciantes, um conjunto de dicionários bilíngues nas duas direções (latim-vernáculo e vernáculo-latim); para iniciados, um dicionário monolíngue com exemplos e citações. A segunda é a preocupação filológica, que se manifesta na necessidade de apoiar o trabalho diretamente em textos de autores clássicos, mais do que na reunião de compilações anteriores. É provável que suas colocações representassem o pensamento geral humanista no início de século XVI; ao menos é o que parece quando se observa a publicação na França, durante a década de 1530, do exato conjunto de três dicionários preconizados por Vives, pelas mãos de Robert Estienne: o monolíngue *Dictionarium, seu Linguae Latinae Thesaurus* (1531), e os dois bilíngues *Dictionarium latinogallicum* (1538) e *Dictionnaire francoislatin* (1539).

Robert Estienne (Robertus Stephanus) era herdeiro de uma oficina tipográfica nas dependências da Universidade de Paris. Sua primeira grande publicação foi a edição crítica da Bíblia em latim (a *Vulgata*) em 1527, cujo sucesso fez dele a figura mais proeminente do mercado livreiro parisiense de seu tempo (Considine, 2008, p.40). Em 1531, Estienne publica seu *Dictionarium, seu Linguae Latinae Thesaurus* com a missão de suprir a necessidade de um dicionário monolíngue que refletisse o uso encontrado nos autores clássicos (Wooldridge, 2002). No prefácio ele menciona a demanda de seus clientes por uma obra de referência sem os erros de citação ou com informação sobre os acentos e nomes próprios, e cita nominalmente Budé como exemplo do rigor metodológico (Stephanus, 1543, fol.*ii.r). Um segundo texto prefacial indica as "principais vantagens da obra" (Stephanus, 1543, fol.*iii.r): nomenclatura em ordem alfabética, incluindo palavras latinas e estrangeirismos; anotação gramatical de gênero dos nomes e pretéritos e supinos dos verbos; informação sobre os acentos, por meio das quantidades silábicas;

[49] No original: "Ex quibus universis confletur dictionarium Latinae Linguae, quod nullum est plenum satis et iustum. Istudque sit duplex, alterum enumeratione tantum vocabulorum, brevi interpretatione adiecta, alterum copiosius dictis authorum intermistis: quod faciet, non solus securus sit lector, et expositione acquiescat, verum ut sciat quoque quemadmodum usurus sit, quod forte eum sine exemplo lateret. Expediet in quaque etiam vulgari lingua geminum pueris tradi, vnum quo Latina verba reddantur vulgaribus, alterum quo vice versa vulgaria Latinis: quod in nostro sermone Antonius Nebrissensis fecit: opus non satis exactum, tyronibus magis, quam prouectioribus vtile."

[50] No original: "Ex quibus dictionarium suum Ambrosius Calepinus congessit, homo congerendis quidem illis idoneus, explendis vero quae deerant, non idoneus."

comprovação do uso de palavras, construções e expressões[51] através de exemplos extraídos dos autores clássicos, também organizados alfabeticamente, com referência bibliográfica precisa (autor, livro, parágrafo ou verso), fruto de suas leituras e atividade editorial intensa; e distribuição dos diversos significados das palavras "não de forma amontoada, com cada um ocupando o início de uma linha".[52]

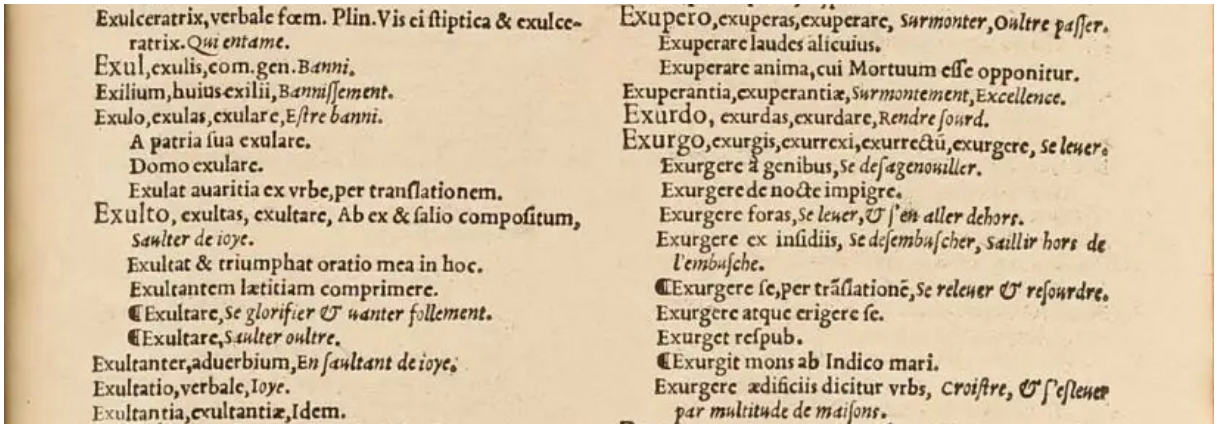
Com efeito, o cuidado e os recursos tipográficos empregados na organização das informações por Estienne eram inéditos na lexicografia até então. Os mesmos recursos foram usados no dicionário bilíngue latim-francês (*Dictionarium latinogallicum*), construído sobre um compêndio das informações do *Thesaurus*, que diminui a quantidade de informações aproximadamente à metade, porém constitui ainda assim um volume de dimensões consideráveis (1430 páginas). Os aspectos gráficos mais visíveis são o uso de recuo de parágrafo, alternância maiúscula/minúscula e de tamanho da letra, cuja associação e aplicação rigorosa permite demarcar o tipo de informação do ponto de vista lexicográfico. Por exemplo, entradas de palavras primitivas têm lema em letras minúsculas, sem recuo de parágrafo, em fonte maior; entradas de palavras derivadas têm a mesma configuração, mas com fonte menor; entradas de nomes próprios são semelhantes a essas últimas, porém com lema grafado em letras maiúsculas. O recuo é utilizado para marcar acepções e exemplos, que se tornam subentradas da entrada principal; as acepções se distinguem dos exemplos pelo uso de uma marca de parágrafo (¶); as indicações aos autores, suprimidas na primeira edição (1538), são recuperadas na segunda edição aumentada (1546) (Wooldridge, 2002). Por fim, toda a parte em francês do dicionário é grafada em itálico. A Figura 7 ilustra essa configuração. No meio da coluna à esquerda se observa, por exemplo, o verbo 'Exulto', considerado primitivo (letra maior), seguido por duas formas flexionadas do lema ('exultas, exultare'), uma explicação etimológica ('Ab ex et salio compositum') e a tradução para o francês da época ('*Saulter de joye*', i.e. 'pular de alegria'); seguem-se duas linhas com exemplos, então duas linhas marcadas por ¶ indicando diferentes acepções, com o lema representado pelo infinitivo ('exultare'); as últimas três linhas da coluna são entradas de palavras derivadas (Exultanter, Exultatio, Exultantia), sem recuo, grafadas em tamanho normal.

O *Dictionarium latinogallicum* (DLG) estabelece um novo parâmetro para dicionários bilíngues latim-vernáculo, até então constituído pelo *Lexicon* de Nebrija e pelo Calepino poliglota. Prova disso é o percurso editorial do primeiro dicionário impresso latim-inglês, compilado por Thomas Elyot. A primeira versão, intitulada *The Dictionary of syr Thomas Eliot knyght*, publicada no mesmo ano do DLG (1538), organiza as entradas como o Calepino, isto é, com acepções e exemplos de expressões apresentadas continuamente em um único bloco de parágrafo correspondente ao verbete. A segunda edição, de 1542, com o título bilíngue *Bibliotheca Eliotae Eliotis*

[51] Estienne exemplifica nesse mesmo texto o que ele entende por *constructio* ('construção') com combinações gramaticais de verbo e pronome "Timeo te, timeo tibi, timeo a te", ao passo que *formula*, *locutio* ou *varietas loquendi* (respectivamente, 'fórmula, expressão, variedade de expressão') são exemplificadas pelas combinações lexicais "Facere argentariam, Facere bellum, Facere castra").

[52] No original: "Vocabulorum variae significationes, non promiscue, et in vnum quasi aceruum congestae sunt, mas ita digestae collocataeque, vt versuum singula capita singulae occupent, quantunuis multiplex fuerit variusque dictionis vsus."

Figura 7 – Excerto do *Dictionarium latinogallicum* de Estienne.



Exulceratrix, verbale fœm. Plin. Vis ei ſtiptica & exulceratrix. *Qui entame.*
Exul, exulis, com. gen. *Banni.*
Exilium, huius exilii, *Banniſſement.*
Exulo, exulas, exulare, *Eſtre banni.*
A patria ſua exulare.
Domo exulare.
Exulat auaritia ex vrbe, per tranſlationem.
Exulto, exultas, exultare, Ab ex & ſalio compoſitum, *Saulter de ioye.*
Exultat & triumphat oratio mea in hoc.
Exultantem lætitiam comprimere.
¶ Exultare, *Se glorifier & vanter follement.*
¶ Exultare, *Saulter oultre.*
Exultanter, aduerbium, *En ſaultant de ioye.*
Exultatio, verbale, *Ioye.*
Exultantia, exultantiæ, Idem.
Exupero, exuperas, exuperare, *Surmonter, Oultre paſſer.*
Exuperare laudes alicuius.
Exuperare anima, cui Mortuum eſſe opponitur.
Exuperantia, exuperantiæ, *Surmontement, Excellence.*
Exurdo, exurdas, exurdare, *Rendre ſourd.*
Exurgo, exurgis, exurrexi, exurrectũ, exurgere, *Se leuer.*
Exurgere à genibus, *Se deſagenouiller.*
Exurgere de nocte impigre.
Exurgere foras, *Se leuer, & ſ'en aller dehors.*
Exurgere ex inſidiis, *Se deſembuſcher, Saillir hors de l'embuſche.*
¶ Exurgere ſe, per trãſlationẽ, *Se releuer & reſourdre.*
Exurgere atque erigere ſe.
Exurget reſpub.
¶ Exurgit mons ab Indico mari.
Exurgere ædificiis dicitur vrbs, *Croiſtre, & ſ'eſleuer par multitude de maiſons.*

Fonte: Stephanus (1538, p.177).

Library, distribui os exemplos como subentradas recuadas e organizadas alfabeticamente, indício de que o contato com a obra de Estienne determinou o formato da nova versão (Stein, 2014, p.348).[53] Não por acaso, o nome de Robertus Stephanus é incluído nominalmente no índice de autores consultados, apresentado logo após o prefácio da segunda edição.[54]

Em suma, pode-se dizer que as obras de Nebrija, Calepino e Estienne constituem os "três grandes fluxos de tradição lexicográfica latina na Europa" (Verdelho, 1995, p.233) e podem servir como balizas para a observação dos dicionários latim-português produzidos pelos lexicógrafos portugueses, apresentados individualmente nas subseções a seguir. Embora aquelas três vertentes tenham se desenvolvido de modo desigual no contexto geral europeu – com Nebrija se mantendo especialmente forte na península Ibérica, Calepino dominando o ambiente católico, em particular entre os Jesuítas, e Estienne exercendo influência sobretudo nas culturas onde a Reforma protestante havia sido bem sucedida –, o caráter internacional do estudo do latim pelos humanistas sugere que diferentes dicionários circulavam pelos mesmos espaços e se sobrepunham no contexto escolar, tornando difícil estabelecer com precisão as fontes utilizadas a cada nova compilação.[55]

[53] Stein (2014), que faz uma descrição minuciosa dos dicionários de Thomas Elyot, menciona a influência persistente de Estienne também na lexicografia latina em língua inglesa praticada nos séculos XVI e XVII pelos sucessores de Elyot, como Thomas Cooper, Thomas Thomas, Adam Littleton e Robert Ainsworth. A mesma influência se observa na lexicografia latina em alemão, em que uma obra com microestrutura inspirada no *Catholicon* como o *Dictionarium Latinogermanicum* de Joannes Serranus (1539) é reformulada por Ioannes Frisius em 1556 sob a evidente influência do DLG.

[54] Essa atitude de reconhecimento já havia sido manifestada por Elyot no prefácio da primeira edição, em uma passagem que demonstra a consciência humanista de pertencer a uma longa tradição lexicográfica – de Festo e Nônio a Calepino, passando por Dionysius Nestor, Giovanni Tortelli, Lorenzo Valla e Guilherme Budé, excluídos evidentemente todos os medievais –, à qual ele pertence e na qual ele se insere criticamente. Para Stein (2014, p.26), a lista de predecessores que Elyot menciona serve para amplificar o argumento da abrangência lexical que sua compilação alcançou, diante da incompletude dos cinco primeiros ou da especificidade da pesquisa dos dois últimos, dedicados à coleção de frases elegantes ou construções gramaticais, não do significado das palavras.

[55] À atitude honesta de Elyot contrasta, por exemplo, a estratégia adotada por Sébastien Gryphe, editor em Lyon na segunda metade do século XVI, de acrescentar à estrutura do Calepino passagens inteiras copiadas de cada nova edição do *Thesaurus* de Estienne; para Furno (2011, 194), isso constituía "um meio cômodo de plagiar Estienne, contornando o risco do privilégio real" concedido ao francês sem perder o apelo comercial que o

Em Portugal, em particular, o movimento humanista despertou o interesse da corte desde a segunda metade do século XV. Se, de um lado, se verifica uma grande presença de estudantes portugueses em universidades italianas, de outro, o rei D. João II contrata um humanista italiano, Cataldo Parísio, para exercer as funções de *orator* do rei (secretário latino, orador político e embaixador) e, sobretudo, para dar uma educação moderna ao príncipe herdeiro, D. Jorge (Ramalho, 2013, p.221). O humanismo português floresceu na primeira metade do século XVI, especialmente durante o reinado de D. João III, grande admirador do pensamento de Erasmo. Nesse período, grandes humanistas como Damião de Góis e André de Resende circulavam nos principais meios intelectuais europeus e um projeto de renovação dos costumes e da cultura segundo os ideais do humanismo cristão era posto em prática através da reforma das instituições e dos currículos educacionais (Rodrigues, 1981, p.31–2).[56]; são exemplos a transferência da Universidade de Lisboa para Coimbra em 1537, a conselho de Juan Luís Vives, e a criação do Real Colégio das Artes e Humanidades nesta mesma cidade, em 1542. A escola e o letramento em latim passam a ser identificados pela classe dirigente como meio de acesso social (Rodrigues, 1981, p.41). Esse processo reformista foi interrompido, ou melhor, reorientado nas décadas seguintes pelas determinações contra-reformistas do Concílio de Trento. Não por caso, a nova política conservadora dominante entregou a escola ao humanismo católico representado pelos Jesuítas, que passam a comandar o Colégio das Artes a partir de 1555 e a nova Universidade de Évora, fundada em 1559. Nesse movimento, segundo Rodrigues (1981, p.48–9), "morreu então, na cultura portuguesa, a doutrina, em larga medida também a prática, que reclamava a guerra à escolástica e à cultura medieval", dando lugar a "um humanismo que pedia ao classicismo os modelos literários e as referências exemplares, mas que recebia das dominâncias da Contra-Reforma a problemática, a dialéctica, a tensão polémica, o sentido do discurso", em suma, "um humanismo apreciado não como cultura – a profana e das coisas profanas –, mas como uma forma ou instrumento ao serviço da doutrina católica recentemente aprovada em Trento". É nesse ambiente político-cultural que se inserem lexicógrafos portugueses dos séculos XVI e XVII.

3.2.2 O *Dictionarium* de Jerônimo Cardoso

Jerônimo Cardoso foi um humanista português, formado pela Universidade de Salamanca, que exerceu atividade como pedagogo da Língua Latina em Lisboa em meados do século XVI. Embora tenha se correspondido com os grandes humanistas portugueses do seu tempo, é provável que tenha lecionado em escolas ou colégios não universitários (Almeida, 1959, p.145).

Sua produção lexicográfica, pioneira em Portugal na forma impressa, parece responder aos preceitos de Vives em relação à necessidade de obras escolares com traduções para o vernáculo,

nome do italiano já portava.

[56] Rodrigues (1981, p.39–40), observando os diversos "humanismos" que se desenvolveram em diferentes regiões fora da Itália, descreve humanismo português como resultado de uma fusão de "três componentes essenciais: o sentido do profano, em termos de valores e de saberes, realçado pelos humanistas de Itália; o anseio evangélico e culturalmente avançado do humanismo cristão; e a consciência das realidades históricas nacionais, sobretudo na dimensão de país descobridor".

voltadas para os mais jovens.

Excetuando-se um primeiro repertório de pretéritos e supinos, provavelmente nos moldes do manuscrito de Alcobaça, a primeira publicação de caráter lexicográfico de Cardoso foi o *Dictionarium Iuventuti studiosae admodum frugiferum* ('Dicionário muito útil para a juventude estudiosa'), impresso em Coimbra por João de Barreira em 1551. Com pouco mais de 3 mil palavras latinas reunidas em seções temáticas dedicadas aos nomes das partes do corpo, das partes da casa, dos utensílios domésticos, dos armamentos, é um dicionário que atendia aos primeiros passos do aprendizado de vocabulário em latim, nos moldes dos dicionários de "coisas" medievais.[57] Já o par de dicionários bilíngues alfabéticos foi publicado na década de 1560: primeiro o *Dictionarium ex Lusitanico in Latinum sermonem* ('Dicionário da língua portuguesa para a língua latina), impresso em Lisboa em 1563 por João Álvares; em seguida, o *Dictionarium latino lusitanicum* ('Dicionário latim-português'), publicado postumamente em Coimbra, em 1569 ou 1570, como parte de um conjunto lexicográfico editado por Sebastião Stockhammer que incluía os dois dicionários bilíngues, um vocabulário de termos eclesiásticos, uma tabela de equivalências de moedas, pesos e medidas e, por fim, um dicionário de nomes próprios (este de autoria de Stockhammer). A finalidade desse conjunto de instrumentos de informação sobre o léxico é auxiliar nas atividades de letramento desenvolvidas pelos humanistas portugueses nos colégios de Lisboa e Évora (Almeida, 1959, p.146). Isso transparece tanto no título do primeiro dicionário que publicou ("para a juventude estudiosa") quanto no prefácio escrito Sebastião Stockhammer para o *Dictionarium latino lusitanicum*, que caracteriza Cardoso como "homem de muita leitura e praticamente nascido para instruir a juventude nas boas letras" (Cardoso, 1570, p.10, tradução própria)[58].

Do conjunto editado por Stockhammer (Cardoso, 1570), o *Dictionarium latino lusitanicum* ocupa a maior parte.[59] O aspecto gráfico é coerente com os recursos de impressão disponíveis em Portugal naquele período.[60] A página é dividida em duas colunas, com indicação de letras iniciais e paginação no cabeçalho. A tipografia é composta de uma única fonte para palavras latinas e portuguesas, com lemas e explicação semântica marcados por uma inicial maiúscula e, quando possível, separados por um espaço em branco muitas vezes generoso. A ortografia admite um uso amplo de abreviaturas da tradição manuscrita, ainda vigentes, tanto no textos em latim (e.g. "atq;" por *atque*, "ardu9" por *arduus*, "fabellę" por *fabellae*, "exuberãtia" por *exuberantia*) quanto na parte em português (e.g. "casamẽto" por casamento). Por fim, ilustrações são empregadas em duas situações: capitulares decorativas demarcam o início de cada conjunto da ordem alfabética; e uma pequena figura de mão com o dedo indicador aponta para certo número

[57] O método de iniciar o ensino de vocabulário pelas coisas cotidianas é atestado por Vives (1551, p.191–2), que menciona a falta de um dicionário desse tipo, deixando a cargo do professor a tarefa de compilar repertórios de palavras necessárias ao uso cotidiano específico para cada idade.

[58] No original: "propemodum nati ad iuuentutem bonis literis instituendam".

[59] Mais precisamente, 262 fólios de um volume total de 421 impresso em formato 4º (21 cm), segundo a descrição do volume fornecida pela Biblioteca Nacional de Portugal, cf. <https://purl.pt/14265>.

[60] Sobre a cultura impressa quinhentista em Portugal, cf. Pedro (2018).

de entradas fraseológicas.[61] A disposição das entradas segue o estilo de Nebrija: palavras simples e combinações lexicais constituem entradas independentes, sem distinção; caso o verbete se estenda por mais de uma linha, o recuo é utilizado da segunda linha em diante. A Figura 8 permite uma visualização dessas características.

A macroestrutura consiste em um total de 26.932 entradas,[62] das quais a maior parte (75%) é constituída por verbetes que descrevem lexias simples; o restante (25%) corresponde a um conjunto considerável de entradas pluriverbais, em grande medida extraídas dos *Adagia* de Erasmo (Verdelho, 1995, p.234);[63] não há entradas para siglas e abreviaturas nem para afixos ou radicais. Da nomenclatura se depreende que se trata de um dicionário geral, sem restrições para classes ou tipos de palavras, havendo lugar para nomes próprios (e.g. *Abdera*), palavras tabu (e.g. *culus*), nomes depreciativos (e.g. *concubinus*) e estrangeirismos (e.g. *agya*, *alleluia*). Casos de homonímia constituem entradas distintas, normalmente diferenciadas por algum tipo de indicação formal (e.g. "Dico, as" vs. "Dico, is, xi, ctum", ou "Excido, is... breue" vs. "Excido, is... longa"). Casos de polissemia recebem tratamento semelhante, com a simples repetição do lema em uma nova entrada, talvez uma reminiscência do estilo das glosas também encontrado em Nebrija (e.g. as duas entradas para "Exuuiae.arum", que encerra a letra E, ou para "Fabula.ae"). O critério para arranjo das entradas é semi-alfabético, com entradas de palavras primitivas precedendo seus respectivos derivados e combinações lexicais (e.g. a primeira entrada da letra F tem como lema *faba* e inicia uma série de entradas relacionadas que incluem o derivado *fabarius*, as expressões "Faba cudetur in me", "Faba munus" e "Fabarum arrosor" e outros quatro derivados, *fabaginus*, *fabacea*, *fabarium*, *fabale*).

Quanto à microestrutura, a imensa maioria dos verbetes é constituído de lema e explicação semântica.

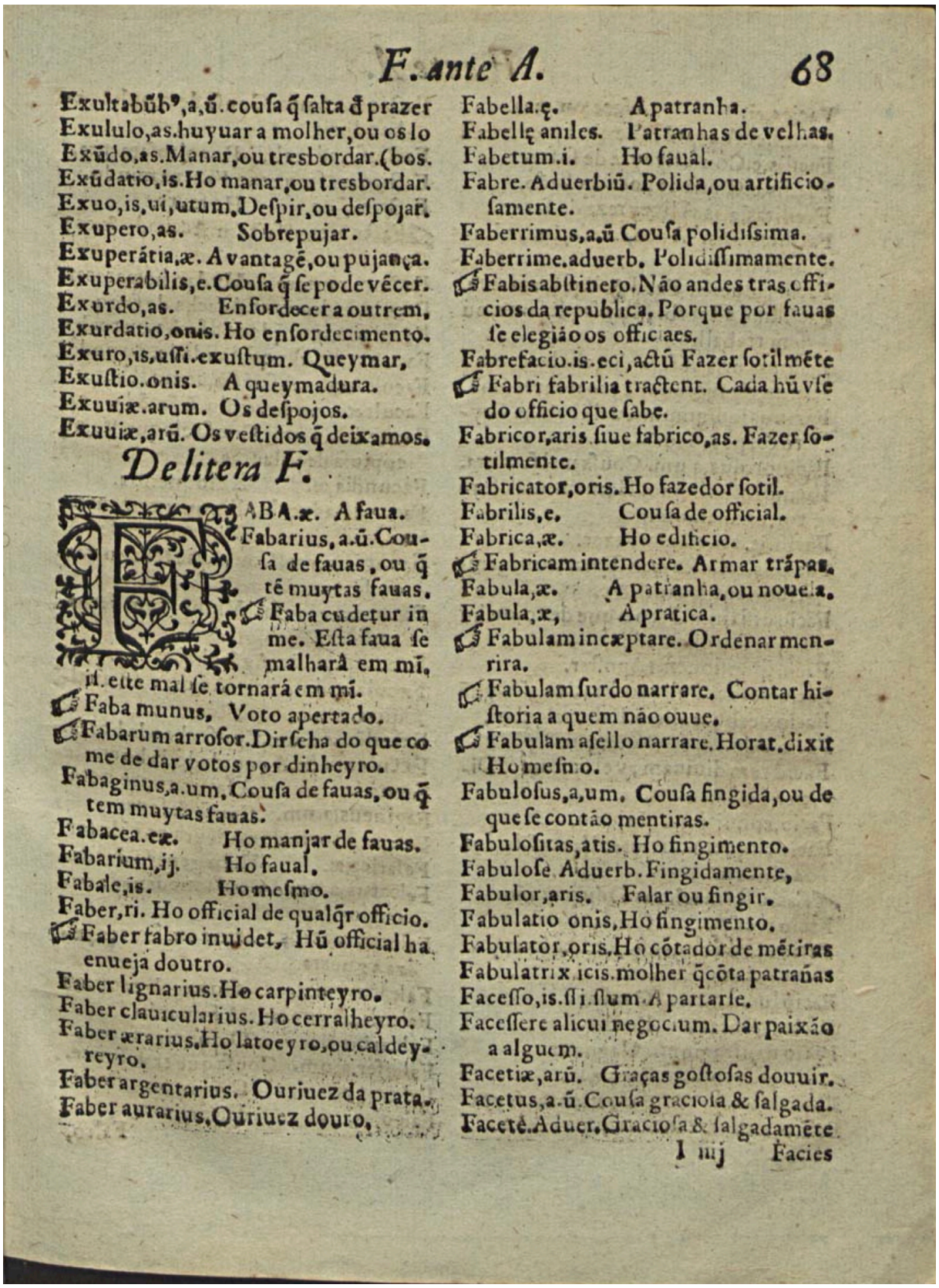
O lema é composto de uma ou mais formas flexionadas, já tornadas canônicas; a quantidade varia conforme a classe e o tipo morfológico do lexema: indeclináveis como advérbios, preposições, conjunções são lematizadas pela sua forma única; substantivos contam com duas formas, normalmente o nominativo e o genitivo singulares (ou plurais, no caso dos *pluralia tantum*); adjetivos são apresentados com duas ou três formas, de modo a recobrir os nominativos singulares de cada gênero (masculino, feminino e neutro), com os chamados adjetivos uniformes lematizados como os substantivos; verbos são sempre lematizados pela primeira pessoa do

[61] Stockhammer afirma no prefácio sua diligência como editor ao " assinalar e ampliar os provérbios elegantes, úteis e muito dignos de saber, dos quais este Léxico é abundante, bem como interpretar certas alegorias mais obscuras" (Cardoso, 1570, p.10, tradução própria do original: "Paroemias elegantes, utiles, scituque dignissimas, quibus hoc Lexicon abundat, …diligenter adscripsi, auxi et obscuriores quasdam allegorias interpretatus sum.")).

[62] Os dados quantitativos apresentados foram computados a partir da versão em texto digital disponível no *Corpus Lexicográfico do Português* (Cardoso, 2002), mediante um processo de estruturação dos dados descrito em detalhes no Capítulo 3.

[63] O título ampliado da obra faz menção aos adágios sem mencionar seu autor, provavelmente pelo fato de Erasmo ter sido proscrito pela Inquisição: *...cum adagiorum ferè omnium iuxta seriem alphabeticam perutili expositione*, isto é, "com a muito útil exposição de quase todos os adágios junto à ordem alfabética" (Cardoso, 1570, p. de rosto, tradução própria).

Figura 8 – Página do *Dictionarium latino lusitanicum* de Jerônimo Cardoso.



F. ante A. 68

Exultabũb⁹,a,ũ. cousa q̃ salta d̃ prazer
Exululo,as.huyuar a molher,ou os lo
Exũdo.as.Manar,ou tresbordar.(bos.
Exũdatio,is.Ho manar,ou tresbordar.
Exuo,is,ui,utum.Despir,ou despojar.
Exupero,as. Sobrepujar.
Exuperátia,æ. Avantagẽ,ou pujança.
Exuperabilis,e.Cousa q̃ se pode vẽcer.
Exurdo,as. Ensordecer a outrem,
Exurdatio,onis. Ho ensordecimento.
Exuro,is,ussi.exustum. Queymar,
Exustio.onis. A queymadura.
Exuuiæ.arum. Os despojos.
Exuuiæ,arũ. Os vestidos q̃ deixamos.

De litera F.

FABA.æ. A faua.
Fabarius,a.ũ.Cousa de fauas,ou q̃ tẽ muytas fauas.
Faba cudetur in me. Esta faua se malharâ em mĩ, i. este mal se tornará em mĩ.
Faba munus, Voto apertado.
Fabarum arrosor.Dirsçha do que come de dar votos por dinheyro.
Fabaginus,a.um, Cousa de fauas, ou q̃ tem muytas fauas.
Fabacea.eæ. Ho manjar de fauas.
Fabarium,ij. Ho faual.
Fabale,is. Ho mesmo.
Faber,ri. Ho official de qualq̃r officio.
Faber fabro inuidet, Hũ official ha enueja doutro.
Faber lignarius.Ho carpinteyro.
Faber clauicularius.Ho cerralheyro.
Faber ærarius,Ho latoeyro,ou caldeyreyro.
Faber argentarius. Ouriuez da prata.
Faber aurarius,Ouriuez douro,
Fabella.ę. A patranha.
Fabellę anites. Patranhas de velhas.
Fabetum.i. Ho faual.
Fabre. Aduerbiũ. Polida,ou artificiosamente.
Faberrimus,a.ũ Cousa polidissima.
Faberrime.aduerb. Polidissimamente.
Fabis abstinero.Não andes tras officios da republica. Porque por fauas se elegião os officiaes.
Fabrefacio.is.eci,actũ Fazer sotilmẽte
Fabri fabrilia tractent. Cada hũ vse do officio que sabe.
Fabricor,aris siue fabrico,as. Fazer sotilmente.
Fabricator,oris. Ho fazedor sotil.
Fabrilis,e. Cousa de official.
Fabrica,æ. Ho edificio.
Fabricam intendere. Armar trãpas.
Fabula,æ. A patranha,ou noueira.
Fabula,æ, A pratica.
Fabulam incæptare. Ordenar mentira.
Fabulam surdo narrare. Contar historia a quem não ouue.
Fabulam asello narrare.Horat.dixit Ho mesmo.
Fabulosus,a,um. Cousa fingida,ou de que se contão mentiras.
Fabulositas,atis. Ho fingimento.
Fabulosè Aduerb. Fingidamente,
Fabulor,aris. Falar ou fingir.
Fabulatio onis,Ho fingimento.
Fabulator,oris,Ho cõtador de mẽtiras
Fabulatrix icis.molher q̃cõta patranas
Facesso,is.ssi.ssum.A partarse.
Facessere alicui negocium. Dar paixão a alguem.
Facetiæ,arũ. Graças gostosas douuir.
Facetus,a.ũ.Cousa graciosa & salgada.
Facetè.Aduer.Graciosa & salgadamẽte

Fonte: Cardoso (1570, p.68).

singular do presente do indicativo (com exceção das chamadas formas impessoais, lematizadas pela terceira pessoa), acompanhadas quase sempre de uma segunda forma (respectivamente, a segunda pessoa do presente e a terceira pessoa do imperfeito), muitas vezes por uma terceira (a primeira pessoa singular do pretérito perfeito) e não raro por uma quarta (o supino no acusativo). Por vezes o lema pode contar com variantes ortográficas e variantes lexicais, geralmente mediadas por uma conjunção latina (e.g. *et*, *siue*, *uel*). No primeiro caso, as grafias concorrentes da forma de entrada são justapostas; não havendo variação morfológica, a segunda forma é comum a ambas (e.g. "Stingo et stinguo.is. inxi"; "Allux siue hallux.ucis"). No segundo caso, a variação implica em mudança do tipo flexional, seja entre lemas homógrafos, quando é marcada no interior das formas adicionais (e.g. "Lupercalia.um vel orum"), seja entre lemas não homógrafos, quando se contrapõem conjuntos completos de formas canônicas (e.g. "Larinx.ingos siue laringa.ae", "Hilaris.e siue hilarus.a.um").

A explicação semântica, por sua vez, consiste basicamente em definições ou equivalentes de tradução em português. No caso de lexias simples, a análise permite detectar alguns padrões de textualização, conforme a classe lexical: verbos geralmente são descritos por meio de equivalentes, às vezes acompanhados de especificação argumental (e.g. "Exululo.as. huyvar a molher ou os lobos"; "Exuo.is.ui.utum. Despir ou despojar"); substantivos recebem equivalentes ou breves definições, iniciados quase sempre por um artigo definido (e.g. "Fabarium.ij. Ho faval"; "Faber.ri. Ho official de qualquer officio"); adjetivos recebem muitas vezes definições iniciadas pelo substantivo português *cousa* (e.g. "Exuperabilis.e. Cousa que se pode vencer"; "Fabulosus.a.um. Cousa fingida, ou de que se contam mentiras"); advérbios recebem equivalentes em português, muitas vezes terminados em -MENTE (e.g. "Fabre. Aduerbium. Polida ou artificiosamente"; "Fabulose. Aduerb. Fingidamente"). No caso das lexias complexas, a grande maioria (82%) recebe um equivalente de tradução, muitas vezes acompanhado de uma informação adicional de caráter pragmático (e.g. "Oculis ferre. Trazer nos olhos. Dir-se-ha das cousas que muyto amamos"); outra parte significativa (15%) conta apenas com o comentário pragmático (e.g. "Pedem in sepulchro habere. Dir-se-ha do muyto velho, que tem ja hum pé na coua"). Por fim, alguns poucos verbetes trazem textos definitórios redigidos exclusivamente em latim, principalmente termos técnicos gregos (e.g. "Metonymia, figura est. i. Transnominatio"; "Prosopopaeia, interpr[etatur]. Personae fictio"), outros estrangeirismos e algumas palavras de baixo calão.

Nos exemplos acima se pode observar algumas informações suplementares sobre a forma ou o significado, adicionadas normalmente para fins de distinção, ora entre o lema e a definição, ora após esta.

Informações adicionais sobre a forma são escassas, cobrindo apenas 7% dos verbetes. A mais frequente (quase 1,5 mil ocorrências) é a informação sobre a classe lexical de palavras invariáveis, veiculada em latim segundo a terminologia da gramática clássica, por extenso ou de forma abreviada; são classificados assim os advérbios ("aduerb."), as conjunções ("coniunctio"), as preposições ("praepositio"), as interjeições ("interiectio"), bem como alguns substantivos indeclináveis ("indeclinabile"). Já a classe das palavras variáveis poderia ser depreendida pela

observação das formas que compunham o lema composto, tornando dispensável sua atribuição explícita; todavia, uma informação adicional frequente (436 ocorrências) diz respeito ao tipo de derivação que caracteriza certos substantivos ("diminutiuum") e verbos ("frequentatiuum").

Com exceção dos comentários pragmáticos a respeito das lexias complexas, vistos acima, outras informações adicionais relativas ao significado são raras, a maioria veiculadas em latim. Entre elas, pouco mais de uma centena traz a indicação do gênero do verbo (i.e. "actiuo", "deponens", "neutro", "absoluto"), que codifica a informação sobre o tipo de complemento (acusativo para os dois primeiros, não acusativo para o terceiro, nenhum para o quarto). Menos frequentes são marcas de uso (e.g. "antiquum", "poeticum", "inusitatum", "prouerbium", "rarum"), indicações de autoridade (e.g. "apud Persium", "inquit Varro", "Terentius") e exemplos introduzidos pela conjunção latina *ut* (e.g. "Appulsus, a, um. Cousa aportada por mar, vt, nauis appulsa est. A náo aportou").

Por fim, cabe destacar outros recursos metalinguísticos utilizados por Cardoso para lidar com a polissemia e a sinonímia. No primeiro caso, a separação de acepções dentro de um mesmo verbete é feita textualmente por meio da conjunção portuguesa *ou* (e.g. "Cado, is, cecidi, casum. Cayr, ou morrer"). No segundo, a equivalência entre palavras latinas é dada pela expressão latina "idem quod" ('o mesmo que…'), que demarca a sinonímia por uma espécie de referência cruzada (e.g. "Abmitto.is. idem quod mitto. Enuiar")[64]; já a expressão "vide…" ('ver…') é usada claramente com a finalidade de remeter a outra entrada do dicionário (e.g. "Leuus, a, um. vide super laeuus cum dipht[ongo]").

O conjunto dicionarístico de Cardoso se estabelece como o grande dicionário bilíngue latim-vernáculo em Portugal, contando com pelo menos dez novas edições publicadas no decorrer do século seguinte, com a última datada de 1694 (Almeida, 1959, p.150–1). Ainda em 1613 o conjunto é ampliado pela adição de uma espécie de dicionário onomasiológico latim-português contendo "modos de dizer e frases elegantíssimas", supostamente extraídas dos textos de Cícero, organizado alfabeticamente por temas. Ao mesmo tempo, o início do século XVII vê nascer, no contexto da educação jesuíta, dois pequenos vocabulários concebidos para servirem como índice de palavras latinas de dois métodos distintos de ensino de latim. Trata-se do *Index totius artis*, publicado por Antonio Velez em 1608 como apêndice ao manual de gramática de Manuel Álvares, e das *Raizes da língua latina*, publicadas por Amaro de Reboredo em 1621 como suplemento ao seu *Methodo grammatical para todas as línguas*. Apesar de diminutos quanto à nomenclatura (não passam de 5 mil entradas), seus diferentes métodos de selecionar e estruturar a informação sobre o léxico merecem uma observação detalhada, como se vê a seguir.

[64] Uma forma semelhante de referência é o uso frequente (perto de 800 ocorrências) da expressão "ho mesmo" para relacionar o significado de um verbete ao conteúdo do anterior (e.g. "Adnato.as. Nadar. | Adno.as.aui.atũ. Ho mesmo."

3.2.3 O *Index* de Antônio Velez

O *Index totius artis* é uma espécie de minidicionário latim-português de pouco mais de 3.600 entradas lexicográficas, cuja história de produção se estende ao longo do século XVII. Sua concepção está ligada a "uma das obras com maior impacto editorial e doutrinal a nível global" (Kemmler, 2013), a gramática latina de Manuel Álvares (*De institutiones grammaticae libri tres*), publicada pela primeira vez em 1572 e instituída pela *Ratio studiorum*, em 1599, como a gramática oficial de todos os colégios jesuítas (Springhetti, 1961–1962; Kemmler, 2018).

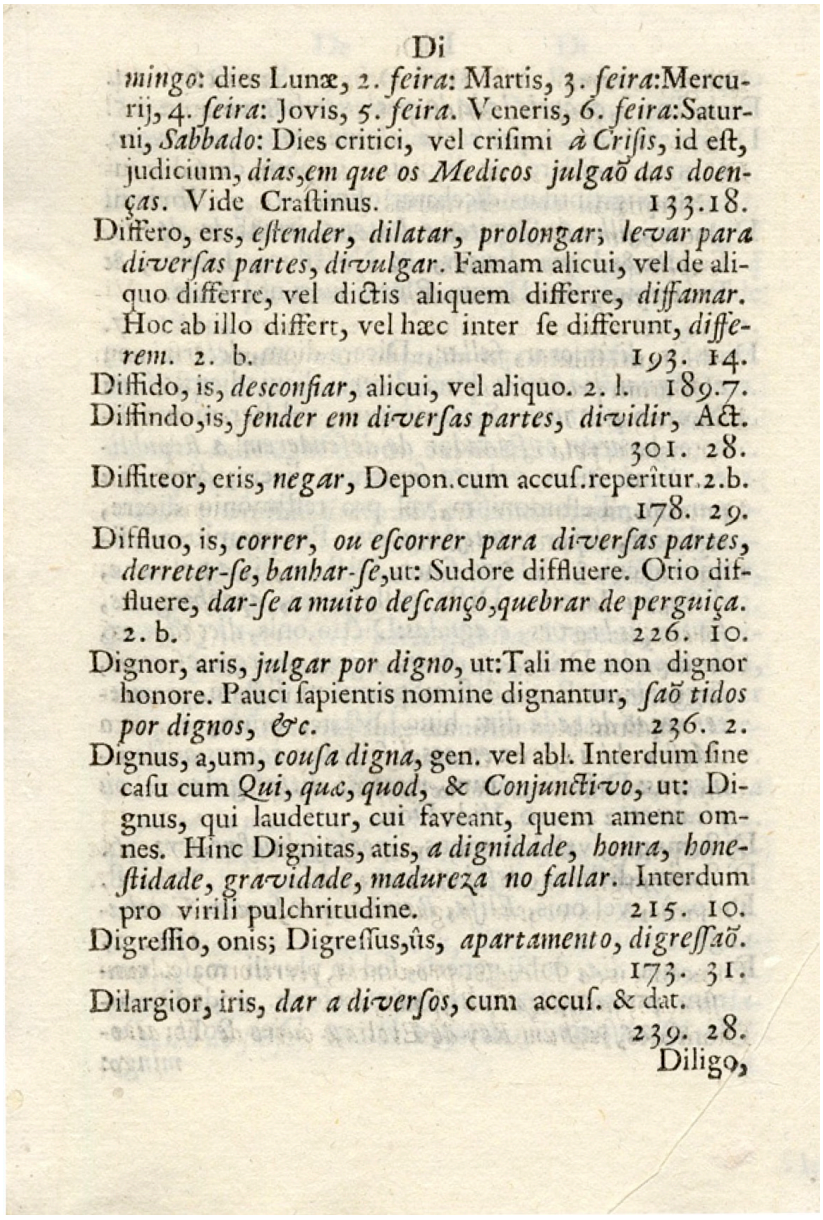
A produção do *Index* se dá basicamente em três etapas (Iken, 2002). A primeira corresponde à publicação, também em 1599, de uma edição revisada da gramática de Álvares, sob os cuidados de António Velez (Antonius Velesius Amiensis), então reitor da Universidade de Évora. Entre as intervenções do revisor está um índice remissivo das palavras latinas que ocorrem na gramática (intitulado *Index totius operis*), cujas entradas consistem normalmente em um lema simples acompanhado da indicação numérica para a página e a linha em que ocorre; eventualmente, o lema vem acompanhado de informações gramaticais como forma flexionada, quantidade silábica ou indicação de regência (Sánchez Salor; Gómez Gómez, 2020). Em 1608, vem à luz a versão compacta da gramática de Álvares (conhecida como 'Arte menor'), também sob os cuidados de Velez. Nessa edição, o índice é renomeado para *Index totius artis* e transformado em dicionário mediante o acréscimo de equivalentes de tradução para o português e, ocasionalmente, definições em latim, exemplos de uso e fraseologia. O motivo para esses acréscimos é informado em uma "Nota ao leitor", escrita em latim, em que se menciona a necessidade de fornecer aos aprendizes um material de apoio ("*promptuarium*") para o estudo do vocabulário que fosse não só mais consistente com o programa gramatical de ensino, mas também mais fácil de consultar e economicamente mais acessível do que os grandes dicionários de seu tempo, como o *Latinae linguae Thesaurus* ou o Calepino. Além disso, indica como público-alvo os aprendizes iniciantes (*tyronibus*), aos quais pretende fornecer "apenas o significado de cada palavra e algum uso" através de exemplos (Velez, 1744, p.366, tradução própria), mas sem transcrever passagens integrais ou referências a seus autores.

Após a morte de Velez, em 1609, a gramática e o *Index* são reeditados muitas vezes sem alterações significativas até que, em 1689, uma nova edição elaborada pela Academia eborense traz um índice com conteúdo bastante ampliado, com adição de novas acepções, exemplos de uso e fraseologia, bem como subentradas relativas a homônimos e formas derivadas. Esse *Index* se tornou a versão padrão para todas as edições posteriores da gramática de Álvares até a última, de 1755, quatro anos antes da expulsão dos jesuítas de Portugal (Iken, 2002). É essa versão, mais precisamente a edição publicada em 1744 pela tipografia da Universidade de Évora, que será descrita a seguir.[65]

A edição de 1744 compreende um volume de 660 páginas impressas em formato 8º (16

[65] A escolha da edição de 1744 se deve à sua disponibilidade em formato texto digital no *Corpus lexicográfico do português* (Velez, 2002), que favorece a extração de dados quantitativos, mediante o trabalho de processamento de dados descrito no Capítulo 3.

Figura 9 – Página do *Index totius Artis* de Antonio Velez.



Di

mingo: dies Lunæ, 2. *feira*: Martis, 3. *feira*:Mercurij, 4. *feira*: Jovis, 5. *feira*. Veneris, 6. *feira*:Saturni, *Sabbado*: Dies critici, vel criſimi *à Criſis*, id eſt, judicium, *dias,em que os Medicos julgaõ das doenças*. Vide Craſtinus. 133.18.

Differo, ers, *eſtender, dilatar, prolongar; levar para diverſas partes, divulgar*. Famam alicui, vel de aliquo differre, vel dictis aliquem differre, *diffamar*. Hoc ab illo differt, vel hæc inter ſe differunt, *differem*. 2. b. 193. 14.

Diffido, is, *desconfiar*, alicui, vel aliquo. 2. l. 189.7.

Diffindo,is, *fender em diverſas partes, dividir*, Act. 301. 28.

Diffiteor, eris, *negar*, Depon.cum accuſ.reperîtur.2.b. 178. 29.

Diffluo, is, *correr, ou eſcorrer para diverſas partes, derreter-ſe, banhar-ſe*,ut: Sudore diffluere. Otio diffluere, *dar-ſe a muito deſcanço,quebrar de perguiça*. 2. b. 226. 10.

Dignor, aris, *julgar por digno*, ut:Tali me non dignor honore. Pauci ſapientis nomine dignantur, *ſaõ tidos por dignos, &c.* 236. 2.

Dignus, a,um, *couſa digna*, gen. vel abl. Interdum ſine caſu cum *Qui, quæ, quod*, & *Conjunctivo*, ut: Dignus, qui laudetur, cui faveant, quem ament omnes. Hinc Dignitas, atis, *a dignidade, honra, honeſtidade, gravidade, madureza no fallar*. Interdum pro virili pulchritudine. 215. 10.

Digreſſio, onis; Digreſſus,ûs, *apartamento, digreſſaõ*. 173. 31.

Dilargior, iris, *dar a diverſos*, cum accuſ. & dat. 239. 28.

Diligo,

Fonte: Velez (1744, s.p.).

cm), das quais 296 (ou 45%) correspondem ao *Index*.[66] Os verbetes são dispostos em uma única coluna e demarcados por um recuo a partir da segunda linha, no estilo do Calepino; de fato, não há distinção gráfica entre acepções, exemplos ou subentradas, com todos esses itens distribuídos em um único bloco de parágrafo. A única informação espacialmente demarcada tem caráter extra-lexicográfico, a saber, os números remissivos a página e linha da gramática, posicionados sempre no final do verbete, alinhados à direita. Quanto aos recursos tipográficos, dois tipos de fonte são usados: romano para texto em latim, itálico para texto em português. A Figura 9 oferece uma página de amostra dessas características.

[66] Para efeito de comparação, na edição de 1643, a porção relativa ao *Index* é de cerca de 28% (135 páginas de um total de 490), o que dá uma dimensão do aporte de informações efetuado na terceira fase, a partir da edição de 1689.

A macroestrutura é constituída por 4.695 itens lexicais distribuídos em 3.619 verbetes. Essa diferença ocorre devido à existência de verbetes com múltiplos lemas, que podem constituir subentradas (em caso de palavras derivadas ou homográficas) ou lemas alternativos à cabeça do verbete (em caso de variantes lexicais ou ortográficas). Nos dois casos, é comum o uso metalinguístico de um conector latino, herdado da tradição manuscrita: para subentradas derivativas, as conjunções *unde* (‘donde’) e *hinc* (‘daí’, cf. “Dignus, a, um […] Hinc Dignitas, atis […]” na Figura 9);[67] para homônimos e parônimos, a conjunção adversativa *at* (‘mas’) ou a fórmula *est item*... (‘também existe […]’); para variantes lexicais e ortográficas, um conjunto compartilhado de conjunções, embora algumas sejam mais frequentes indicando as primeiras (e.g. *vel* ‘ou também’, *et* ‘e’, esta normalmente representada pelo símbolo &), outras indicando as segundas (e.g. *sive* ‘ou então’, *vel potius* ‘ou melhor’).

A nomenclatura tem caráter geral, incluindo itens lexicais de todas as classes de palavras – com predominância de substantivos (44%) e verbos (30%) –, além de uma dezena de afixos e dois pares de expressões idiomáticas. Por suas dimensões compactas, pode-se pensar que se trata de um vocabulário selecionado entre os itens mais frequentes da língua,[68] mas uma análise quantitativa das classes contempladas pela nomenclatura leva a outras conclusões. Por um lado, o confronto entre a nomenclatura do *Index* e uma lista dos mil lemas mais frequentes no latim clássico (Francese, 2020) revela uma alta taxa de cobertura para palavras funcionais (100% dos pronomes e preposições e acima de 85% das conjunções e advérbios), cobertura média para verbos e adjetivos (79% e 63%, respectivamente) e um valor baixo para substantivos (46%, para uma taxa média geral de 69%). Por outro lado, o levantamento dos capítulos da gramática em que o lema ocorre, possível mediante o mapeamento dos números de página indicados nos verbetes, mostra que a maior parte dos substantivos é mencionada no capítulo sobre o gênero dos nomes (38%), seguido pelo capítulo sobre os *anomala sive inaequalia* (28%), que descreve palavras com declinação variável ou parcial; verbos e adjetivos, por sua vez, são majoritariamente (53% e 65%, respectivamente) extraídos do capítulo sobre a sintaxe (*de constructione*), provavelmente devido ao seu caráter de núcleo definidor de regência. Ambas as perspectivas demonstram o vínculo do vocabulário com a natureza gramatical das informações que ele indexa, mais do que uma relação direta com qualquer *corpora* de textos clássicos.

Na microestrutura se encontram lemas, acepções e exemplos. Cada um desses componentes pode contar com informações adicionais, com função distintiva ou suplementar.

Lemas são compostos pela forma canônica principal, conforme a tradição (nominativo singular para formas nominais, 1ª pessoa do presente para formas verbais), geralmente acompanhada de uma única forma suplementar (genitivo para nomes, 2ª pessoa para verbos), uma

[67] Esse estilo textualizado de organização do verbete pode incluir o uso metalinguístico de orações; um exemplo é o verbete encabeçado por “Dico,is”, em que um conjunto de cinco lemas derivados (a saber, *dictum, dictio, dicto, dictata, dicteria*) são introduzidos pela expressão “A Dico fiunt […]” (‘A partir de *Dico* se produzem […]’) (Velez, 1744, s.v. Dico).

[68] Verdelho (1995, p.461), por exemplo, considera o *Index* “uma espécie de vocabulário do latim básico ou latim fundamental, escolhido naturalmente entre os escritores mais elegantes e de boa nota”. Iken (2002, p.58) endossa essa apreciação.

combinação suficiente para informar, de modo implícito, a classe de palavras à qual pertence; no caso das palavras invariáveis, o lema consiste na forma única acompanhada da indicação explícita da classe (e.g. "Adverb.", "Praep.", "Conjunct."). Palavras derivadas podem contar com informação sobre a origem e o tipo de derivação (e.g. o verbo frequentativo "Intento, as, Frequent. ab Intendo, *intentar, ameaçar*"). A informação etimológica pode ser usada para distinguir lemas homônimos ou homográficos, como em "Concido, is, à Cado, *cahir*…" *vs.* "Concîdo, à Caedo, *cortar em partes pequenas*…" (aqui secundada por um sinal gráfico no lema indicando a sílaba acentuada). Outras informações gramaticais ocorrem aparentemente com o mesmo objetivo: classes de palavras, como em "Par, paris, substant. *hum par*, adject. cousa igual…"; gênero do substantivo (e.g. "Populus, i, masc. *o povo*…" *vs.* "Populus, i, faem. o alemo, arvore"); informação prosódica, como em "occido, is, *matar*…2. l." (isto é, segunda sílaba longa) *vs.* "occido, is, *morrer*…2. b." (isto é, segunda sílaba breve).[69]

A informação sobre o significado é expressa geralmente por meio de uma definição ou de equivalentes de tradução em português. Algumas fórmulas padronizadas de enunciação do significado estão associadas a certas classes; adjetivos, por exemplo, são frequentemente definidos por uma expressão da forma *cousa* + Adjetivo (e.g. "Dignus, a, um, *cousa digna*", cf. Figura 9); verbos podem ter um equivalente especificado pela indicação de complemento (e.g. "Dilargior, iris, *dar a diversos*", cf. Figura 9); substantivos equivalentes são muitas vezes precedidos de artigo (e.g. "Spes, spei, *a esperança*", "Nesis, idis, *huma ilha de Campania*", "Maiores, um, *os antepassados*"); é comum advérbios serem traduzidos por um advérbio em -MENTE (e.g. "Nempe, Adverb. *certamente*"). Há casos, porém, em que a explicação do significado é feita em latim, como em um dicionário monolíngue. Nesses casos, a informação semântica pode ser veiculada por uma definição ou por um sinônimo, o qual pode ser precedido de expressões metalinguísticas como *id est* ('isto é'), *idem quod* ('o mesmo que') ou *latine* ('em latim') – esta particularmente usada em casos de estrangeirismos (e.g. "Phase, latinè transitus, *a Paschoa*").

A polissemia é tratada basicamente de duas maneiras, podendo tanto ser codificada na forma de uma única série de equivalentes em português, eventualmente separados por conjunção ('e', 'ou'), quanto ser distribuída por seções distintas do verbete; nesse caso, é comum o uso de marcadores metalexicográficos precedendo cada significado alternativo, normalmente em latim (e.g. "significat item", "ponitur pro"), eventualmente em português (respectivamente, "significa também" e "toma-se por"), podendo ser precedidos, nos dois casos, por indicadores de frequência como *interdum* ('às vezes') ou *aliquando* ('algumas vezes').

Diversos tipos de informação adicional acompanham eventualmente as acepções. Talvez o tipo mais característico do *Index* sejam as indicações de construção que acompanham principalmente verbos e adjetivos e descrevem as características gramaticais de seus complementos, sobretudo o caso morfológico implicado pela regência. Normalmente enunciados após a infor-

[69] A informação prosódica comparece principalmente nas últimas edições do *Index*, provavelmente por influência da *Prosodia* de Bento Pereira, publicada em meados do século XVII, em que tal informação é sistemática, como se verá mais adiante.

mação semântica, tais padrões sintáticos podem ser expressos de três maneiras: (1) através da indicação explícita do caso dos complementos, normalmente de forma abreviada e às vezes marcada pela preposição *cum* (e.g. "accus.", "cum dat.", "sine casu"); (2) através da indicação do 'gênero' dos verbos, normalmente de forma abreviada ("Act." para verbos ativos, "Neut." para neutros, "Depon." para depoentes), na medida em que estão associados implicitamente a um padrão de complementação;[70] (3) através de um pronome indefinido representando a forma do complemento (e.g. "aliquid alicui", "aliquid ab aliquo").[71] As três formas estão ilustradas na Figura 9, nos verbetes relativos aos lemas *dilargior*, *diffindo* e *diffido*, respectivamente.

Um segundo tipo de informação adicional corresponde ao que se chama atualmente de marcas de uso, normalmente expressas por um advérbio ou locução latina, podendo caracterizar uma forma, um significado ou uma construção quanto a sua frequência (marcadas por advérbios como *saepe* 'frequentemente' ou *raro* 'raramente', *usitatius* 'mais usado', *inusitatus* 'pouco usado'), diacronia (marcadas por expressões como "apud antiquos", 'entre os antigos'), atestação (marcada por expressões como "apud Poetas", "apud Comicos", "apud Jurisconsultos", "apud Oratores"), entre outras.

Exemplos de uso do lema podem estar associados a uma determinada acepção ou constituir um componente autônomo no verbete. No primeiro caso, são normalmente precedidos pela partícula latina *ut* ('como' ou 'por exemplo') e ilustram o significado; no segundo, costumam vir depois de um ponto-final e retratam normalmente uma expressão com caráter idiomático, na tradição das *locutiones* catalogadas amplamente por Estienne. Uma amostra dessa divisão pode ser observada na Figura 9, em que o verbete referente a *diffluo* traz um exemplo de uso para a acepção 'banhar-se' ("ut: Sudore diffluere", i.e., 'banhar-se de suor') e, em seguida, uma expressão que faz uso metafórico do lema ("Otio diffluere", lit. 'banhar-se de ócio', mas traduzido no *Index* por expressões idiomáticas em português, como "quebrar de preguiça"). Além disso, expressões idiomáticas podem ser codificadas pelo uso de pronomes como representantes de um dos complementos, como se observa no verbete para *differo* na Figura 9, em que se apresentam duas expressões equivalentes a 'difamar', ambas tendo o lema como núcleo e o substantivo *famam* como primeiro complemento, variando com relação ao segundo complemento, que pode assumir a forma de dativo ("Famam alicui [diferre]") ou ablativo preposicionado ("Famam de aliquo diferre").

Como se vê, o *Index* se destaca pela atenção especial à informação sintática, especialmente derivada do capítulo gramatical sobre a construção. Seu uso extensivo de articuladores sintáticos para delimitar os componentes do verbete, especialmente conectores latinos, é uma

[70] Segundo Colombat (2003, p.73), na vertente gramatical a que pertence Álvares, verbos ativos são aqueles que se constroem com complemento acusativo, podendo admitir um segundo complemento em outro caso; verbos neutros nunca se constroem com complemento acusativo, a não ser que este ocorra dentro de um sintagma preposicionado; já os depoentes podem se construir de diferentes maneiras, normalmente indicadas (cf. o verbete para *diffiteor* na Figura 9).

[71] Há casos em que o caso do complemento é associado a um tipo semântico, como no verbete para o adjetivo *conscius*, de que se afirma que se constrói com dativo quando se trata de complemento humano ("si referatur ad personam, dat.") ou por genitivo quando se refere a algo inanimado ("ad rem, gen.").

demonstração do quanto a técnica lexicográfica desenvolvida pela tradição manuscrita continuava influente no início do século XVIII ou, nas palavras de Verdelho (1995, p.196), de que "a semiótica da lexicografia moderna foi elaborada ao longo de séculos, na exercitação do texto manuscrito medieval".[72] Por acompanhar o manual oficial do ensino de latim pelos jesuítas, o *Index* representa a metodologia dominante no século XVII e na primeira metade do século XVIII. Todavia, isso não impediu o surgimento de propostas de ensino alternativas ao método gramatical de Álvares e ao método lexicográfico de Velez, como se vê a seguir.

3.2.4 As *Raizes* de Amaro de Roboredo

O predomínio do ensino baseado na gramática de Álvares foi um fator de silenciamento de propostas alternativas para o ensino de línguas que emergiram em Portugal na primeira metade do século XVII (Romeo, 1996).[73] Entre elas, destaca-se a tríade de obras didáticas publicados por Amaro de Roboredo, que correspondem ao seu esforço de adaptar para o português o método de ensino inovador desenvolvido pelos jesuítas irlandeses do Real Colégio de Salamanca, publicado em 1611 sob o título *Ianua linguarum* ('Porta de línguas'). Portanto, antes de apresentar Roboredo, convém fazer uma síntese das características desse método.

A *Ianua* é compilação de 1141 sentenças de temática moral (divididas em doze centúrias, sendo deficitária a décima primeira), numeradas e traduzidas para o espanhol, seguida de um vocabulário latim-espanhol de quase 5 mil entradas, organizado alfabeticamente, em que cada lema é acompanhado por um equivalente de tradução e pelo número remissivo da sentença em que ocorre (Fernandes, 2004). Segundo o prefácio da obra, ela resulta da busca por um método eficaz e universal de ensino de idiomas, que pudesse facilitar o aprendizado das línguas clássicas por estudantes de diferentes línguas maternas bem como auxiliar os missionários espalhados pelo mundo a "aprender as línguas bárbaras […] e espalhar a fé nas regiões dos infiéis" (Bathe, 1611, p.7). Para tanto, seus autores propõem um método novo de ensino-aprendizagem a partir da fusão ("o caminho do meio") dos dois métodos conhecidos à época para o ensino de idiomas: o método "regular", baseado no ensino das regras da gramática; e o método "irregular", baseado na prática direta da leitura e conversação (Bathe, 1611, p.13). Nesse processo, segundo Fernandes (2004, p.169–70), os irlandeses de Salamanca desenvolvem uma espécie de "esboço da teoria do método indutivo nos processos de ensino aprendizagem".

A questão principal gira em torno da impossibilidade de se adquirir vocabulário do mesmo modo com que se aprende a gramática, isto é, memorizando a descrição fornecida pelo ma-

[72] Estudos detalhados sobre o uso metalinguístico de conectores e expressões latinas em textos lexicográficos medievais e do início da era moderna (Elyot) são encontrados, respectivamente, em Merrilees (1990) e Stein (2014).

[73] Romeo (1996) menciona quatro obras gramaticais que, influenciadas pelos métodos experimentalistas de Francis Bacon, assumem uma postura crítica aos métodos alvaresianos, utilizando-se da língua portuguesa para explicar a gramática latina através de métodos contrastivos. São elas: a *Arte de Grammatica para em breve se saber latim, composta em linguagem, e verso portuguez*, de Pedro Sanches de Paredes (1610); o *Methodo grammatical para todas as linguas*, de Amaro de Roboredo (1619); a *Arte de Grammatica Latina Portugueza Benedictina*, de Frutuoso Pereira da Feira (1636); e as *Margens da sintaxe* de João Nunes Freire (1644).

nual (no caso, o dicionário). As causas dessa impossibilidade são identificadas como sendo três (Bathe, 1611, p.14): (1) a existência de "muitas palavras raras, inúteis para o propósito de muitos"; (2) o princípio da derivação, ou melhor, um certo "parentesco" existente entre as palavras que permite saber muitas delas a partir do "conhecimento de uma única fundamental, espécie de fonte das outras"; e, principalmente, (3) o fato de que, do modo como são dispostas no dicionário, "as palavras não têm nenhum sentido", o que prejudica sua fixação na memória "destituída do recurso ao intelecto".

A solução proposta se baseia no enfrentamento dessas dificuldades, por um método igualmente construído em três etapas (Bathe, 1611, p.14). A primeira consiste em "dividir o Calepino em duas partes", para compor um índice apenas com palavras familiares, excluídas as palavras raras (entendidas respectivamente como "as mais frequentes e as menos frequentes"). A segunda consiste em reduzir o índice às palavras "fundamentais", isto é, às formas primitivas, excluindo as formas derivadas. A terceira é a composição de sentenças breves e adequadas à memória, praticamente "sem repetir nenhuma palavra", de modo que sejam representativas do vocabulário mais frequente da língua. Em seguida, a tradução das sentenças para o vernáculo tornaria possível a um único manual atender a um duplo objetivo: de um lado, permitir que o falante do vernáculo aprendesse latim; de outro, permitir que o conhecedor de latim aprendesse uma língua estrangeira (Bathe, 1611, p.21). Por fim, um vocabulário bilíngue, com números remissivos para a sentença em que cada palavra ocorre, seria um último suporte para o aprendizado.[74]

Ainda na década de 1610, o método estabelecido pela *Ianua* de Salamanca foi aplicado para outras línguas, primeiro uma versão latim-inglês por Wiliam Welde em 1615, depois uma versão quadrilíngue latim-inglês-francês-espanhol por Jean Barbier em 1617, contemplando as quatro línguas mais faladas na Europa (Fernandes, 2004, p.179). Em Portugal, como foi dito, o responsável por sua adaptação é Amaro de Roboredo, que o fez através da publicação em separado das duas partes: as sentenças no *Methodo grammatical para todas as linguas*, de 1619; e o vocabulário nas *Raízes da língua latina*, de 1621. Nos dois casos, ele incrementou o material com outros recursos didáticos e de apoio, formando dois conjuntos didáticos interdependentes.

O *Methodo* de Roboredo consta de três partes, cada uma respondendo de forma específica às necessidades de aprendizado das três propriedades que definem qualquer língua, como afirma no prólogo: "a língua consta de Grammatica, Copia e Frase" (Roboredo, 1619, fol. b1r).[75] Assim, a primeira parte é um compêndio de gramática latina explicada em português, com exemplos em ambas as línguas. A segunda reproduz as 1141 sentenças da *Ianua* que, embora não traduzidas,

[74] A ênfase na aquisição do vocabulário não significa que o método não contemple outras partes do ensino de idioma, especialmente a gramática: para seus autores, "tudo o que é prescrito nas regras da gramática pode ser ensinado por meio de sentenças, que desse modo se fixam mais facilmente e rapidamente nas mentes do que através da exposição nua de regras de gramática" (Bathe, 1611, p.19).

[75] Nessa concepção de língua, a gramática é entendida como o conjunto de "regras válidas para todas" as línguas; a cópia e a frase são específicas de cada idioma, consistindo, respectivamente, no conjunto de palavras que forma o léxico e no "particular modo de fallar de cada língua segundo a pronunciação e ajuntamento de palavras, por certa *collocação* delas" (Roboredo, 1619, p.182, grifo adicionado). Para uma discussão aprofundada das ideias linguísticas e pedagógicas de Roboredo, cf. Assunção e Fernandes (2007), Kossárik (1997), Kossarik (2002).

são enriquecidas com um sistema de anotação interlinear de caráter duplo: (1) as palavras são numeradas conforme a ordem natural da sintaxe portuguesa (SVO), de modo a indicar a sequência pela qual a estrutura sintática da frase latina deveria ser abordada (Fernandes, 2004); (2) as palavras recebem uma indicação abreviada das formas canônicas que compõem seus respectivos lemas, viabilizando assim a consulta ao dicionário (e.g. para a forma *valeat* a indicação de lema verbal '-eo,es,ui,itum'; para a forma *encomiis*, a indicação de lema substantivo '-um,i', cf. Roboredo (1619, p.121)). Na terceira parte, enumera e explica diversos aspectos envolvidos no estudo da fraseologia – a saber, sentido metafórico, polissemia e homonímia, elipse, uso excepcional de casos, conjunções e advérbios, uso das preposições –, exemplificando com frases em latim e em português.

As *Raízes da língua Latina*, por sua vez, são a versão ampliada do vocabulário da *Ianua*. O título completo descreve a obra em detalhes:

> Compêndio do Calepino, ou melhor, do *Thesaurus linguae Latinae*, com a tradução portuguesa e espanhola, excluídas as palavras que mais raramente são empregadas. As palavras são apresentadas com os números correspondentes às sentenças da *Ianua* em que ocorrem, e também são indicadas abreviadamente, com base naqueles [dicionários], as quantidades silábicas, as frases mais polidas, a composição a partir das simples, a derivação a partir das primitivas, bem como, eventualmente, os gêneros dos nomes, os pretéritos e os regimes dos verbos (Roboredo, 1621, p.35, tradução própria).[76]

Como o título sugere, Roboredo parte do vocabulário simplificado da *Ianua* e, compilando informações dos dois maiores dicionários monolíngues da época, cria uma espécie de dicionário compacto completo. De fato, ele acrescenta novos itens à nomenclatura e reorganiza o léxico por famílias, posiciona palavras compostas e derivadas como subentradas (marcadas por recuo de parágrafo), fornece informações morfológicas e sintáticas e indica as possibilidades de combinação fraseológica para um bom número de itens.

O texto prefacial, apresentado em português e latim em colunas paralelas, é uma espécie de guia de uso do dicionário. Nele, o autor explica seu método lexicográfico em detalhes, iniciando pelo propósito da obra: "mostrar as raízes da língua Latina, as quais não são outra cousa mais que as palavras simples e primitivas" (Roboredo, 1621, p.1). O objetivo pressuposto é auxiliar o aprendiz a adquirir vocabulário ('cópia de palavras') através de um sistema econômico de apresentação fundamentado na ideia da composicionalidade do significado das palavras. A principal estratégia para atingir esse objetivo passa pelo estudo da "composição, derivação e ortografia das palavras", apresentado na forma de um tratado que precede o dicionário propriamente dito. Esse tratado consiste basicamente em um repertório especializado de prefixos e sufixos (chamados de "preposições que fazem composição" e "as outras partículas componen-

[76] No original em latim: "Compendium Calepini, vel potius Thesauri Linguae Latinae cum interpretatione Lusitanica, et Hispanica omissis vocabulis, quae rarius in usum vocantur. Voces in sententiis Ianuae Linguarum collocatae suis numeris, ipsarum vocum et ab illis colligibilium quantitas et phrasis illustrior cum lima, simplicium compositio, primitivarum derivatio, necnon Nominum genera, Verborum praeterita atque regimina breviter indicantur."

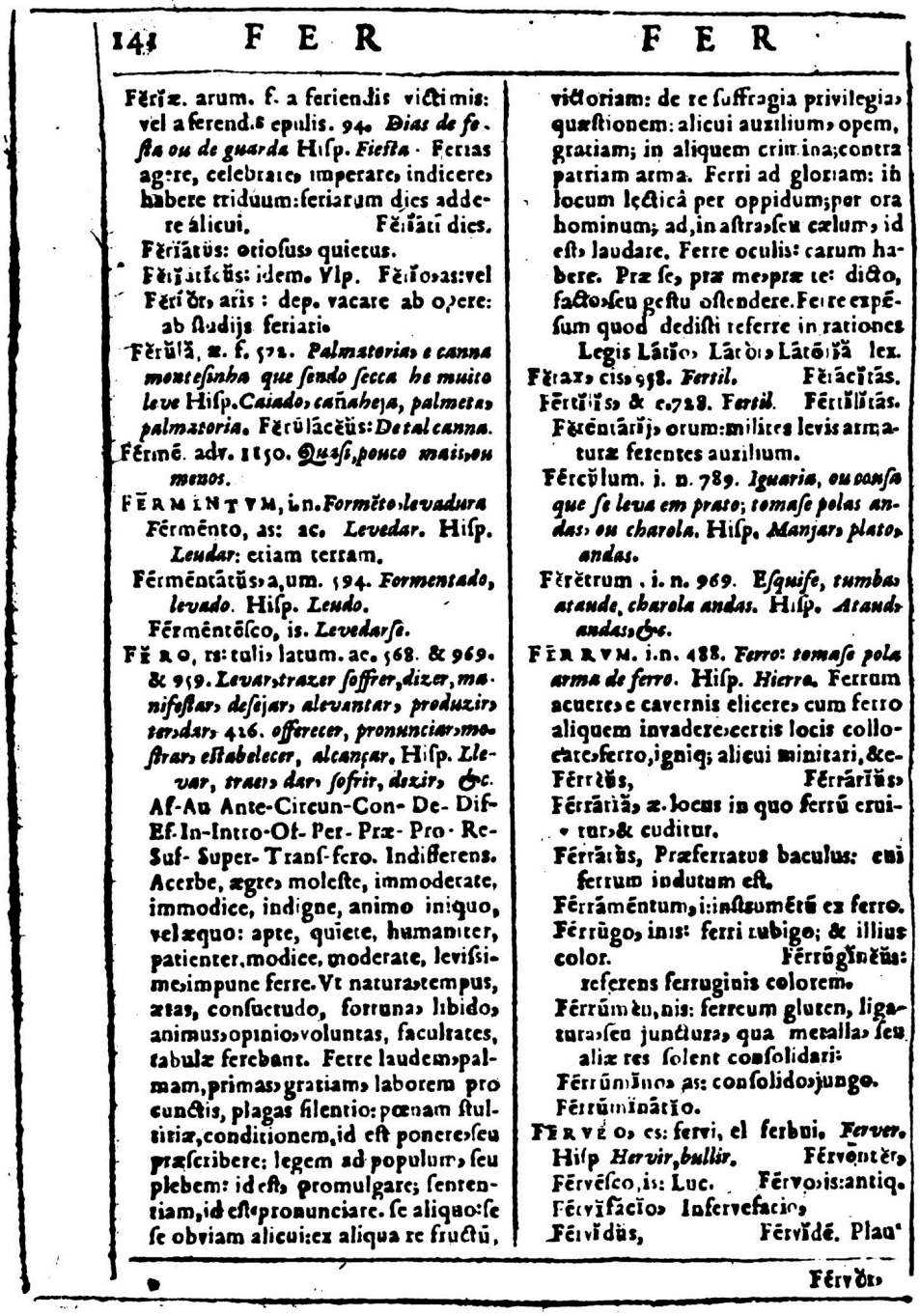
tes", respectivamente),[77] organizados alfabeticamente e descritos quanto a seus usos e significados. O conhecimento prévio desses componentes permite ao usuário deduzir o significado da palavra a partir do significado de seus elementos formadores e, ao compilador, adotar um estilo compacto de apresentação das possibilidades de composição e derivação no interior do verbete. Este consiste na indicação de tais possibilidades por meio de séries de prefixos e sufixos acompanhando cada raiz, que devem ser interpretadas da seguinte maneira: "Ad-De-Per-Red-AMO, isto he, *Adamo, Deamo, Peramo, Redamo*. AM-anter-atio-ator-atorius, etc. quer dizer, *Amanter, Amatio, Amator, Amatorius, etc.*" (Roboredo, 1621, p.1).

O prefácio também explica outras escolhas de Roboredo com relação aos elementos lexicográficos que compõem os verbetes. Sobre os lemas, aponta casos que fogem aos padrões convencionais de enunciação das formas canônicas (e.g. nomes terminados em -us e -is sem indicação de genitivo são adjetivos, verbos da primeira conjugação regular não trazem formas de pretérito e supino, considerados desnecessários). Quanto às acepções, afirma que a significação própria vem sempre em primeiro lugar, seguida pelas significações "a que se estende por semelhança"; caso uma palavra não tenha indicação de significado, este deve ser interpretado com base no significado regular do sufixo, disponível no tratado precedente. Indicações de autoridade são acrescentadas apenas a lemas ou significados raros; o sinal † indica que não se encontrou autoridade latina para tal uso. A fraseologia recebe um tratamento semelhante ao destinado à composição e derivação, com as possibilidades de combinação lexical do lema indicadas "sumariamente, para que o mestre as explique […] como, por exemplo, em *Aculeus* acharás *aculeum habere, emittere, excutere, etc.* e tu dirás *aculeum habere, aculeum emittere, aculeum excutere, etc.*" (Roboredo, 1621, p.3). Nesses casos, Roboredo sugere que o professor ensine aos aprendizes que só usem da significação própria ou da mais usada "até que meditem as metafóricas entretecidas nos autores" (ibid.). Complementos extras de certos verbos são muitas vezes representados por pronomes indefinidos, como em "aemulus alicuius de aliqua re", uma fórmula que, segundo o autor, poderia ser usada para compor em latim a frase "Cartago foi competidora de Roma sobre o mundo" (Roboredo, 1621, p.4). Por fim, informa os padrões tipográficos utilizados, quais sejam: letras "capítulas" (isto é, maiúsculas) são usadas para grafar palavras simples e primitivas ou para destacar as raízes no caso das compostas; a brevidade e a "longura" das vogais são indicadas por sinais gráficos (Ă ă e Ā ā, respectivamente); três níveis de recuo ("três fileiras") são utilizados para indicar subentradas compositivas ou derivativas.

A Figura 10 oferece uma amostra do dicionário de Roboredo. O primeiro verbete tem como lema "Fĕrĭae. arum" e recebe indicação de gênero ("f.", i.e. feminino) e etimologia ("a feriendis victimis vel a ferendis epulis"); em seguida, o número da sentença do *Methodo* ("94.")

[77] Ao contrário dos prefixos, considerados usos especiais das preposições pela gramática antiga, aparentemente não havia uma termo técnico que designasse os sufixos envolvidos nas operações de derivação; na tradição gramatical, é comum que esses sejam tratados em termos de 'letras' ou 'sílabas finais', como se vê, por exemplo, neste excerto do texto de Diomedes, gramático do século IV: "nam quaedam ex nominibus appellatiuis **deriuata in ter syllabam exeunt**, quotiens datiuus casus i littera terminatur, ut molli molliter" (Keil, 1857, p.406, grifo acrescentado).

Figura 10 – Página de *Raizes da lingua Latina* de Amaro de Roboredo.



141 F E R F E R

Fĕrĭæ. arum. f. a feriendis victimis: vel a ferendis epulis. 94. *Dias de feſta ou de guarda* Hiſp. *Fieſta*. Ferias agere, celebrare, imperare, indicere, habere triduum: feriarum dies addere alicui. Fĕrĭātī dies.

Fĕrĭātŭs: otioſus, quietus.

Fĕrĭātĭcŭs: idem. Vlp. Fĕrĭo, as: vel Fĕrĭŏr, aris: dep. vacare ab opere: ab ſtudijs feriari.

Fĕrŭlă, æ. f. 572. *Palmatoria, e canna monteſinha que ſendo ſecca he muito leve* Hiſp. *Caiado, cañaheja, palmeta, palmatoria*. Fĕrŭlăcĕŭs: *De tal canna*.

Fĕrmĕ. adv. 1150. *Quaſi, pouco mais, ou menos*.

FĔRMĔNTVM, i. n. *Formẽto, levadura* Fĕrmĕnto, as: ac. *Levedar*. Hiſp. *Leudar*: etiam terram.

Fĕrmĕntātŭs, a, um. 594. *Formentado, levado*. Hiſp. *Leudo*.

Fĕrmĕntēſco, is. *Levedarſe*.

FĔRO, rs: tuli, latum. ac. 568. & 969. & 959. *Levar, trazer ſoffrer, dizer, manifeſtar, deſejar, alevantar, produzir, tẽr, dar*, 426. *offerecer, pronunciar, moſtrar, eſtabelecer, alcançar*. Hiſp. *Llevar, traer, dar, ſofrir, dezir*, &c.

Af-Au Ante-Circun-Con- De- Dif- Ef- In-Intro-Ob- Per- Præ- Pro- Re- Suf- Super- Tranſ-fero. Indifferens. Acerbe, ægre, moleſte, immoderate, immodice, indigne, animo iniquo, vel æquo: apte, quiete, humaniter, patienter, modice, moderate, leviſsime, impune ferre. Vt natura, tempus, ætas, conſuetudo, fortuna, libido, animus, opinio, voluntas, facultates, tabulæ ferebant. Ferre laudem, palmam, primas, gratiam, laborem pro cunctis, plagas ſilentio: pœnam ſtultitiæ, conditionem, id eſt ponere, ſeu præſcribere: legem ad populum, ſeu plebem: id eſt, promulgare; ſententiam, id eſt, pronunciare. ſe aliquo: ſe ſe obviam alicui: ex aliqua re fructũ, victoriam: de re ſuffragia privilegia, quæſtionem: alicui auxilium, opem, gratiam; in aliquem crimina; contra patriam arma. Ferri ad gloriam: in locum lectica per oppidum; per ora hominum; ad, in aſtra, ſeu cælum, id eſt, laudare. Ferre oculis: carum habere. Præ ſe, præ me, præ te: dicto, facto, ſeu geſtu oſtendere. Ferre expẽſum quod dediſti referre in rationes Legis Lātĭo, Lātŏr, Lātōrĭă lex.

Fĕrax, cis. 958. *Fertil*. Fĕrācĭtās.

Fĕrtĭlis, & e. 728. *Fertil*. Fĕrtĭlĭtās.

Fĕrentārĭj, orum: milites levis armaturæ ferentes auxilium.

Fĕrcŭlum. i. n. 789. *Iguaria, ou conſa que ſe leva em prato; tomaſe polas andas, ou charola*. Hiſp. *Manjar, plato, andas*.

Fĕrĕtrum. i. n. 969. *Eſquife, tumba, atande, charola andas*. Hiſp. *Atande, andas*, &c.

FĔRRVM. i. n. 488. *Ferro: tomaſe pola arma de ferro*. Hiſp. *Hierro*. Ferrum acuere, e cavernis elicere, cum ferro aliquem invadere, certis locis collocare, ferro, igniq; alicui minitari, &c.

Fĕrrĕŭs, Fĕrrārĭŭs.

Fĕrrārĭă, æ. locus in quo ferrũ eruitur, & cuditur.

Fĕrrātŭs, Præferratus baculus: cui ferrum indutum eſt.

Fĕrrāmĕntum, i: inſtrumẽtũ ex ferro.

Fĕrrūgo, inis: ferri rubigo; & illius color. Fĕrrūgĭnĕŭs: referens ferruginis colorem.

Fĕrrūmĕn, nis: ferreum gluten, ligatura, ſeu junctura, qua metalla, ſeu aliæ res ſolent conſolidari:

Fĕrrūmĭno, ās: conſolido, jungo.

Fĕrrūmĭnātĭo.

FĔRVĕO, es: fervi, el ferbui. *Ferver*. Hiſp *Hervir, bullir*. Fĕrvēntĕr, Fĕrvēſco, is: Luc. Fĕrvo, is: antiq. Fĕrvĭfăcĭo, Infervefacio, Fĕrvĭdŭs, Fĕrvĭdē. Plau-

Fĕrvŏr,

Fonte: (Roboredo, 1621, p.141).

e a tradução para o português ("*Dias de festa ou de guarda*") e para o espanhol ("Hisp. *Fiesta*"); seguem-se as combinações lexicais, começando com a forma acusativa servindo como complemento verbal ("Ferias agere, celebrare, imperare…") ou nominal ("feriarum dies addere alicui"); por fim, uma série de derivados acompanhados de explicação do significado em latim (e.g. "Feriatus: otiosus, quietus"). A produtividade de certos lemas resulta em longos verbetes, como é o caso do verbo *fero*, que ocupa metade da primeira coluna e o quarto inicial da segunda; nesse caso, o lema é composto de mais formas (incluindo perfeito e supino), conta com informação sobre regência ('ac.', i.e. rege acusativo),[78] tem seu significado explicado por uma série de equivalentes representando diversas acepções (em itálico), para então ter indicadas suas possibilidades de composição – representadas respectivamente por uma série de 18 prefixos ("Af-Au-Ante-Circun-…") – e de combinação fraseológica, organizadas por forma sintática: são 16 advérbios; 11 substantivos no nominativo, 22 substantivos no acusativo (alguns acompanhados de outros complementos e explicação do significado, e.g. "[ferre] sententiam, id est, pronunciare"), além de sete sintagmas preposicionados.

Pode-se dizer que o esquema de organização empregado por Roboredo não só permite repertoriar um grande número de palavras em um espaço reduzido, mas também favorece a associação mental entre os diversos itens originados da mesma raiz, auxiliando na memorização. O autor publicou, ainda, uma versão completa da *Ianua* traduzida para o português, intitulada *Porta de línguas* (Roboredo, 1623). A despeito das promessas de aprendizado rápido e seguro, nenhuma de suas obras foi reeditada em Portugal, o que indica que o método universal não conseguiu fazer frente ao domínio da metodologia de ensino baseado em gramática.

3.2.5 A *Prosodia* de Bento Pereira

Os pequenos dicionários de Velez e Roboredo eram vocabulários de métodos de ensino e não tinham a pretensão de substituir o *Dictionarium* de Cardoso na função de dicionário geral latim-português. Essa tarefa coube a Bento Pereira, padre jesuíta e professor de retórica e teologia na Universidade de Évora. Sua produção bibliográfica, profundamente vinculada às necessidades pedagógicas da escola, inclui compêndios (*promptuaria*) de religião e teologia, estudos jurídicos e morais; uma gramática de português para estrangeiros escrita em latim;[79] e, especialmente, o conjunto lexicográfico conhecido como *Prosodia*. Constituído de dois dicionários bilíngues (latim-português e português-latim) e uma terceira parte dedicada à fraseologia, esse conjunto teve ampla difusão por mais de um século, tornando-se uma espécie de *vade-mecum* de estudantes e professores que nele buscavam suporte tanto para a leitura quanto para a escrita de textos em latim (Verdelho, 1982).

Assim como ocorreu com o *Index* de Velez, visto há pouco, a compilação da *Prosodia*

[78] Conforme lista de abreviaturas informada após o prefácio, cf. Roboredo (1621, p.5).

[79] Publicada em 1672 com o título *Ars grammaticae pro lingua lusitana addiscenda latino Idiomate proponitur*, foi escrita durante a estadia de Pereira em Roma, onde desempenhou a função de revisor geral dos livros da Companhia de Jesus (Fernandes, 2008).

também foi marcada por diversas revisões e ampliações ao longo do século XVII. Sua história editorial pode ser dividida em três ciclos distintos (Freire Cameron, 2012). O primeiro se inicia em 1634, com a publicação, em Évora, do dicionário trilíngue latim-português-espanhol (*Prosodia in Vocabularium trilingue, Latinum, Lusitanicum et Hispanicum digesta*), custeada pelo irmão do autor,[80] e vai até a sua sexta edição, publicada em Lisboa em 1683, dois anos após sua morte; durante esse período, agregam-se sucessivamente ao volume o dicionário português-latim[81] (2.ed., Lisboa, 1553) e a seção sobre fraseologia [82] (3.ed., Lisboa, 1661). O segundo começa em 1697, com a publicação de uma versão bilíngue latim-português (*Prosodia in Vocabularium bilingue, Latinum et Lusitanum digesta*), renovada e ampliada por uma equipe de professores da Universidade de Évora, sob a coordenação de Matias de São Germano; essa versão se reproduz nas edições e reimpressões seguintes, todas realizadas pela tipografia da Academia de Évora, até a última edição, em 1750. O terceiro ciclo, por fim, consiste em dois esforços de reformulação da *Prosodia*, empreendidos de forma simultânea em meados do século XVIII pelo padre José Caeiro e pelos lexicógrafos de Évora, ambos interrompidos pela expulsão dos jesuítas de Portugal em 1759; sua repercussão na produção lexicográfica posterior será abordada mais adiante.

Desde o início, Pereira se mostra consciente de sua missão de substituir o dicionário de Cardoso por uma obra mais ampla. Em um prefácio ao leitor escrito em latim (*Ad lectorem*),[83] ele menciona entre suas motivações a necessidade de suprir a "penúria de vocabulário com que os portugueses trabalhavam" e afirma a superioridade de sua obra ao comparar o tamanho das nomenclaturas de Cardoso ("22.167 vocábulos") com as "cerca de 50 mil palavras" contempladas pela *Prosodia*. Também na folha de rosto da primeira edição a obra se apresenta como "totalmente necessária a professores de letras, médicos, juristas e estudiosos de todas as faculdades",[84] e especifica as três vantagens (*utilia*) da compilação:

1. o significado das palavras foi traduzido para o português e o espanhol mediante consulta

[80] Segundo informa a folha de rosto da obra, reproduzida por Freire Cameron (2012, p.71): "Fecit sumptus Dominicus Pereyra da Sylva authoris frater". Autores pagarem pela publicação de suas obras era prática comum no período, de modo que o fato de a *Prosodia* ter sido publicada, já a partir da segunda edição, às expensas dos tipógrafos é um indício de seu sucesso editorial.

[81] O dicionário português-latim, intitulado (*Thesouro da Lingoa Portuguesa*) tinha sido publicado isoladamente em 1647. Com aproximadamente 24 mil entradas, trata-se de uma nomenclatura portuguesa mais ampla do que se encontra nos dicionários anteriores de Cardoso e Barbosa (Freire Cameron, 2012, p.17).

[82] Antes de se integrar ao volume da *Prosodia*, o repertório fraseológico intitulado *Florilegio dos modos de fallar e adagios da lingua portugueza*) havia sido publicado isoladamente em 1655. Considerado "um dicionário contextual bilíngue" português-latim (Schäfer-Prieß, 2019, p.22), fornece três tipos de modelo para a composição retórica de textos em latim: (1) um repertório de "frases portuguesas" traduzidas com "as mais puras e elegantes frases latinas", isto é, uma compilação de milhares de expressões e modos de falar organizados em torno de conceitos-chave (aproximadamente 600 tópicos), dispostos em ordem alfabética e referentes principalmente a valores morais da época (Verdelho, 1982); (2) um repertório de adágios portugueses acompanhados de suas versões latinas; (3) exemplos de descrições, em latim, compostas pelo autor ou coligidas de "autores aprovados", também organizadas alfabeticamente por conceitos-chave.

[83] Cf. texto latino reproduzido por Freire Cameron (2012, p.50).

[84] No original: "Opus omnino necessarium professoribus sacrarum, et humaniorum litterarum, medicis, iuristis, et omnibus cuiuscunque facultatis studiosis" (transcrito por Freire Cameron (2012, p.70), tradução própria).

a peritos em língua materna, farmacêuticos experientes e todo tipo de profissional com grande autoridade em seu ofício;

2. a quantidade das sílabas de cada palavra é estabelecida com base em critérios como ocorrência nos poemas, derivação e composição, origem grega ou hebraica e, em último caso, no julgamento dos homens doutos;

3. a inclusão de muitas palavras frequentes nos textos eclesiásticos e jurídicos ou usadas por autores clássicos que são negligenciadas tanto pelo Calepino quanto pelos compiladores do *Thesaurus* e de quaisquer outros vocabulários.

De fato, Pereira parece orientado por um ideal de exaustividade, que se aplica tanto à nomenclatura quanto, especialmente, à informação sobre pronúncia (daí o título "prosódia"). No caso da nomenclatura, constituem evidência as declarações, veiculadas nas folhas de rosto e nos textos prefaciais das sucessivas edições,[85] de que o número de palavras descritas não só era bem maior do que aquele encontrado em Cardoso, mas também contemplava vocábulos ausentes em qualquer dicionário da época, incluindo o Calepino e o próprio *Thesaurus Linguae Latinae* (ed. 1551), principal referência de dicionário monolíngue de latim: na primeira edição fala-se em 5 mil palavras exclusivas, marcadas na nomenclatura por um asterisco (*); a segunda traria outras 12 mil, sinalizadas com asterisco (*) estas e as anteriores com uma cruz (†); da terceira à sexta teriam sido "milhares", com o asterisco sendo usado para marcar os novos acréscimos. No caso da informação sobre pronúncia, a obra pretenderia responder a certa expectativa da "república literária" por uma obra que considerasse as quantidades de todas as sílabas de cada palavra.

O método utilizado por Pereira para realizar esse enorme projeto de compilação é a leitura, igualmente exaustiva, de (1) "todos os dicionários de latim" (e menciona o *Thesaurus*, o Calepino, vocabulários especializados em termos eclesiásticos, médicos e jurídicos, bem como bilíngues latim-francês, inglês, italiano, espanhol e português); (2) "todos os livros que tratam de quantidade das sílabas" (e menciona Manuel Álvares, Despautério, Nebrija, Textor e certos "manuscritos de Velez"); e, por fim, (3) "todos os poetas antigos e alguns recentes de renome", os quais não são especificados no prefácio mas constam do interior de alguns verbetes, junto com um verso que serve de abonação para as quantidades silábicas do lema em questão.

Esse ideal se mantém na sétima edição, publicada em 1697 e realizada pela equipe de lexicógrafos de Évora. Em primeiro lugar, há um novo aporte de 24 mil palavras a partir da consulta a diversas fontes,[86] incluindo novos itens pertencentes às respectivas famílias lexicais, ao vocabulário enciclopédico (antropônimos e topônimos) e terminologia especializada (Freire

[85] Reproduzidos em Freire Cameron (2012, pp.71–107).

[86] Um novo texto prefacial, assinado coletivamente pela "Academia Eborense", menciona um trabalho de compilação feito "a partir dos latinos, a saber, os latino-gregos, os latino-bárbaros, os críticos, os antiquários, os tesouros, os léxicos, os onomásticos, os glossários e os químicos, como a partir daqueles que cultivam o conhecimento, ambas as jurisprudências, a medicina e outras ciências" (Freire Cameron, 2012, p.53–54).

Cameron, 2012, p.114), bem como palavras sem abonação nos autores clássicos, indicadas por um asterisco para que sejam usadas "com cautela" (*caute usurpanda*, cf. Pereira (1697, rosto)). Além disso, são acrescentadas sistematicamente, ao final de cada verbete, a indicação abreviada da fonte de informação sobre o lema e a descrição apresentada (especialmente a quantidade silábica); tais abreviaturas são explicitadas em um catálogo de fontes, localizado entre os textos prefaciais, contendo os nomes dos 314 autores consultados, que incluem não apenas autores clássicos da literatura como Cícero, Virgílio, Ovídio, Horácio, Plauto, mas também a *Bíblia*, grandes dicionários contemporâneos como a *Amalthea Onomastica* de Laurenzi (ed. 1664), o Calepino, o *Graecum lexicon* e até compiladores medievais como Papias e Isidoro (Freire Cameron, 2012, p.140). Por fim, a explicação do significado é renovada mediante a remoção completa dos termos em espanhol, ao passo que as glosas portuguesas foram renovadas e ampliadas, não só pela inclusão de equivalentes para as novas entradas, mas também pela inclusão de diversas novas acepções para as formas já existentes (Freire Cameron, 2012, p.114).

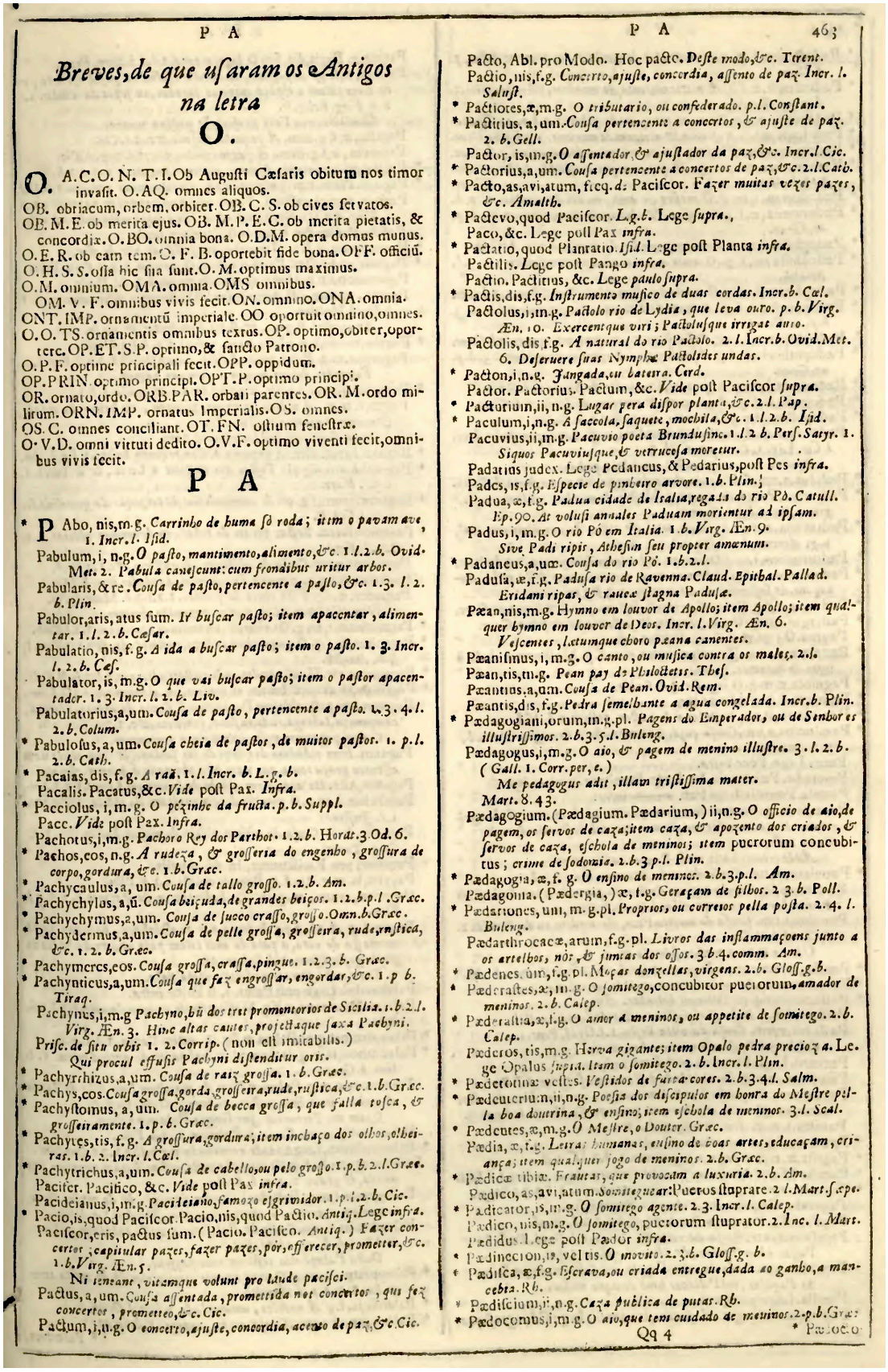
O resultado é um volume de 736 páginas em formato in-folio (22 x 32 cm), com aproximadamente 76 mil entradas organizadas semi-alfabeticamente (já que palavras primitivas normalmente antecedem suas derivadas) e distribuídas em duas colunas, separadas por uma linha contínua e encabeçadas pelas duas primeiras letras da palavra que as inicia, como se observa na Figura 11. Tal quantidade de entradas, considerada extraordinária para os padrões dicionarísticos da época (Freire Cameron, 2012), é formada majoritariamente de substantivos (44,5 mil entradas, ou 60%),[87] incluindo muitos nomes próprios e estrangeirismos (especialmente palavras gregas do vocabulário técnico das ciências naturais). Mesmo as expressões multipalavra constam principalmente de substantivos compostos, ao contrário de Cardoso e Velez, que dedicavam mais espaço às expressões com núcleo verbal. Além disso, cada seção dedicada a uma letra do alfabeto termina com um bloco específico para itens lexicais do tipo abreviatura (cf. a seção "Breves de que usaram os Antigos na letra O", no alto da Figura 11).

A microestrutura da *Prosodia* tem, como elementos fundamentais: lema, informação gramatical, explicação do significado e informação prosódica.

O lema segue o padrão já estabelecido, sendo composto por um conjunto de formas canônicas segundo a classe de palavras: para substantivos, o nominativo e o genitivo singular abreviado (e.g. 'Pabo, nis', 'Pabulum, i'); para adjetivos, o nominativo singular de cada gênero, quando este se distingue morfologicamente (e.g. 'Pabulosus, a, um', 'Pabularis, & re') ou nominativo e genitivo quando o gênero não se distingue pela forma (e.g. 'Pachos, eos'); para verbos, a sequência de formas no indicativo singular, composta pela primeira e segunda pessoas do presente, a primeira do perfeito e a forma nominal do supino (e.g. 'Pacto, as, avi, atum'). Quase 6 mil entradas possuem formas variantes do lema, eventualmente demarcadas entre parênteses, por um conectivo latino (e.g. *et*, *sive*, *vel*) ou, em algumas dezenas de casos, pela conjunção portuguesa

[87] Os dados quantitativos apresentados a seguir foram computados a partir da versão em texto digital disponível no *Corpus Lexicográfico do Português* (Pereira, 2002), mediante processo de estruturação parcial dos dados, descrito em detalhes no Capítulo 3.

Figura 11 – Página da *Prosodia in Vocabularium bilingue Latinum et Lusitanum* de Bento Pereira



P A

Breves, de que uſaram os Antigos na letra

O.

O. A. C. O. N. T. I. Ob Auguſti Cæſaris obitum nos timor invaſit. O. AQ. omnes aliquos.
OB. obriacum, orbem, orbiter. OB. C. S. ob cives ſervatos.
OB. M. E. ob merita ejus. OB. M. P. E. C. ob merita pietatis, & concordiæ. O. BO. omnia bona. O. D. M. opera domus munus.
O. E. R. ob eam rem. O. F. B. oportebit fide bona. OFF. officiũ.
O. H. S. S. oſſa hic ſita ſunt. O. M. optimus maximus.
O. M. omnium. OMA. omnia. OMS omnibus.
OM. V. F. omnibus vivis fecit. ON. omnino. ONA. omnia.
ONT. IMP. ornamentũ imperiale. OO oportuit omnino, omnes.
O. O. TS. ornamentis omnibus textus. OP. optimo, obiter, oportere. OP. ET. S. P. optimo, & ſanctо Patrono.
O. P. F. optime principali fecit. OPP. oppidum.
OP. PRIN optimo principi. OPT. P. optimo principi.
OR. ornato, ordo. ORB. PAR. orbati parentes. OR. M. ordo militum. ORN. IMP. ornatus Imperialis. OS. omnes.
OS. C. omnes conciliant. OT. FN. oſtium feneſtræ.
O. V. D. omni virtuti dedito. O. V. F. optimo viventi fecit, omnibus vivis fecit.

P A

* PAbo, nis, m. g. *Carrinho de huma ſó roda; item o pavam ave*, 1. *Incr. l. Iſid.*
Pabulum, i, n. g. *O paſto, mantimento, alimento, &c.* 1. *l.* 2. *b. Ovid. Met.* 2. *Pabula caneſcunt: cum frondibus uritur arbor.*
Pabularis, & re. *Couſa de paſto, pertencente a paſto, &c.* 1. 3. *l.* 2. *b. Plin.*
Pabulor, aris, atus ſum. *Ir buſcar paſto; item apacentar, alimentar.* 1. *l.* 2. *b. Cæſar.*
Pabulatio, nis, f. g. *A ida a buſcar paſto; item o paſto.* 1. 3. *Incr. l.* 2. *b. Cæſ.*
Pabulator, is, m. g. *O que vai buſcar paſto; item o paſtor apacentador.* 1. 3. *Incr. l.* 2. *b. Liv.*
Pabulatorius, a, um. *Couſa de paſto, pertencente a paſto.* 1. 3. 4. *l.* 2. *b. Colum.*
* Pabuloſus, a, um. *Couſa cheia de paſtos, de muitos paſtos.* 1. *p. l.* 2. *b. Cath.*
* Pacaias, dis, f. g. *A raã.* 1. *l. Incr. b. L. g. b.*
Pacalis. Pacatus, &c. *Vide* poſt Pax. *Infra.*
* Pacciolus, i, m. g. *O pezinho da fructa. p. b. Suppl.*
Pacc. *Vide* poſt Pax. *Infra.*
Pachorus, i, m. g. *Pachoro Rey dos Parthos.* 1. 2. *b. Horat.* 3 *Od.* 6.
* Pachos, cos, n. g. *A rudeza, & groſſeria do engenho, groſſura de corpo, gordura, &c.* 1. *b. Græc.*
* Pachycaulus, a, um. *Couſa de tallo groſſo.* 1. 2. *b. Am.*
* Pachychylus, a, ũ. *Couſa beiçuda, de grandes beiços.* 1. 2. *b. p. l. Græc.*
* Pachychymus, a, um. *Couſa de ſucco craſſo, groſſo. Omn. b. Græc.*
* Pachydermus, a, um. *Couſa de pelle groſſa, groſſeira, rude, rnſtica, &c.* 1. 2. *b. Græc.*
* Pachymeres, eos. *Couſa groſſa, craſſa, pingue.* 1. 2. 3. *b. Græc.*
* Pachynticus, a, um. *Couſa que faz engroſſar, engordar, &c.* 1. *p b. Tiraq.*
Pachynus, i, m. g *Pachyno, hũ dos tres promontorios de Sicilia.* 1. *b.* 2. *l. Virg. Æn.* 3. *Hinc altas cautes, projectaque ſaxa Pachyni.*
Priſc. de ſitu orbis 1. 2. *Corrip.* (non eſt imitabilis.)
Qui procul effuſis Pachyni diſtenditur oris.
* Pachyrrhizus, a, um. *Couſa de raiz groſſa.* 1. *b. Græc.*
* Pachys, cos. *Couſa groſſa, gorda, groſſeira, rude, ruſtica, &c.* 1. *b. Græc.*
* Pachyſtomus, a, um. *Couſa de becca groſſa, que falla toſca, & groſſeiramente.* 1. *p. b. Græc.*
* Pachytes, tis, f. g. *A groſſura, gordura; item inchaço dos olhos, olheiras.* 1. *b.* 2. *Incr. l. Cæl.*
* Pachytrichus, a, um. *Couſa de cabello, ou pelo groſſo.* 1. *p. b.* 2. *l. Græc.*
Pacifer. Pacifico, &c. *Vide* poſt Pax *infra.*
Pacidianus, i, m. g. *Pacidiano, famozo eſgrimidor.* 1. *p. l.* 2. *b. Cic.*
* Pacio, is, quod Paciſcor. Pacio, nis, quod Pactio. *Antiq.* Lege *infra.*
Paciſcor, eris, pactus ſum. (Pacio. Paciſco. *Antiq.*) *Fazer concerto, capitular pazes, fazer pazes, pôr, offerecer, prometter, &c.* 1. *b. Virg. Æn.* 5.
Ni teneant, utcumque volunt pro laude paciſci.
Pactus, a, um. *Couſa aſſentada, promettida nos concertos, que fez concerto, prometteo, &c. Cic.*
Pactum, i, n. g. *O concerto, ajuſte, concordia, aceſſo de paz, &c. Cic.*

P A 463

Pacto, Abl. pro Modo. Hoc pacto. *Deſte modo, &c. Terent.*
Pactio, nis, f. g. *Concerto, ajuſte, concordia, aſſento de paz. Incr. l. Saluſt.*
* Pactiotes, æ, m. g. *O tributario, ou confederado. p. l. Conſtant.*
* Pactitius, a, um. *Couſa pertencente a concertos, & ajuſte de paz.* 2. *b. Gell.*
Pactor, is, m. g. *O aſſentador & ajuſtador da paz, &c. Incr. l. Cic.*
* Pactorius, a, um. *Couſa pertencente a concertos de paz, &c.* 2. *l. Cath.*
* Pacto, as, avi, atum, freq. de Paciſcor. *Fazer muitas vezes pazes, &c. Amalth.*
* Pactevo, quod Paciſcor. *L. g. b.* Lege *ſupra.*,
Paco, &c. Lege poſt Pax *infra.*
* Pactatio, quod Plantatio. *Iſid.* Lege poſt Planta *infra.*
Pactilis. Lege poſt Pango *infra.*
Pactio. Pactitius, &c. Lege *paulo ſupra.*
* Pactis, dis, f. g. *Inſtrumento muſico de duas cordas. Incr. b. Cæl.*
Pactolus, i, m. g. *Pactolo rio de Lydia, que leva ouro. p. b. Virg. Æn.* 10. *Exercentque viri; Pactoluſque irrigat aurо.*
Pactolis, dis, f. g. *A natural do rio Pactolo.* 2. *l. Incr. b. Ovid. Met.* 6. *Deſeruere ſuas Nymphæ Pactolides undas.*
* Pacton, i, n. g. *Jangada, ou bateira. Cord.*
Pactor. Pactorius. Pactum, &c. *Vide* poſt Paciſcor *ſupra.*
* Pacturium, ii, n. g. *Lugar pera diſpor planta, &c.* 2. *l. Pap.*
* Paculum, i, n. g. *A ſaccola, ſaquete, mochila, &c.* 1. *l.* 2. *b. Iſid.*
Pacuvius, ii, m. g. *Pacuvio poeta Brunduſino.* 1. *l.* 2 *b. Perſ. Satyr.* 1. *Siquos Pacuviuſque, & verruceſa moretur.*
Padarius judex. Lege Pedaneus, & Pedarius, poſt Pes *infra.*
Pades, is, f. g. *Eſpecie de pinheiro arvore.* 1. *b. Plin.*
Padua, æ, f. g. *Padua cidade de Italia, regada do rio Pô. Catull. Ep.* 90. *At voluſi annales Paduam morientur ad ipſam.*
Padus, i, m. g. *O rio Pó em Italia.* 1. *b. Virg. Æn.* 9.
Sive Padi ripis, Atheſim ſeu propter amœnum.
* Padaneus, a, uœ. *Couſa do rio Pó.* 1. *b.* 2. *l.*
Paduſa, æ, f. g. *Paduſa rio de Ravenna. Claud. Epithal. Pallad. Eridani ripas, & rauceæ ſtagna Paduſæ.*
Pæan, nis, m. g. *Hymno em louvor de Apollo; item Apollo; item qualquer hymno em louvor de Deos. Incr. l. Virg. Æn.* 6.
Veſcentes, lætumque choro pæana canentes.
Pæaniſmus, i, m. g. *O canto, ou muſica contra os males.* 2. *l.*
Pæan, tis, m. g. *Pean pay de Philoctetis. Theſ.*
Pæantius. a, um. *Couſa de Pean. Ovid. Rem.*
Pæantis, dis, f. g. *Pedra ſemelhante a agua congelada. Incr. b. Plin.*
* Pædagogiani, orum, m. g. pl. *Pagens do Emperador, ou de Senhores illuſtriſſimos.* 2. *b.* 3. 5. *l. Buleng.*
Pædagogus, i, m. g. *O aio, & pagem de menino illuſtre.* 3. *l.* 2. *b. (Gall.* 1. *Corr. per, e.)*
Me pedagogus adit, illam triſtiſſima mater.
Mart. 8. 43.
Pædagogium. (Pædagium. Pædarium,) ii, n. g. *O officio de aio, de pagem, os ſervos de caza; item caza, & apozento dos criados, & ſervos de caza, eſchola de meninos; item* pucrorum concubitus; *crime de ſodomia.* 2. *b.* 3 *p. l. Plin.*
* Pædagogia, æ, f. g. *O enſino de meninos.* 2. *b.* 3. *p. l. Am.*
Pædagogia. (Pædergia,) æ, f. g. *Geraçam de filhos.* 2 3. *b. Poll.*
* Pædariones, uni, m. g. pl. *Proprios, ou correios pella poſta.* 2. 4. *l. Buleng.*
Pædarthrocaceæ, arum, f. g. pl. *Livros das inflammaçoens junto a os artelhos, nôs, & juntas dos oſſos.* 3 *b.* 4. *comm. Am.*
* Pædenes. ũm, f. g. pl. *Moças donzellas, virgens.* 2. *b. Gloſſ. g. b.*
* Pæderaſtes, æ, m. g. *O ſomitego, concubitor puerorum, amador de meninos.* 2. *b. Calep.*
* Pæderaſtia, æ, f. g. *O amor a meninos, ou appetite de ſomitego.* 2. *b. Calep.*
Pæderos, tis, m. g. *Herva gigante; item Opalo pedra preciоſa.* Lege Opalus *ſupra. Item o ſomitego.* 2. *b. Incr. l. Plin.*
* Pædetotinæ veſtes. *Veſtidos de furta-cores.* 2. *b.* 3. 4. *l. Salm.*
* Pædeuterium, ii, n. g. *Poeſia dos diſcipulos em honra do Meſtre pella boa doutrina, & enſino; item eſchola de meninos.* 3. *l. Scal.*
* Pædeutes, æ, m. g. *O Meſtre, o Doutor. Græc.*
Pædia, æ, f. g. *Letras humanas, enſino de boas artes, educaçam, criança; item qualquer jogo de meninos.* 2. *b. Græc.*
* Pædicæ tibiæ. *Frautas, que provocam a luxuria.* 2. *b. Am.*
Pædico, as, avi, atum. *Somitegucar:* Pueros ſtuprare. 2 *l. Mart. ſæpe.*
* Pædicator, is, m. g. *O ſomitego agente.* 2. 3. *Incr. l. Calep.*
Pædico, nis, m. g. *O ſomitego,* puerorum ſtuprator. 2. *Inc. l. Mart.*
Pædidus. Lege poſt Pædor *infra.*
* Pædinection, is, vel tis. *O movito.* 2. 3. *b. Gloſſ. g. b.*
* Pædiſca, æ, f. g. *Eſcrava, ou criada entregue, dada ao ganho, a mancebia. Rb.*
* Pædiſcium, ii, n. g. *Caza publica de putas. Rb.*
* Pædocomus, i, m. g. *O aio, que tem cuidado de meninos.* 2. *p. b. Græc.*

Fonte: (Pereira, 1697, p.463). A imagem se encontra reduzida para aprox. 80% de seu tamanho real.

'ou' (quase todos na letra A, e.g. 'Aethiops, is, ou Aethiopus, i').

A informação gramatical é sistemática apenas para substantivos e palavras invariáveis. Os primeiros recebem indicação de gênero por meio de abreviaturas dos termos latinos (*masculini*, *feminini* e *neutri generis*, abreviados 'm.g.', 'f.g.' e 'n.g.', respectivamente); eventualmente, essa informação é acompanhada de informação derivativa (e.g. Abaculus, i, m. g. dimin. De Abacus'), contrastiva (quando o gênero da palavra grega é distinto) ou restritiva (geralmente para indicar palavras mais usada no plural ou indeclináveis). As segundas são caracterizadas pela abreviatura referente à classe, muitas vezes acompanhadas por uma caracterização adicional (e.g. 'Adv.', 'praep. accus.', 'conjunct. causal.', 'Interj.'). Verbos e adjetivos recebem muito raramente informação gramaticais, aparentemente usadas apenas para formas derivadas (verbos frequentativos e incoativos, adjetivos nos graus comparativo e superlativo) ou casos especiais (verbos impessoais e defectivos, adjetivos participiais ou usados apenas no plural).

A explicação do significado é feita normalmente por meio de definições ou equivalentes em português; há poucos casos em que a informação semântica é veiculada ou complementada em latim, em geral por pudor do lexicógrafo diante de uma palavra-tabu (Silvestre, 2003). Algumas palavras recebem séries de equivalentes separados por pontuação (vírgula e ponto-e-vírgula), que, segundo Freire Cameron (2012, p.119), devem ser interpretados como diferentes acepções do lema; não raro, porém, as diferentes acepções são delimitadas pela partícula latina *item* ('também'). A autora aponta ainda alguns padrões de explicação do significado associados a determinadas classes de palavras: para adjetivos, o uso da forma instrumental "cousa" seguida por um ou mais adjetivos em português; para substantivos comuns, além das séries sinonímicas, também ocorrem definições construídas com recurso a frases ou complementos preposicionais; já para topônimos ou nomes próprios com significação histórico-cultural, a definição pode ser mais explicativa e mais alongada, seguindo a vocação escolar e enciclopédica da obra; por fim, verbos são descritos simplesmente por meio de equivalentes, que recobrem suas várias acepções, geralmente ordenadas a partir do significado mais comum.

Por fim, a informação prosódica está presente na maior parte das entradas e é indicada por meio de uma combinação de dois tipos de abreviaturas: o primeiro indica a posição relativa de cada sílaba na palavra, representada por um dos seguintes símbolos, '1.2.3.p.Inc.', que significam, respectivamente, primeira, segunda, terceira, penúltima e sílaba incremental (isto é, a sílaba extra que os nomes da terceira declinação têm no genitivo em relação ao nominativo); o segundo indica a quantidade, representada por 'b.' (breve), 'l.' (longa) ou 'comm.' (comum); assim, o lema 'Pabulatio, nis' tem a quantidade descrita pela expressão "1.3.Incr.l.2.b", indicando que as sílabas primeira, terceira e incremental ('nis') são longas e a segunda sílaba é breve. Além disso, praticamente todas as indicações são abonadas por uma referência bibliográfica, em alguns casos acompanhada de um verso cuja escansão permite comprovar a quantidade silábica informada. Quanto a marcas de uso, por fim, a principal é o asterisco posicionado à esquerda das entradas para indicar os itens de vocabulário não clássico; outras marcas de uso existem mas não chegam a duas centenas, podendo ser de tipo diacrônico ('antiq.' indica um arcaísmo),

diatécnico (‘Eccles.’ indica palavra usada por autores eclesiásticos), diafrequente (‘inusit.’ indica uma palavra desusada), ou então marcar a origem dos estrangeirismos (‘Arab.’, ‘Graec.’, ‘Hebr.’), entre outras possibilidades menos frequentes. Informações sintagmáticas, por sua vez, são raríssimas.

A nova versão da *Prosodia* se fixou como grande dicionário latim-português na primeira metade do século XVIII. Em meados do século, porém, passou a ser criticada tanto por suas dimensões excessivas, tidas como inapropriadas para o uso dos estudantes, e principalmente pela quantidade de erros encontrados. Dois textos são especialmente relevantes. O primeiro é do educador Luís António Verney, que, em sua dissertação sobre *O verdadeiro método de estudar* (1746), aponta como erros da *Prosodia* seu uso inconsistente do asterisco para assinalar os termos bárbaros, o fato de não distinguir “as idades dos vocábulos”, não explicar “a força das vozes” nem mostrar “com exemplos os significados próprios e figurados das palavras”, e sugere sua substituição da por um “dicionário pequeno para os rapazes”, ainda por ser feito (Borges, 2008, p.1082). O segundo é de Antonio Pereira de Figueiredo, professor de latim e retórica e crítico dos métodos jesuítas,[88] cujo *Apparato critico para a correcção do diccionario intitulado Prosodia…* (1755) pode ser considerado a primeira crítica sistemática de um dicionário publicada em Portugal (Almeida, 1969); nesse opúsculo, Figueiredo reconhece a dificuldade de se compilar um léxico e, a fim de contribuir com a nova edição em preparação na Universidade de Évora, elenca doze pontos em que a *Prosodia* é falha, entre os quais estão as “muitas palavras bárbaras” incluídas na macroestrutura, a começar pelo título: para Figueiredo, “o nome *Vocabularium* é indubitavelmente bárbaro”, sugerindo em seu lugar o uso “da voz grega”, referindo-se provavelmente a *lexicon* (Almeida, 1969, p.30).

Como foi dito anteriormente, os jesuítas já estavam trabalhando paralelamente em duas reformulações da *Prosodia* (Silvestre; Borges, 2012), ambas com o objetivo de resolver o problema da presença de palavras consideradas bárbaras na nomenclatura.

Na Universidade de Évora, uma equipe de lexicógrafos propõe a compilação de uma obra em duas partes, assim descrita na folha de rosto do manuscrito (reproduzida por Silvestre e Borges (2012, p.324)): a primeira contendo “somente as palavras rigorosamente latinas e usadas dos Autores da 1ª e 2ª classe para uso e segurança dos que pretendem falar e escrever com pureza a língua latina”; a segunda trazendo “todas as palavras latinas bárbaras de que usaram quaesquer [sic] Autores que escreveram na língua latina desde o século argênteo da mesma língua até o nono”. Os modelos utilizados em tal empresa foram o dicionário latim-francês de Pierre Danet (1711) e, posteriormente, a versão revisada e ampliada do Calepino realizada por Facciolatti (1718) – esta conhecida pelo rigor filológico e particularmente abundantes em citações de autores clássicos, necessárias para que se conheça os “modos de falar dos autores clássicos” (ibid.).[89]

[88] Figueiredo havia publicado, em 1753, o *Novo Método da Gramática Latina para uso das Escolas da Congregaçaõ do Oratorio na Real Casa das Necessidades*, em cujo prólogo de 107 páginas lista cem erros na gramática de Manuel Álvares, pondo em causa a obra central do ensino jesuíta (Freire Cameron, 2012, p.149).

[89] Vale notar que essa ampliação do conteúdo fraseológico é bastante semelhante à que se observou no *Index* de

Em Lisboa, por sua vez, o padre José Caeiro é encarregado de fazer uma versão escolar da *Prosodia*, diversa da orientação tesaurística da equipe de Évora. Para tanto, adota como modelo o dicionário latino-italiano de Pasini (1731), cuja nomenclatura também era livre de "barbarismos" mas tinha a vantagem de selecionar, do extenso corpus de citações encontrado em Facciolatti, aquelas consideradas mais pertinentes para a aprendizagem.

A despeito de estarem em fase adiantada de produção (a de Caeiro tinha sido impressa até a letra S), ambas as reformulações foram interrompidas pela expulsão dos jesuítas de Portugal, em 1759, em seguida à proibição de seu ensino e de todas as suas obras escolares, no contexto das reformas pombalinas. Com isso, Portugal se vê totalmente desprovido de um dicionário de latim,[90] tornando urgente a tarefa de produzir um novo dicionário para as escolas. Ironicamente, os dois dicionários produzidos com essa finalidade se basearam, direta ou indiretamente, no trabalho de atualização realizado pelos jesuítas, como se vê a seguir.

3.2.6 O *Parvum lexicon* de Fonseca

O primeiro encarregado de produzir um dicionário para suprir as necessidades da escola em Portugal foi Pedro José da Fonseca, professor de retórica e poética no Real Colégio dos Nobres de Lisboa, em um período em que se assiste, na Europa, a uma "maior preocupação com a educação da nobreza", decorrente das novas ideias iluministas voltadas para a preparação da classe dirigente (Borges, 2011, p.45).[91] Pela primeira vez, a produção de um dicionário era regida por princípios que eram determinações de Estado.[92] A urgência do trabalho conduziu o compilador a escrever seu dicionário com base em uma tradução do dicionário de Pasini, mencionado acima, mas sabe-se que entre as obras consultadas constavam Danet, Facciolati, Estienne, bem como a parte já impressa da *Prosodia* de Caeiro (Silvestre; Borges, 2012, p.332). O resultado sai em 1762 com o título *Parvum Lexicon Latinum Lusitana Interpretatione Adiecta ad*

Velez a partir da edição de 1689, também realizada pela Universidade de Évora, sugerindo uma tendência da lexicografia na passagem do século XVII para o XVIII que merece investigação.

[90] Há dois testemunhos interessantes, mencionados por Freire Cameron (2012, p.44), de como portugueses e brasileiros lidaram com essa proibição. De um lado, as dezenas de exemplares da *Prosodia* encontrados com a folha de rosto arrancada revelam uma espécie de tática subversiva para ocultar a natureza da obra e salvá-la da destruição. De outro, a decisão do desembargador de Cachoeira, na Bahia, de não aceitar a devolução dos exemplares que lhe faziam religiosos, magistrados e médicos, em obediência ao novo decreto, postergando sua recolha "para o tempo em que houverem Diccionarios", parece sugerir que o dicionário de latim era item imprescindível para certos grupos sociais.

[91] Processo análogo ocorrera, por exemplo, na Itália, onde, entre 1729 e 1731, "a reorganização dos estudos em Piemonte teve como trabalhos prioritários o provimento de professores, a redação de um novo regulamento dos estudos e a composição de um bom dicionário latino, […] [que] deveria reunir o léxico dos melhores escritores da língua latina e declarar, com uma suficiente cópia de exemplos, as várias significações de cada entrada e a regência dos verbos e das preposições latinas" (Borges, 2011, p.116). O dicionário resultante são os *Vocabula Latini Italique Sermonis*, de Giuseppe Pasini (1731), citado acima como fonte da *Prosodia* de Caeiro.

[92] As *Instrucçoens para os Professores de Grammatica Latina, Grega, Hebraica, e de Rhetorica*, documento oficial publicado em 1759 que determinava o novo programa de estudos oficiais e a metodologia didática de cada disciplina a constar dos manuais escolares, dizem acerca dos dicionários: "Para uso dos Estudantes se tem escolhido hum Diccionario proporcionado aos seus principios; no qual, sem amonthoar authoridades, breve, e summariamente se lhes declarem as significaçoens naturaes, e figuradas, que são mais frequentes nos Authores, que lerem…", cf. Borges (2008, p.1081–2).

Vsum Lusitanorum Adolescentium, quer dizer, 'Pequeno léxico latino acompanhado da tradução portuguesa para uso dos jovens portugueses'.

O volume se inicia com um prólogo "Ao leitor" que descreve as escolhas feitas por Fonseca em ambos os níveis da macro e da microestrutura, explicadas no prefácio da obra (Fonseca, 1798, p.ii-vi). Condicionado pelo tempo escasso (escrita no "espaço de menos de hum anno") e orientado pelo objetivo de fazer um dicionário para uso de principiantes e com "preço moderado", o autor afirma ter "descarregado" o volume de tudo o que considerou supérfluo às primeiras classes, reduzindo a nomenclatura às palavras que ocorrem nos textos que os estudantes devem ler no princípio de seus estudos. Ficam assim excluídas as palavras que só se encontram em glossários, em escritores eclesiásticos ou da baixa latinidade, bem como palavras de significação obscena, consideradas inadequadas para "a mocidade", e todos os nomes próprios, deixados "à explicação dos Mestres"; termos técnicos ligados às artes e às ciências (e.g. pedras preciosas, animais, plantas, instrumentos, entre outros) são mantidos, mas sem "gastar tempo em largas descrições" de natureza enciclopédica, sempre em nome da brevidade; apenas os termos considerados como tendo "profundo significado cultural" (a saber, "costumes, dignidades e ofícios dos Romanos, do que pertence à sua religião e milicia") mereceram uma explicação mais detalhada. Por outro lado, afirma incluir na nomenclatura as formas verbais de pretérito e particípio, bem como certos casos oblíquos dos nomes que, "por se afastarem algum tanto da sua raiz, causam embaraço aos principiantes, que de ordinário os buscam segundo a ordem alfabética". A perspectiva pedagógica também impele Fonseca a "ajuntar a cada uma das diversas significações [das palavras] […] os autores de primeira e maior graduação", para acostumar os estudantes desde o início a conhecer "o merecimento das palavras, fazendo-lhes ver o seu Autor"; nesse sentido, reconhece ter recorrido a diversas fontes para obter tais informações, incluindo grandes dicionários como os de Estienne, Gesner, Facciolati. Igualmente digna de nota é a percepção que Fonseca revela a respeito da fraseologia, particularmente aquelas com caráter idiomático, pelo que implicam nas tarefas relacionadas à tradução:

> Ajuntei a muitas dicções aqueles usos e expressões particulares da língua Latina, que vulgarmente chamamos *frases*, para ir dando […] a conhecer o gênio das duas línguas, Latina e Portuguesa, e costumando os principiantes a percebê-lo no conhecimento de que é impraticável o expor no nosso idioma palavra por palavra inumeráveis termos e expressões latinas. (Fonseca, 1798, p.iv (verso), com ortografia atualizada)

Por fim, demonstra consciência de que a indicação do significado por meio de sinônimos é limitada, visto "não terem todos igual força nem se poderem tomar do mesmo modo ou servir-nos deles indiferentemente"; não obstante, afirma a utilidade de se recorrer a esse expediente, considerando que muitas vezes a visão conjunta de dois ou mais sinônimos favorece a compreensão do significado.

Segue-se ao prefácio um índice cronológico de autores latinos, escrito em latim, de autoria de Andrea Stubelio e presente em algumas edições do *Thesaurus*, que abrange um período de

quinze séculos, dividindo os autores em quatro "idades": de ouro (de 217 a.C. a 14 d.C.); de prata (de 14 a 117 d.C.); de bronze (de 117 a 410 d.C.); e de ferro (de 410 a 1182 d.C.) (Almeida, 1969; Borges, 2011).

A macroestrutura do dicionário se apresenta contendo quase 33 mil entradas – sendo 14 mil substantivos, 9,5 adjetivos, 6,5 mil verbos, 2,5 mil advérbios, aproximadamente –,[93] distribuídas por 826 páginas em formato *in-octavo* (aprox. 14x21cm). Uma página de amostra é reproduzida na Figura 12.

A microestrutura traz algumas novidades com relação aos dicionários latim-português anteriores, importadas do modelo de Pasini, mediante adaptações. No que respeita aos lemas, as sílabas têm a quantidade indicada por meio de sinais de longa e breve posicionados sobre as vogais, em vez da descrição por abreviaturas usada por Velez e Pereira;[94] porém, quase todas as sílabas recebem essa marcação, como na *Prosodia*, ao passo que em Pasini ela incide sobre sílabas selecionadas. Algumas classes contam com formas adicionais em sua estrutura lemática, especialmente os verbos, que passam a ter cinco formas, acrescentado o infinitivo àquelas tradicionalmente utilizadas, e os adjetivos, que passam a incluir o comparativo e o superlativo em seus lemas (em Pasini, essas formas são posicionadas após a explicação do significado).

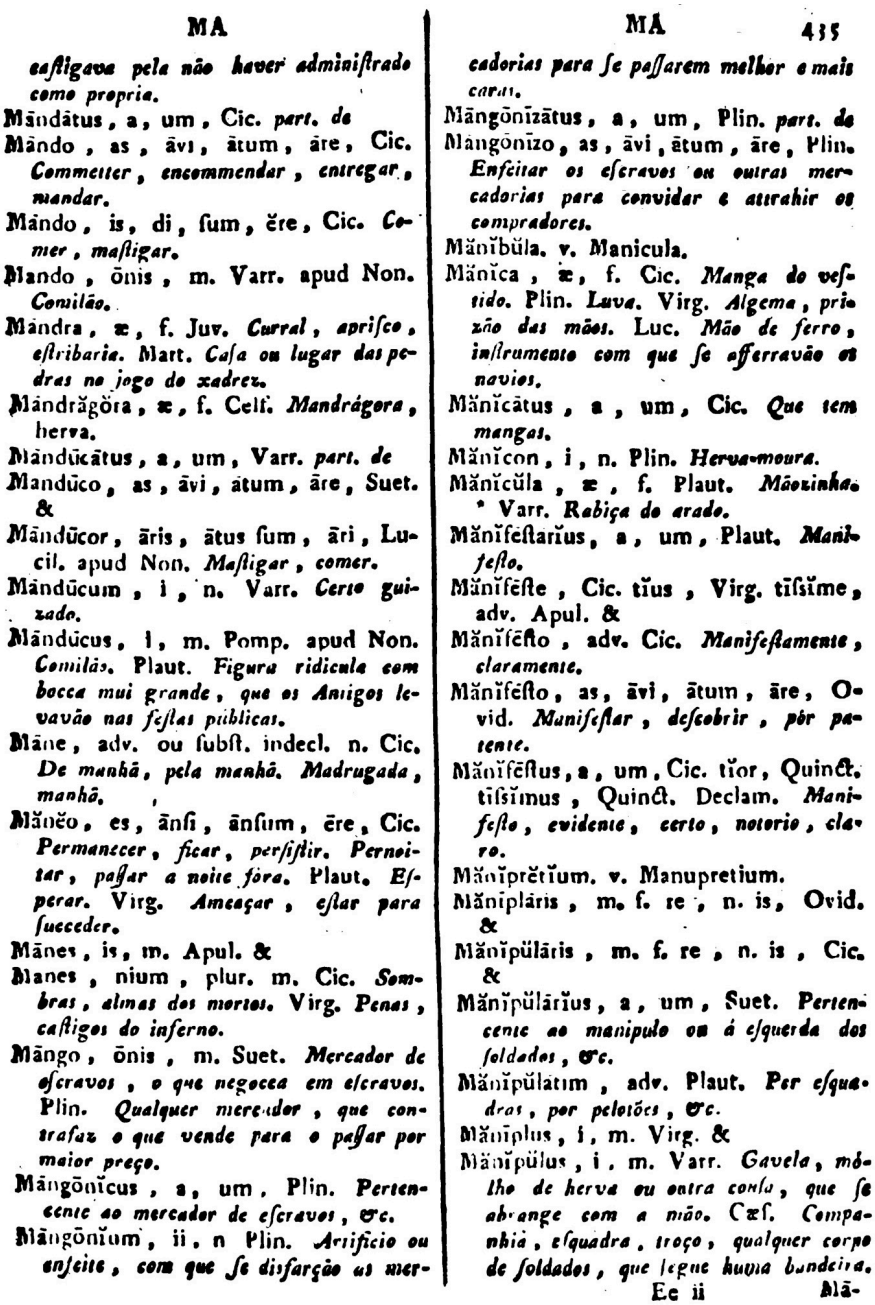
Quanto às informações gramaticais, referem-se principalmente à indicação dos gêneros nominais segundo a terminologia latina, de forma abreviada: 'f.' para feminino, 'm.' para masculino, 'n.' para neutro; no caso dos adjetivos triformes e biformes de segunda classe, cada forma de nominativo é especificada (e.g. 'Vŏlŭcer, m. cris, f. cre, n.' e 'Vŏlātĭlis, m. f. le, n. is', respectivamente), com os uniformes sendo acompanhados da indicação *omne genus*, isto é, "todos os gêneros" (e.g. 'Vŏcĭfĭcans, āntis, omn. gen.'). Além disso, há entradas de caráter remissivo, que lematizam formas flexionadas, como previsto pelo prólogo, incluindo na nomenclatura algumas formas de perfeito e de particípio, acompanhadas da devida indicação gramatical (e.g. 'Abduxi, *pret. de* Abduco' e 'Admissus, a, um, Cic. *part. de* Admitto', respectivamente). Vale notar que, quando a remissão se refere à entrada imediatamente posterior, o lema principal é omitido, implicando a entrada seguinte (e.g. 'Mandatus, *part. de* | Mando, as […]'). Já a indicação de classe de palavras só ocorre para as classes invariáveis ('adv.', 'prep. de ablat.', 'conj.', 'interj.'), seguindo a tradição estabelecida.

A descrição do significado, por sua vez, emprega uma técnica de definição geralmente baseada na acumulação sinonímica, expressa na forma de séries de equivalentes; a novidade aqui é o uso significativo e sistemático da pontuação para delimitar acepções (ponto-final) e equivalentes no interior de cada grupo (vírgula). Algumas vezes se inclui uma definição de tipo explicativo entre os equivalentes (e.g. 'Virtus, ūtis. f. Cic. *Virtude. Fortaleza, valentia. Propriedade, perfeição do homem tanto no corpo, como no animo.*'). Outras vezes, a aparente falta de

[93] Os dados quantitativos apresentados foram computados a partir da versão em texto digital disponível no *Corpus Lexicográfico do Português* (Fonseca, 2002), mediante um processo de estruturação dos dados descrito em detalhes no Capítulo 3.

[94] Ao contrário do que afirma Almeida (1969, p.207), essa marcação não era inédita em dicionários portugueses, tendo sido usada por Roboredo em *Raízes da língua Latina* (cf. *supra*).

Figura 12 – Página do *Parvum Lexicon Latinum* de Pedro José da Fonseca.



MA

caſtigava pela não haver adminiſtrado como propria.

Mandātus, a, um, Cic. *part. de*

Mando, as, āvi, ātum, āre, Cic. *Commetter, encommendar, entregar, mandar.*

Mando, is, di, ſum, ĕre, Cic. *Comer, maſtigar.*

Mando, ōnis, m. Varr. apud Non. *Comilão.*

Mandra, æ, f. Juv. *Curral, apriſco, eſtribaria.* Mart. *Caſa ou lugar das pedras no jogo do xadrez.*

Mandrăgŏra, æ, f. Celſ. *Mandrágora,* herva.

Mandūcātus, a, um, Varr. *part. de*

Mandūco, as, āvi, ātum, āre, Suet. &

Mandūcor, āris, ātus ſum, āri, Lucil. apud Non. *Maſtigar, comer.*

Mandūcum, i, n. Varr. *Certo guizado.*

Mandūcus, i, m. Pomp. apud Non. *Comilão.* Plaut. *Figura ridicula com bocca mui grande, que os Antigos levavão nas feſtas publicas.*

Mane, adv. ou ſubſt. indecl. n. Cic. *De manhã, pela manhã. Madrugada, manhã.*

Mănĕo, es, ānſi, ānſum, ēre, Cic. *Permanecer, ficar, perſiſtir. Pernoitar, paſſar a noite fóra.* Plaut. *Esperar.* Virg. *Ameaçar, eſtar para ſuceeder.*

Mānes, is, m. Apul. &

Manes, nium, plur. m. Cic. *Sombras, almas dos mortos.* Virg. *Penas, caſtigos do inferno.*

Māngo, ōnis, m. Suet. *Mercador de eſcravos, o que negoceа em eſcravos.* Plin. *Qualquer mercador, que contrafaz o que vende para o paſſar por maior preço.*

Māngōnĭcus, a, um, Plin. *Pertencente ao mercador de eſcravos, &c.*

Māngōnĭum, ii, n Plin. *Artificio ou enjeite, com que ſe disfarção as mercadorias para ſe paſſarem melhor e mais caras.*

MA 435

Māngōnīzātus, a, um, Plin. *part. de*

Māngōnīzo, as, āvi, ātum, āre, Plin. *Enfeitar os eſcravos ou outras mercadorias para convidar e attrahir os compradores.*

Mănĭbŭla. v. Manicula.

Mănĭca, æ, f. Cic. *Manga do veſtido.* Plin. *Luva.* Virg. *Algema, prizão das mãos.* Luc. *Mão de ferro, inſtrumento com que ſe aferravão os navios.*

Mănĭcātus, a, um, Cic. *Que tem mangas.*

Mănĭcon, i, n. Plin. *Herva-moura.*

Mănĭcŭla, æ, f. Plaut. *Mãozinha.* * Varr. *Rabiça do arado.*

Mănĭfeſtārĭus, a, um, Plaut. *Manifeſto.*

Mănĭfeſte, Cic. tĭus, Virg. tĭſſĭme, adv. Apul. &

Mănĭfēſto, adv. Cic. *Manifeſtamente, claramente.*

Mănĭfēſto, as, āvi, ātum, āre, Ovid. *Manifeſtar, deſcobrir, pôr patente.*

Mănĭfēſtus, a, um, Cic. tĭor, Quinct. tĭſſĭmus, Quinct. Declam. *Manifeſto, evidente, certo, notorio, claro.*

Mănĭprĕtĭum. v. Manupretium.

Mănĭplāris, m. f. re, n. is, Ovid. &

Mănĭpŭlāris, m. f. re, n. is, Cic. &

Mănĭpŭlārĭus, a, um, Suet. *Pertencente ao manipulo ou á eſquerda dos ſoldados, &c.*

Mănĭpŭlātim, adv. Plaut. *Por eſquadras, por pelotões, &c.*

Mănĭplus, i, m. Virg. &

Mănĭpŭlus, i, m. Varr. *Gavela, môlho de herva ou outra coiſa, que ſe abrange com a mão.* Cæſ. *Companhia, eſquadra, troço, qualquer corpo de ſoldados, que ſegue huma bandeira.*

Ee ii Mā-

Fonte: Fonseca (1798, p.435).

equivalente em português leva à utilização de definições hiperonímicas (Borges, 2011, p.140), usadas especialmente para lemas relativos às ciências naturais e expressas por meio de formulações como "espécie de […]", "gênero de […]" ou "certo/a […]" (e.g. 'Acorna, ae, f. Plin. *Genero de cardo*. | Acorum, i, n. Plin. *Certa herva medicinal*. | […] Acrocorium, ii, n. Plin. *Huma especie de cebolla*.').

Quanto ao uso de classificadores semânticos, este é bastante raro. Borges (2011, p.146) observa o emprego de dois tipos. Um deles é o uso metalinguístico da expressão "falando de …" para esclarecer a situação de comunicação (e.g. 'Calor, oris, m. Cic. *Calor*. Ter. *Impeto, fervor, vivacidade*, **fallando do animo**.', grifo adicionado); por vezes, o recurso é usado para suprir a eliminação dos itens fraseológicos, abundantes no dicionário de Pasini. O outro consiste em umas poucas marcas de uso, não passando de algumas dezenas, normalmente formuladas com uma expressão iniciada por "Termo […]", seja de caráter diatécnico (e.g. 'Participium, ii, n. Quinct. *Participio*, termo dos Grammaticos.'), seja de caráter diaestilístico (e.g. 'Belua, ae, f. Cic. *Féra. Homem cruel, selvagem*, termo injurioso.').

Outra característica nova, imitada de Pasini, é a abonação específica dos diversos elementos de informação veiculados na descrição lexicográfica, que recebem uma marca de autoridade que consiste na referência abreviada ao nome de um autor clássico, de preferência pertencente às idades de ouro e de prata.[95] Com isso, o recurso aparece diversas vezes num mesmo verbete, podendo ser usado para autorizar as diferentes acepções (e.g. 'Manica, ae, f. **Cic.** *Manga do vestido*. **Plin.** *Luva*. **Virg.** *Algema, prizão das mãos*. **Luc.** *Mão de ferro, instrumento com que se afferravão os navios*.', grifo adicionado), bem como casos de propriedades gramaticais alternativas (e.g. 'Vŏlucris, is, m. Cic. f. Virg.'), formas derivadas presentes no lema (e.g. 'Probatus, a, um, Cic. tior, Liv. tissimus, Cic.'), variantes lexicais (e.g. 'Manes, is, m. Apul. & | Manes, nium, plur. m. Cic. *Sombras dos mortos* […]')[96] ou itens fraseológicos, ainda que escassos: normalmente apresentados em bloco, ao final do verbetes, contemplam apenas 2 mil entradas, totalizando aproximadamente 3,6 mil frases.

Segundo Silvestre e Borges (2012), "a concisão do dicionário de Fonseca era sua principal limitação", especialmente no que se refere à omissão do vocabulário onomástico (que Pasini, aliás, fornecia em repertório à parte). Ainda assim, foi reeditado seis vezes em Portugal, ganhando, a partir da quarta edição (1798), um apêndice contendo a tradução do "Dicionário da fábula" de Chompré (que havia sido publicado separadamente em 1785 e contemplava os nomes próprios que faltavam ao dicionário geral)[97] e, na última edição (1847), um acréscimo de

[95] Segundo Borges (2008, p.1084), "os cinco mais citados – Cícero, Plínio, Plauto, Virgílio e Lívio – correspondem à tradicional hierarquia de autores que integram o cânone da 'latinidade pura'".

[96] Vale notar que, nesse último caso, as variantes são muitas vezes veiculadas em entradas distintas, com um conectivo latino abreviado (& = *et*, l. = *vel*) posicionado no final da entrada precedente para vincular o conteúdo semântico; estes representam um resquício da metalinguagem em latim no dicionário de Fonseca, junto com as indicações de gênero, vistas acima, e da preposição *apud*, usada ocasionalmente para indicar autoria.

[97] A descrição deste dicionário, precedida de um panorama da tradição renascentista de vocabulários especializados em nomes próprios ligados tanto a fatos históricos e culturais quanto aos enredos mitológicos encontrados nos textos latinos, encontra-se em Borges (2011, p.231–265) .

aproximadamente 5 mil vocábulos (Almeida, 1969, p.207).

3.2.7 O *Magnum lexicum* de Manuel de Pina Cabral

No final da década de 1770, o afastamento do Marquês de Pombal permitiu que se retomasse a reformulação da *Prosodia* preparada por Caeiro e parcialmente impressa quando da expulsão dos jesuítas. Segundo Silvestre e Borges (2012, p.333), o trabalho de Caeiro era superior ao executado por Fonseca "no número de entradas, na extensão das definições e das descrições em vernáculo, com um leque sinonímico amplo e traduções de excertos de autores latinos do período clássico". A incumbência de terminar a obra foi dada à Ordem Terceira de São Francisco, cabendo ao frei Manuel de Pina Cabral o encargo de escrever a parte faltante (das letras SAP– em diante) (Borges, 2011, p.148). Sua apreciação do dicionário incompleto dos jesuítas ficou registrada:

> He este Diccionario bastantemente copioso; mas sem superfluidade: nelle se poem por extenso todas as abbreviaturas romanas; explicam-se em Lingua Portugueza os significados proprios, e translatos de todas as palavras da pura Latinidade, as Frases, e os Proverbios; notaõ-se os vocabulos, em que ha variedade de liçaõ; poem-se os nomes proprios de Regioens, Ilhas, Provincias, Reinos, rios, lagos, montes, os dos homens, e Deoses, em huma palavra tudo o que pertence á Mythologia, á antiga Historia, e Geographia, que se encontra nos Poetas, Historiadores, e mais Authores Romanos.[98]

Aparentemente, Cabral finaliza o trabalho em poucos meses e o *Magnum lexicon Latinum et Lusitanum* vem à luz já em 1780.

Nesta primeira edição, a folha de rosto omite qualquer atribuição de autoria, trazendo apenas a caracterização do dicionário como "produzido a partir de sólidas observações dos filólogos mais célebres e eruditos e adequado à mais completa tradução dos escritores latinos";[99] um breve prólogo "Ao Leitor", escrito em português, menciona a ordem real dada à Ordem de São Francisco de se completar o dicionário inacabado (sem se referir aos jesuítas como seus primeiros compiladores) e o fato de ter sido a tarefa executada mediante "um vigilante cuidado de se consultarem não só os mais famosos Dicionários e os Críticos mais atendíveis, mas também as Edições mais corretas dos Autores Latinos" (reproduzido em Almeida (1969, p.214)). O trabalho de Cabral passa a ser reconhecido a partir da segunda edição, de 1802, que acrescenta à folha de rosto a seguinte indicação: "de longe mais abundante e corrigida que a anterior, por obra e esforço do Reverendo Padre Mestre Frei Manuel de Pina Cabral" (Cabral, 1802, tradução própria).[100]

---

[98] Carta de Pina Cabral a Fr. Manuel do Cenáculo, de 28 de Dezembro de 1779, reproduzida em Borges (2011, p.148).

[99] Tradução própria do original: "ex diuturnis celeberrimorum eruditissimorumque Philologorum observationibus depromptum, ad plenissimam scriptorum latinorum interpretatione accommodatum" (reproduzido por Almeida (1969, p.212).

[100] No original: "Editio Altera, priori longe auctior et emendatior opera et studio R. P. M. Fr. Emmanuelis Pinii Cabralii".

Nesta edição, a macroestrutura inicia logo após o prólogo, reproduzido da edição anterior, e se distribui por 727 páginas em formato grande (aprox. 18x28cm) numeradas e divididas em duas colunas, encabeçadas pelas três letras iniciais da primeira palavra; a mesma tríade é usada no interior das colunas para separar cada grupo alfabético em subconjuntos. Cada verbete ocupa um parágrafo, marcado pelo recuo de primeira linha e inicial maiúscula no lema. A Figura 13 fornece uma página de amostra.

A nomenclatura, contendo entre 40 e 45 mil entradas,[101] aparentemente segue o plano da *Prosodia* de Caeiro, já que mescla o vocabulário comum e uma diversidade de nomes próprios. Já os elementos da microestrutura seguem de perto a organização de Pasini, não apresentando nem as inovações de Fonseca (como os comparativos e superlativos anexados ao lema dos adjetivos, por exemplo) e nem sua redução drástica da fraseologia.[102] A explicação do significado é semelhante à de Fonseca, com abonações inseridas entre os grupos de equivalentes (característica, aliás, não encontrada em Pasini); uma peculiaridade de Cabral é o fato de ter preservado a fórmula de definição de adjetivos iniciada por "cousa […]", abreviada por 'c.' (e.g. 'Amplus, a, um. Cic. *c. ampla, grande, espaçosa, larga* […]').

A obra de Cabral teve grande aceitação no espaço escolar português. A quarta edição (1833) é ampliada, trazendo na folha de rosto o nome de José Antônio Ramalho, religioso da mesma congregação e professor de retórica, como editor e revisor. No mesmo ano, inicia-se uma série de edições impressas em Paris, com folha de rosto que reproduz as indicações anteriores mas com a autoria agora atribuída a Manuel José Ferreira. Em um prólogo ("Advertência"), este reconhece o trabalho dos padres Cabral e Ramalho como base da obra, mas afirma ter consultado "os melhores e mais modernos Diccionarios Latinos", citando as edições recentes de Gesner, Nolten, Wailly, Ainsworth, Noel e "mui particularmente da novissima edição Italiana do *Magnum totius Latinitatis Lexicon* de Forcelini"; além disso, afirma ter acrescentado "mais de cinco mil vocábulos latinos", marcados com asteriscos, e "muitas locuções dos clássicos, nas quaes a significação dos verbos ou adjectivos hé figurada", de modo a contemplar os eruditos e, principalmente, a "mocidade, a quem são destinados com mais particularidade os dicionários", dedicando a obra à "mocidade Portugueza e Brasiliense" (Cabral; Ramalho; Ferreira, 1867, s.p.). Essa edição parisiense foi reeditada por cinco vezes, sendo a última datada de 1867; já a edição portuguesa conheceu sua última edição em 1857, ampliada com vocábulos e frases extraídas da edição francesa e de outros dicionários.[103] Assim ampliado, o *Magnum Lexicon* de Cabral se torna o principal dicionário latino-português de uso escolar a partir da segunda metade do

[101] Estimativa baseada em duas fontes: (1) a afirmação de Borges (2011, p.149) de que "o *Magnum lexicon* contém mais cerca de 7.000 vocábulos" em relação aos 33 mil de Fonseca; (2) a contagem de linhas iniciadas por letra maiúscula no texto digital obtido por reconhecimento automático de caracteres da edição de 1867 (aprox. 50 mil), excluídas quase 6 mil entradas exclusivas desta edição, marcadas com asterisco (cf. <https://archive.org/details/dicionario-latim-portugues-magnum-lexicon-novissimum-latinum-et-lusitanum-1867>).

[102] A título de exemplo, a entrada para o verbo *maneo* contém 12 equivalentes e 17 frases em Pasini (1735, p.355); destes, Cabral (1802, p.376) aproveita todos os equivalentes e 12 frases, contra oito equivalentes e nenhuma frase em Fonseca (1798, p.435)

[103] A menção na folha de rosto, citada em Almeida (1969, p.215), inclui "Forcellino, Noel, Fonseca, Ferreira, Freund et aliis depromptis locupletata".

Figura 13 – Página do *Magnum lexicon Latinum et Lusitanum* de Cabral.

Oxycedros, i. f. Plin. *huma eſpecie de cedro, de páo mais duro, e cheo de nós.*

Oxygăla, ætis. n. Plin. *certa bebida feita de leite azedo, &c.*

Oxygăla, æ. f. Col. *o meſmo.*

Oxygărum, i. n. Mart. *huma caſta de molho envinagrado.*

Oxylăpăthum, i. n. Plin. *a labaça azeda, e de folha aguda.*

Oxymeli, n. *indecl.* Plin. *oximel, xarope prepado de mel cozido com vinagre, &c.*

Oxymēlis, ītis. f. Plin. *o meſmo.*

Oxymirsĭne, es. f. Plin. *a gilbarbeira, arbuſto.*

Oxypōrus, a, um. Plin. *c. envinagrada, azeda, temperada com vinagre.*

Oxys, yos. f. Plin. *trevo azedo, herva.* Plin. *huma eſpecie de junco marinho.*

Oxiſchœnŭs, i. m. Plin. *o junco marinho.*

Oxytrĭphyllon, i. n. Plin. *o trevo agudo, ou azédo, herva.*

O Z

Ozæna, æ. f. Plin. *huma eſpecie de polvo, peixe.* Plin. *a ozna, chaga, que ſe gera dentro do nariz, o polypo.*

Ozænītis, ĭdis. f. Plin. *eſpecie de nardo Indico, herva.*

Ozymum, i. *v.* Ocimum.

P

P Littera. P. *decima quarta letra do Alfabeto Latino, e huma das mutas. Nas Notas Romanas* P. pace. pactum. paſſus. pater. patria. pecunia. pedes perpetuus. pius. plebs. pontifex. populus. poſſeſſor. poſuit. poteſtas. præſes. prætor. pridie. princeps. provincia. publicus. publius. puer. pupillus. P. A. pius auguſtus. poſuit aram. PA. DI. patriciatus dignitas. PAL. palatina. palilia. P. ÆL. AVG. LIB. Publius Ælius Auguſti libertus. PAP. papia. PAR. parenti. parilia. Parthicus. PAR. INF. P. parentes infeliciſſimi poſuerunt.

P A B

Păbŭlans, antis. *adj. part.* Val. Max. *c. que anda á forragem.*

Păbŭlāris, re. Col. *c. de paſto, pertencente a paſto.*

Păbŭlātio, ōnis. f. Varr. *o paſto, a paſtagem.* Cæſ. *a ida a buſcar paſto, o ir á forragem.*

Păbŭlātor, ōris. m. Liv. *o que vai buſcar paſto, ou a forragear.*

Păbŭlātōrius, a, um. Col. *c. que ſerve para conſervar paſto, pertencente a paſto.*

Păbŭlor, āris, ātus, āri. Col. *apaſcentar-ſe, paſtar.* Cæſ. *forragear, ir, ou andar á forragem.* Pabulari oleas ſimo: Col. *eſtercar as oliveiras.*

Păbŭlōſus, a, um. Solin. *c. abundante de paſtos.*

Păbŭlum, i. n. Cic. *o paſto, a paſtagem, o ſuſtento dos brutos.* Pabulum ſtudii: Cic. *o goſto de eſtudar.* Morbi: Ovid. *o que ſerve de augmentar a enfermidade.*

P A C

Păcālis, le. Ovid. *c. de paz, pacifica.*

Păcandus, a, um. *adj. part.* Cæſ. *c. que ſe ha de pacificar.*

Păcāte. *adv.* Cic. de Conſolat. *pacificamente, ſocegadamente.* Pacatiùs. *comp.* Petr.

Păcātor, ōris. m. Sen. Trag. *o pacificador.*

Păcatus, a, um. *adj. part.* Cic. *c. pacificada, quieta, ſocegada, pacifica.* Pacatior. *comp.* Cic. Pacatiſſimus. *ſup.* Cic.

Pacenſis, le. Pacenſis colonia: Plin. *Béja, cid. de Portugal, que ſe chamou antigamente* Pax Julia.

Păchynum, i. n. Plin. *Cabo Paſſaro, hum dos tres promontorios de Sicilia.*

Păchynus, i. m. Ovid. *o meſmo.*

Păcĭfer, ĕra, ĕrum. Virg. *c. que traz paz, ou a ſignifica.*

Păcĭficans, antis. *adj. part.* Sall. *c. que pacifica.*

Păcĭficatio, ōnis. f. C.c. *a pacificação, o fazer pazes.*

Păcĭfĭcātor, ōris. m. Liv. *o pacificador.* Pacificator habitus in ſtatuis: Quint. *a acção da mão direita levantada nas eſtatuas principalmente dos Emperadores.*

Păcĭfĭcātōrius, a, um. Cic. *c. de paz, pertencente a paz.*

Păcĭfĭcātus, a, um. *adj. part.* Plaut. *c. pacificada, poſta em paz.*

Păcĭfĭco, as, āvi, ātum, āre. Liv. *pacificar, fazer pazes, pôr em paz.* Sil. *aplacar, moderar, mitigar.*

Păcĭfĭcus, a, um. Cic. *c. pacifica, que faz pazes, amiga de paz, quieta, ſocegada.*

Păcio, ōnis. f. Feſt. *v.* Pactio. *Antiq.*

Păciſco, is, ĕre. Næv. ap. Non. *fazer pacto. Antiq.*

Păciſcor, ĕris, pactus, iſci. Cic. *pactear, fazer pacto, concerto, ajuſte, ou convenção.* Pactus eſt pecuniam: Mart. *prometteo dar dinheiro.* Paciſci vitam ab hoſte: Sall. *render-ſe ſalvando a vida.* Vitam pro laude: Virg. *dar a vida por conſeguir gloria.* Filiam alicui: Cic. *prometter em matrimonio ſua filha a alguem.*

Păco, as, āvi, ātum, āre. Cæſ. *pacificar, aplacar, mitigar, moderar.* Sylvæ pacantur vomere: Hor. *com a lavaura ſe fazem os matos fecundos.*

Pacta, æ. f. Paterc. *donzella ajuſtada para caſar.*

Pactĭlis, le. Plin. *c. feita de partes diverſas, ou de pedaços.* Pactilis corona: Plin. *grinalda compoſta de diverſas caſtas de flores, ou de ramos.*

Pactio, ōnis. f. Cic. *o pacto, concerto, ajuſte, aſſento de paz, compromiſſo, a convenção.* Cic. *as peitas, o dinheiro por que o Juiz ſe deixa corromper.* Nuptialis pactio: Liv. *o ajuſte de caſamento.* Pactio verborum: Cic. *a fumula.*

Pactĭtius, a, um. Gell. *c. pactcada, concertada, ajuſtada.*

Pactōlis, ĭdis. f. Ovid. *a filha do rio Pactólo.*

Pactōlus, i. m. Hor. *Pactólo, hoje Sarabat, rio de Lydia.*

Pactor, ōris. m. Cic. *o que ajuſta pactos, &c.*

Pactum, i. n. Cic. *o pacto, ajuſte, concerto, aſſento de paz, a convenção, &c.* Hoc pacto: Plaut. Eo pacto: Ter. *deſta maneira.* Quonam pacto? Cic. *de que ſorte.*

Pactus, ûs. m. Petr. *o meſmo.*

Pactus, a, um. *adj. part.* Cic. *c. que pacteou, ajuſtou, concertou.* Cic. *c. pacteada, ajuſtada, concertada, promettida.* Pacta conjux: Virg. *promettida para caſar.*

Pactus, a, um. *adj. part.* de Pango. Ovid. *c. pregada, fincada.*

Păcŭviānus, a, um. Cic. *c. de Pacuvio.*

Păcŭvius, ii. m. Hor. *Pacuvio, Poeta Tragico antigo.*

P A D

Păde, is. f. Plin. *eſpecie de arvore reſinenta. L. var.*

Pădus, i. f. Plin. *o pinheiro alvar, arvore.*

Pă

Fonte: (Cabral, 1802, p.445).

século XIX, relegando os dicionários anteriores para a classe de preciosidades (senão raridades) bibliográficas (Borges, 2011, p.150).[104]

3.2.8 O *Novissimo diccionário* de F. R. dos Santos Saraiva

A referência acima à "mocidade Brasiliense" como distinta da portuguesa é certamente reflexo do *status* do Brasil como nação independente. Não por acaso, é ao contexto educacional brasileiro, particularmente o do Colégio D. Pedro de Alcântara (atualmente Colégio Pedro II), no Rio de Janeiro, que se vincula o primeiro dicionário latim-português escrito no Brasil: o *Novíssimo Dicionário Latino-Português* de Francisco Rodrigues dos Santos Saraiva.

Saraiva era um padre português, filho de um rabino convertido ao cristianismo, poliglota e conhecedor das línguas ocidentais e orientais (Rabuske, 1976). Formado pela Universidade de Coimbra, com passagem por Londres, onde teria adquirido conhecimentos de diversos idiomas contemporâneos e clássicos do Ocidente e do Oriente, vem ao Brasil por volta de 1860, numa espécie de auto-exílio.[105] Instalado no Rio de Janeiro a partir de 1870, é chamado por Dom Pedro II para uma entrevista sobre conhecimentos eruditos, em que teria impressionado o imperador com seus conhecimentos e sua erudição. Incentivador dos estudos filológicos no Brasil,[106] foi nomeado reitor do Colégio D. Pedro de Alcântara, função que desempenha entre 1870 e 1875. Nesse mesmo período, trabalha em uma versão ampliada da gramática latina do padre Figueiredo e, em especial, na elaboração do dicionário latim-português, sob contrato com a editora francesa dos Irmãos Garnier.[107]

---

[104] Almeida (1969, p.216) menciona ainda o *Diccionario Portuguez-Latino* de Fonseca, publicado de 1771 a 1879, como segundo elemento do par de dicionários escolares em Portugal nesse período. Além disso, pode ter entrado em concorrência com o *Magnum Lexicon* a nova edição do dicionário latim-português de Fonseca, publicada em 1847 com novo título (*Lexicon Latinum*) e conteúdo bastante ampliado em resposta à solicitação repentina de compêndios escolares gerada pela entrada em funcionamento dos liceus portugueses em 1844 (Borges, 2011, p.150).

[105] Sacramento Blake (1883, pp.106–7), um de seus raros biógrafos, assim caracteriza sua formação linguística e uma certa atitude que, talvez, tenha feito dele uma *persona non grata* perante a intelectualidade portuguesa e motivado sua saída de Portugal: "um dos homens mais extraordinários, sem exceptuar os sábios mais celebres da Europa, como disse o imperador D. Pedro de Alcântara; memoria admirável, intelligencia rara, sendo instruído por seu pae nas línguas semíticas e todo dedicado á linguistica desde os estudos ecclesiasticos, […] passou à Londres, onde conviveu com os mais notáveis orientalistas, adquiriu conhecimentos excepcionaes do hebraico, do sanscrito, do árabe, syriaco, latim, grego e idiomas do norte da Europa que falla e escreve correctamente, e algumas noções do chinez, que lhe permittem decifrar os trezentos caracteres radicaes da lingua litteraria e clássica; versado em summa nas línguas que teem códices escriptos, o é também na phenicia, lingua morta, de que mais de uma vez serviu-se em Portugal para confundir os semi-sabios, rindo-se dos equívocos em que os apanhava."

[106] Saraiva escreveu um memorial ao imperador brasileiro intitulado "Acerca da necessidade e utilidade do estudo das línguas bíblicas no Império do Brasil, como poderoso auxílio das Ciências e da Filologia" (Rabuske, 1976).

[107] Fato curioso, Rabuske (1976, p.41–42) menciona que a confecção do dicionário latino-português, "maior empreendimento de sua vida", foi "concluído ao fim de cinco longos e laboriosos anos" e teria causado tal esgotamento que Saraiva entra em crise, abandona o sacerdócio e a sociedade, vende sua riquíssima biblioteca e parte para as Picadas do Norte (hoje São José), em Santa Catarina, declarando, em sua partida, que ia "viver no mato" como um desconhecido; nessas condições, recebe em 1879 as provas para revisão do dicionário e, pelo que se conta, teria sido induzido pela Casa Garnier a se desfazer dos direitos de propriedade sobre a obra, em troca de uma compensação irrisória.

A primeira edição sai em 1881, impressa na França, em um volume de quase 1300 páginas em formato in-folio (21x28cm) divididas em três colunas (ver reprodução na Figura 14). A folha de rosto traz as seguintes informações:

> Novissimo Diccionario Latino-Portuguez etymologico, prosodico, historico, geographico, mythologico, biographico, etc.
>
> No qual são aproveitados os trabalhos de philologia e lexicographia mais recentes, redigido segundo o plano de L. QUICHERAT e precedido d'uma Lista dos Auctores e Monumentos latinos citados no volume e das principaes Siglas usadas na Lingua latina.

A referência aos avanços recentes da Filologia e da Lexicografia não é gratuita. A virada do século XVIII para o XIX tinha sido palco de mudanças significativas em ambas as disciplinas: de um lado, novas edições críticas de textos latinos exigiam uma atualização constante da seleção e descrição do léxico pelos dicionários; de outro, o uso de princípios históricos afirmava-se como lógica de organização dos verbetes, especialmente no que se refere à ordenação das acepções (Considine, 2014; Considine, 2016; Considine; Ogilvie; Safran, 2020).

No âmbito da lexicografia latina, dois nomes são especialmente relevantes: o italiano Egidio Forcellini e o alemão Wilhelm Freund. O primeiro é responsável por estabelecer, com a publicação do *Totius Latinitatis Lexicon* em 1771, uma nova base para a produção de dicionários de latim. Como ocorrera com Estienne duzentos anos antes, seu trabalho começa pela incumbência de fazer uma ampla revisão do dicionário de Calepino e, percebendo a insuficiência deste, termina com a produção de um dicionário monolíngue independente com pretensão de exaustividade tanto na nomenclatura quanto nas possibilidades de significado de cada palavra. Além disso, cada significado descrito era chancelado por passagens de autores devidamente referenciadas, incluindo a edição utilizada, o que conferia maior precisão e permitia que o leitor fizesse seu próprio julgamento. Seu único defeito, segundo Beck (1851, p.87), residia na "falta de um princípio claro e racional de organização" desses significados.

A resposta para esse problema não era de todo desconhecida. Viu-se acima que Roboredo, no início do século XVII, pretende apresentar a "significação própria" antes daquelas "que se estendem por semelhança"; no século seguinte, Ainsworth pretende, em seu dicionário latim-inglês (1736), seguir uma "ordem natural, que comece com a significação mais próxima da etimologia da palavra (quando ela é conhecida) e prossiga gradualmente aos sentidos mais distantes do original, mesmo que, por vezes, estes sejam os mais usuais" (Beck, 1851, p.83, tradução própria).[108] Um século mais tarde, a aplicação do princípio histórico na lexicografia latina se consolida pelas mãos de Freund, com a publicação, em 1844, do *Wörterbuch der lateinischen Sprache nach historisch-genetischen Principien* (isto é, "Dicionário da língua latina de acordo com os princípios histórico-genéticos"). Segundo as informações do prefácio, resumidas

[108] No original: "…a natural order, beginning with that signification which cometh nearest to the etymology of the word, where that is known, and proceeding gradually to those which are more remote from the original sense, though sometimes the most usual".

Figura 14 – Página do *Novissimo diccionario* de Saraiva.



14 ACC ACC ACE

mesa. *Vestimenta accubitoria.* PETR. Vestidos de estar á mesa.

Accŭbĭtŭm, ī, *app n.* (de *accubare*). LAMPR. Leito de mesa, especie de sofá para uma só pessoa.

Accŭbĭtŭs, ūs, *app. m.* (de *accubare*). STAT. Acção de deitar-se ou assentar-se á mesa. § STAT. Acção de pesar em ou sobre. § ISID. Leito, cama.

Accŭbŏ, ās, ŭī, ĭtŭm ārĕ, *v. intrans.* (de *ad* e *cubo*). Estar deitado perto, ou ao pé de alguem. *Hute bini custodes accubant.* PLAUT. Estão sempre ao lado d'este dois guardas. § Estar adjacente, encostado. *Theatrum monti accubans.* SUET. Theatro encostado a uma montanha. § Deitar-se para comer, pôr-se á mesa. — *aliquem.* PLAUT. — *alicui.* CIC. Estar á mesa, pôr-se á mesa junto a alguem. — *apud aliquem.* CIC. Comer em casa de alguem.

Accŭbŭī, *pret. perf.* de *Accubo* e *Accumbo*.

***Accŭbŭŏ,** *adv.* (de *accubare*). palav. comica. PLAUT. Em a postura d'uma pessoa deitada, encostada (á mesa).

Accūdŏ, ĭs, ūdī, ūsŭm, dĕrĕ, *v. trans.* (de *ad* e *cudo*). PLAUT. *Litterlt.* Forjar, bater em cima (fallando da moeda); bater, cunhar moeda. § PLAUT. Ajunctar, accrescentar a uma somma.

***Accŭmbĭtōrĭŭm, ī,** *app. n.* (de *accumbere*). INSCR. Salla para comer. Equivale a nossa salla de jantar.

Accŭmbŏ, ĭs, cŭbŭī, cŭbĭtŭm, ŭmbĕrĕ, *v. intrans.* (de *ad* e *cumbo*). *Litterlt.* Tomar logar em um leito (para comer), deitar-se, estender-se em um logar; encostar-se, accommodar-se (á mesa de comer); estar deitado. *Accumbere summis in undis.* MANIL. Estirar-se, estender-se na superficie das aguas. § *Accumbere in lecto.* PLAUT. — *lecto.* LIV. Deitar-se no leito, tomar logar á mesa. *Accumbere in summo.* PLAUT. — *summum.* HOR. Occupar á mesa o alto do leito, tomar o logar principal (o da esquerda). — *epulis divûm* VIRG. Tomar logar, fazer parte no banquete dos deuses — *cœnam.* APUL. Pôr-se á mesa para jantar.

Accŭmŭlātē, *adv.* (de *accumulatus*) AD HER. Accumuladamente, abundantemente. § **Accŭmŭlātissimē,** *superl.* AD HER.

Accŭmŭlātĭō, ōnĭs, *app. f.* (de *accumulare*). *term. jard.* PLIN. Accumulação, o amontoar a terra ao pé das arvores e arbustos.

Accŭmŭlātŏr, ōrĭs, *app. m.* (do mesmo. TAC. O que accumula, amontoa.

Accŭmŭlātŭs, ă, ŭm, *part. p.* do seg.

Accŭmŭlŏ, ās, āvī, ātŭm, ārĕ, *v. trans.* (de *ad* e *cumulo*). CIC. Accumular, amontoar, pôr em montão. — *vineas.* PLIN. Arreudar as vinhas, caval-as, chegando a terra para o pé da cépa. § *Accumulare cædem cædi.* LUCR. Amontoar morte sobre morte. § Augmentar. — *curas.* OV. — *metus.* ALBINOV. Augmentar os cuidados, os temores, receios. § VIRG. Encher alguem de. § SIL. Pôr o remate a ou em.

* **Accūrāssŏ,** *arch.* por *Accuravero*, de *Accuro*. PLAUT.

Accūrātē, *adv.* (de *accuratus*). Cuidadosamente, com cuidado, exactidão. — *fallere.* TER. Enganar destramente. — *habere.* SALL. Tractar com respeito, attenção. § **Accūrātŭs,** *comp.* CIC. — **Issimē,** *superl.* NEP.

Accūrātĭō, ōnĭs, *app. f.* (de *accuratus*). CIC. Cuidado, exactidão, diligencia. *Ad omnem accurationem.* VEG. Com todo o cuidado possivel.

Accūrātŭs, ă, ŭm, *part. p.* de *Accuro*. Feito com cuidado, diligencia, exactidão, bem trabalhado, aprimorado. *Accurata malitia.* PLAUT. Malicia completa, refinada. *Accuratæ rupæ.* APIC. Rabanetes bem preparados, guizados. *Accuratissimæ litteræ.* CIC. Carta circumstanciada, que entra em todos os pormenores. § *Accurato opus est.* PLAUT. E' preciso cuidado. § **Accūrātĭŏr,** *comp.* CIC. — **Issimŭs,** *superl.* CIC.

Accūrŏ, ās, āvī, ātŭm, ārĕ, *v. trans.* (de *ad* e *curo*). Cuidar, empregar seus cuidados em alg. coisa, occupar-se cuidadosamente d'uma coisa. *Accurare prandium alicujus.* PLAUT. Preparar o jantar de alguem. — *hospites.* PLAUT. Occupar-se com seus hospedes, tractar bem d'elles. — *victum humanum.* CIC. Prover á alimentação do homem.

Accūrrŏ, ĭs, cūrrī (cŭcūrrī, *arch.*), cūrsŭm, cūrrĕrĕ, *v. intrans.* (de *ad* e *curro*). Correr para alg. logar. correr para, correr em direcção a, vir correndo. — *Romam.* CIC. Correr a Roma. — *in Tusculanum.* CIC. Correr a Tusculo. — *ad prætorem.* CIC. Dirigir-se ao pretor. § *Fig. Istæ imagines accurrunt.* CIC. Estas imagens apresentam-se repentinamente. § *Passiv. Accurritur ab universis.* TAC. Todos acodem, i. é, correm apressadamente para um ponto.

Accūrsŭs, ūs, *app. m.* (de *accurrere*). TAC. Acção de acudir, de correr para um ponto, concurso para alguma parte.

Accūsābĭlĭs, ĕ, *adj.* (de *accusare*). CIC. Digno de ser accusado, censuravel, reprehensivel.

Accūsātĭō, ōnĭs, *app. f.* (de *accusatus*). CIC. Accusação. *Accusationem capessere.* TAC. Encarregar-se d'uma accusação. — *comparare, constituere, factitare, instruere.* CIC. Estabelecer, preparar uma accusação. — *exsequi.* TAC. Continuar uma accusação. *Accusationem dimittere.* VICT. *Accusatione desistere.* CIC. Renunciar a uma accusação. *Accusationes exercere.* TAC. Fazer officio de accusador. § *Accusationes Annibalis.* LIV. Accusação feitas a Annibal, ou de que elle foi alvo.

Accūsātīvŭs, ă, ŭm, *adj.* (de *accusare*). *term. gram. Accusativus,* subent. *casus.* QUINT. Accusativo, quarto caso dos nomes latinos. § *Accusativi præpositio.* ISID. Preposição do accusativo.

Accūsātŏr, ōrĭs, *app m.* (do mesmo). CIC. Accusador, o que accusa. *Sustinere nomen accusatoris.* TAC. Dar-se como accusador (em vez d'outro). § SUET. Delator, denunciador.

Accūsātōrĭē, *adv.* (de *accusatorius*). LIV. Como accusador, a modo de accusador, com paixão e encarniçamento de accusador.

Accūsātōrĭŭs, ă, ŭm, *adj.* (de *accusator*). Que pertence ou se refere ao accusador ou accusação. *Accusatorius animus.* CIC. Espirito de accusador, animosidade. *Accusatoriam vitam vivere.* QUINT. Viver de accusações.

Accūsātrīx, ĭcĭs, *app. f.* de *Accusator*. PLIN. Accusadora, a que reprehende. § PLAUT. A que provoca explicações.

Accūsātŭs, ă, ŭm, *part. p.* de *Accuso*. CIC. Accusado.

Accūsĭtŏ, ās, āvī, ātŭm, ārĕ, *v. trans. freq.* de *Accuso*. PLAUT. Accusar, reprehender frequentemente, tomar a culpa a.

Accūsŏ, ās, āvī, ātŭm, ārĕ, *v. trans.* (de *ad* e *causa*). 1° Accusar, censurar, reprehender; 2° Accusar em juizo, obrigar alg. a justificar-se, citar em juizo; pôr em litigio. § 1° *Accusemus gravius.* PLAUT. Façamos mais graves accusações. *Accusare inertiam.* CIC. Arguir, censurar, lançar em rosto a ociosidade. — *de negligentiâ epistolarum.* CIC. — de ser negligente em escrever. *Si id me non accusas.* TER. Se tu me não censuras isto. *In quo te accuso.* CIC. O ponto em que eu te accuso. *Accusare quòd...* CIC. Accusar de... — *crimine invidiæ.* NEP. Accusar de inveja. § 2° *Accusare ambitûs.* CIC. Accusar de ambitus. — *de vi.* CIC. — de violencia. — *inter sicarios.* CIC. Accusar de assassinato, attribuir a alg. este crime por estar na classe dos assassinos. § *Casus accusandi.* VARR. O accusativo, 4° caso dos nomes latinos.

Acdĕī, ōrŭm, *pr. m. plur.* PLIN. Acdeos, povo da Scythia.

Acdēstĭs, ĭs, ou **Agdēstĭs,** *pr. m.* ARN. Acdeste, monstro nascido d'um sonho de Jupiter.

Acdēstīnŭs, ă, ŭm, ou **Acdēstĭŭs, ă, ŭm,** *adj.* (de *Acdestis*). ARN. De Acdeste, que se refere a Acdeste.

Acē, ēs, *pr. f.* (Ἄκκη), (hebr. עכו, areia ardente pelos raios do sol) NEP. PLIN., depois Ptolemais ou Acca, cid. da Galilea, hoje S. João d'Acre.

Acēdĭă, æ, *app. f.* (ἀκηδία). *Eccl.* Indolencia, negligencia, desleixo, langor, abatimento, tristeza.

***Acēdĭŏr, ārĭs, ārī,** *v. dep.* (ἀκηδιάω). HIER. Ser negligente, moroso; entristecer-se, contristar-se; desanimar-se, cair em langor; soffrer, levar com impaciencia.

** **Acēldāmă** ou **Hăcēldāmă,** *app. indecl.* (forma syro-haldaica do hebr. חלקדמ, bhêlek-dam, sorte de sangue, i. é, campo de sangue). BIBL. PRUD. Pedaço de terreno, perto de Jerusalem, comprado pelos trinta dinheiros por que Judas vendeu ao Salvador.

Acelum, ī, *pr. n.* PLIN. Acelo, cid. da Venecia, hoje Asola.

Acenippo. Ved. *Acinippo.*

Acēntētă, ōrŭm, *app. n plur.* (ἀκίνητα). PLIN. Cristaes puros, sem pontos ou outros defeitos.

Acĕō, ēs, ăcŭī, ērĕ, *v. intrans.* (da raiz primitiva Ac, que exprime a ideia de azedume) CAT. Estar azedo, acido, azedar, envinagrar-se (fallando do vinho).

Acĕphălŭs, ī, *app. m.* (ἀκέφαλος, sem cabeça), ISID. Acéphalo, que não tem chefe. — *versus.* TER. MAUR. Verso acephalo, que parece não ter cabeça (quando o hexametro começa por uma syllaba breve.

Acĕphălī, ōrŭm, *pr. m. plur.* (do preced.). ISID. Acephalos, nom. d'uma seita heretica.

Acĕr, ĕrĭs, *app. n.* (de *ac*, raiz primit.). OV. Bordo, arvore.

Ācĕr, *m.* **ācrĭs,** *f.* **ācrĕ,** *n. adj.* (ἄκρος). Propriamente: Tudo que acaba em ponta aguda, ou extremidade fina; donde: Pontudo, agudo, cortante. 1° Em ger. Tudo que impressiona fortemente os sentidos, azedo ao paladar, aspero, duro ao tacto, estridente ao ouvido, vivo á vista, activo ao olfato; 2° Forte, grande, veliemente; 3° Activo, penetrante, vivo (fallando dos sentidos ou da intelligencia); 4° Activo, resoluto, animoso; energico, rigoroso; irascivel; fogoso, terrivel; 5° Aspero, duro, severo, cruel; 6° Attento, cuidadoso, desvelado, solicito. § 1° *Acer gustatu.* PLIN. Acido, azedo, agro ao paladar. — *stomachus.* HOR. Estomago que tem amargores. *Acria viscerum.* PLIN. Agrura dos humores. *Acer splendor.* LUCR. Vivo esplendor. *Colores acerrimi ruboris.* SEN. Côr do mais vivo ou brilhante vermelho. *Acer hinnitus.* VIRG. Relincho agudo. — *odor.* PLIN. Cheiro forte, penetrante. *Acris tibia.* HOR. Flauta cujo som é agudo, penetrante. — *hiems.* HOR. Inverno rigoroso. § 2° *Acris sitis.* TIB. Sêde ardente. *Acer dolor.* LUCR. Dôr aguda, pungente. — *metus.* VIRG. Grande medo, medo mortal. *Acri sole siccare.* PLIN. Fazer seccar ao sol forte, ardente. § 3° *Acres oculi.* CIC. Olhos de lynce, vista penetrante. *Acris memoria.* CIC. Memoria firme, segura, tenaz. *Acres imagines.* CIC. Vivas imagens. *Acerrimus æstimator.* CIC. Avaliador recto, excellente juiz. § 4° *Acres potores.* HOR. Bebedores denodados,

Fonte: Saraiva (1927, p.14).

por Beck (1851, pp.73–5), Freund concebe como objetivo da lexicografia latina "apresentar a natureza de cada palavra da língua latina em todos os períodos de sua existência", ou melhor, apresentar "a história de cada palavra latina durante o tempo em que o latim foi a língua nacional do povo romano, isto é, dos primeiros tempos ao fim do Império".[109] Para Freund, a "natureza" da palavra tem um aspecto exterior (a forma) e outro interior (o significado) e cada verbete de um dicionário é uma espécie de "monografia histórica de uma palavra particular" elaborada de modo a contemplar esses dois aspectos a partir de sete elementos descritivos e quatro princípios de organização. Os sete elementos descritivos se dividem em dois grupos; de um lado, a história exterior é descrita em termos de (1) informação gramatical, restrita ao que pertence à lexicografia, e (2) a informação etimológica, relativa aos radicais de que a palavra é formada; do outro, a história interior consiste em (3) informação exegética, que mostra o significado da palavra, (4) informação sinonímica, que compara e distingue significados próximos, (5) informação cronológica, que indica o período em que a palavra é usada, (6) informação retórica, relacionada com o estilo em que a palavra é usada (poético ou técnico, baixo ou elevado) e (7) informação estatística, que especifica se a palavra é usada frequentemente ou raramente. Os quatro princípios de organização, por sua vez, são os seguintes: (1) entre os diversos significados, considerar como sendo o original aquele que se depreende da etimologia da palavra; (2) posicionar o sentido próprio (entendido como original) antes dos sentidos figurados (entendidos como derivados), e subdividir esses últimos da mesma maneira; (3) organizar as passagens citadas de autores como abonação das acepções também cronologicamente, com prosadores precedendo poetas, textos autênticos precedendo textos não genuínos; e (4) basear todas as afirmações anteriores na leitura dos textos dos autores, não de índices.

A obra de Freund passa a ser recebida como um novo padrão de dicionário bilíngue, sendo traduzida para o inglês por Ethan A. Andrews em 1851 (Beck, 1851) e para o francês por Napoléon Theil em 1855 (Flobert, 2003). A relação de Saraiva com esse novo modo de conceber dicionários se dá, porém, de modo indireto, através de Louis-Marie Quicherat, como ele mesmo reconhece no frontispício de sua obra. Uma vez que não compôs um prefácio, a compreensão do plano de seu dicionário passa pelo conhecimento de seu modelo.

A primeira edição do *Dictionnaire Latin-Français* de Quicherat, na verdade escrito a quatro mãos com Amédée Daveluy, tido como o melhor professor de retórica de Paris, é publicada em 1844. Nesse tempo, ele já era conhecido por uma série de manuais, edições de textos clássicos e, especialmente, pelo *Thesaurus poeticus*, um repertório completo do vocabulário poético latino voltado para exercícios escolares de versificação, rico em informações sobre prosódia, sinônimos, antônimos e fraseologia coletadas, segundo ele, por meio de leitura direta "de todos os poetas latinos, de Lívio Andrônico a Fortunato", como menciona Flobert (2001, p.390). O prefácio da primeira edição informa os princípios que orientaram a composição (Quicherat;

[109] Segundo Kaltner e Santos (2024, p.7), a descrição da língua (ou, neste caso, do léxico) em sua historicidade, isto é, como um objeto de estudos "relativo a um determinado contexto histórico social, ou a um povo e território", é característica do método histórico-comparativo da Filologia do século XIX.

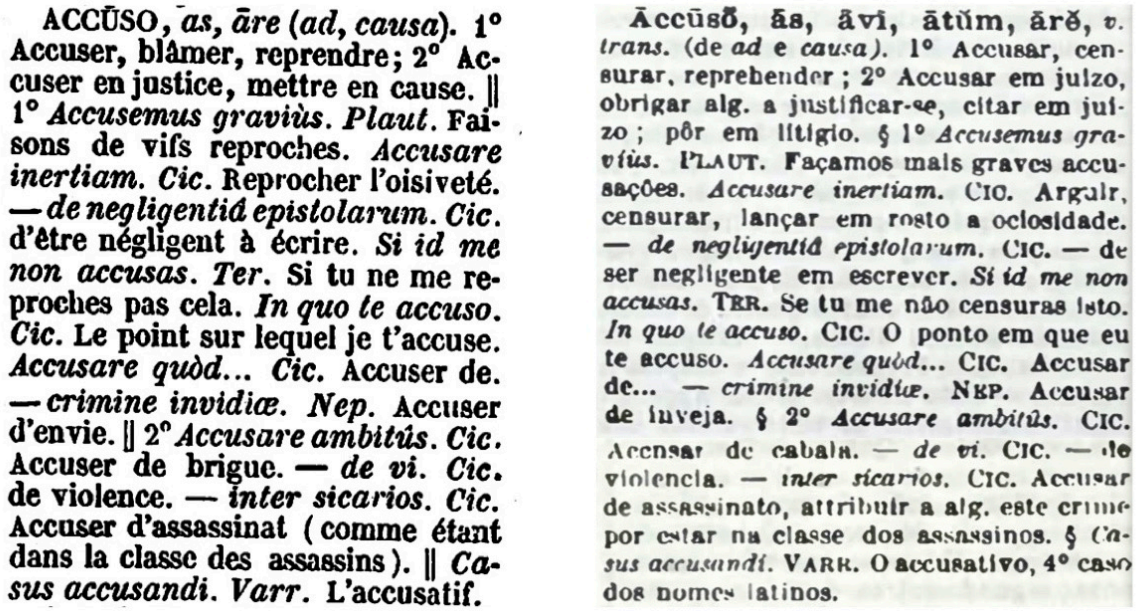
Daveluy, 1865). Com o objetivo de fazer um dicionário para uso das classes, os compiladores afirmam ter seguido de perto Forcellini e Freund, mas verificando as informações com base nos trabalhos recentes da crítica textual. O conteúdo pretende ser exaustivo tanto na nomenclatura quanto na descrição de significados. Assim, ressaltam ter acrescentado 1,5 mil palavras à nomenclatura de Forcellini (aparentemente um feito inédito até então), a maior parte encontrada em tratados gramaticais e autores eclesiásticos; por outro lado, suprimem todas as palavras que ocorrem exclusivamente em escritores posteriores ao século VI (exceção feita a Isidoro e a palavras "escolhidas" dos glossários) e todas as palavras gregas que aparecem transliteradas nos manuscritos, por "ignorância dos copistas"; além disso, inserem na nomenclatura certas formas flexionadas (perfeitos e genitivos) consideradas mais difíceis e tratam os particípios passados como entradas distintas dos verbos (já que os primeiros não herdam todos os significados do segundo e podem ter significados exclusivos). Quanto à descrição, fornecem criteriosamente uma série de indicações formais como a quantidade das sílabas quando não é dedutível pelas regras gerais e a etimologia apenas quando certificada pela crítica moderna, além de marcar formas arcaicas e de grafia duvidosa. O significado, por sua vez, é descrito sempre a partir do sentido primitivo aos sentidos derivados, mesmo quando o primeiro é pouco usado; nos casos de verbetes muito extensos, agrupam-se no início todos os significados acompanhados de numeração, para então serem fornecidos todos os exemplos, igualmente numerados, em um segundo bloco (uma técnica, segundo os autores, inspirada no dicionário grego-francês de M. Alexandre). A esse respeito, aliás, os autores reconhecem que não transcreveram todas as frases dos grandes léxicos, porém tudo o que transcreveram foi traduzido (algo que os grandes dicionários não fazem), eventualmente de diversas maneiras, seja nos casos em que o francês possui mais de uma locução, seja quando a tradução se distancia do texto, requerendo uma explicação literal; nesse trabalho, teriam consultado as melhores traduções, de modo que o dicionário é capaz de informar não só a tradução mais conveniente, mas também o melhor método. Por fim, toda palavra e todo significado seriam referenciados com pelo menos uma autoridade (um prosador e um poeta quando possível) com "extremo rigor", buscando não simplesmente uma atestação, mas um emprego autêntico, com cada autor sendo designado por uma abreviatura exclusiva.

Saraiva segue bem de perto esse modelo, mas insere algumas modificações. No que respeita à macroestrutura, a principal é que ele agrega na mesma nomenclatura o "Vocabulário de nomes geográficos, mitológicos e históricos" que vinha à parte na edição francesa, somando assim aproximadamente 75 mil entradas;[110] também acrescenta, antes do dicionário principal, um repertório de formas abreviadas, um tipo de item lexical totalmente ignorado pelo seu modelo, retomando assim aparentemente a tradição da *Prosodia* refletida também no *Magnum lexicon*. No que respeita à microestrutura, o conteúdo é praticamente o mesmo, mas pequenas diferenças mostram que a imitação de Saraiva não é totalmente servil como quis o revisor de Quicherat na edição de 1892.[111] A título de comparação, a Figura 15 reproduz lado a lado os respectivos

[110] Valor obtido por cálculo impreciso, baseado na contagem manual das entradas em quinze páginas aleatórias. Em tempo: o vocabulário de nomes próprios só foi integrado ao dicionário francês na edição revisada de 1892.

[111] Trata-se de seu sobrinho, Émile Chatelain, que no prefácio da referida edição comenta o amplo sucesso do

Figura 15 – Montagem comparativa dos verbetes relativos ao verbo *accuso* nos dicionários de Quicherat e Saraiva, respectivamente.



ACCŪSO, ās, āre (ad, causa). 1° Accuser, blâmer, reprendre; 2° Accuser en justice, mettre en cause. || 1° *Accusemus graviùs. Plaut.* Faisons de vifs reproches. *Accusare inertiam. Cic.* Reprocher l'oisiveté. — *de negligentiâ epistolarum. Cic.* d'être négligent à écrire. *Si id me non accusas. Ter.* Si tu ne me reproches pas cela. *In quo te accuso. Cic.* Le point sur lequel je t'accuse. *Accusare quòd... Cic.* Accuser de. — *crimine invidiæ. Nep.* Accuser d'envie. || 2° *Accusare ambitûs. Cic.* Accuser de brigue. — *de vi. Cic.* de violence. — *inter sicarios. Cic.* Accuser d'assassinat (comme étant dans la classe des assassins). || *Casus accusandi. Varr.* L'accusatif.

Accūsō, ās, āvī, ātŭm, ārĕ, *v. trans.* (de *ad* e *causa*). 1° Accusar, censurar, reprehender; 2° Accusar em juizo, obrigar alg. a justificar-se, citar em juizo; pôr em litigio. § 1° *Accusemus graviùs.* PLAUT. Façamos mais graves accusações. *Accusare inertiam.* CIC. Arguir, censurar, lançar em rosto a ociosidade. — *de negligentiâ epistolarum.* CIC. — de ser negligente em escrever. *Si id me non accusas.* TER. Se tu me não censuras isto. *In quo te accuso.* CIC. O ponto em que eu te accuso. *Accusare quòd...* CIC. Accusar de... — *crimine invidiæ.* NEP. Accusar de inveja. § 2° *Accusare ambitûs.* CIC. Accusar de cabala. — *de vi.* CIC. — de violencia. — *inter sicarios.* CIC. Accusar de assassinato, attribuir a alg. este crime por estar na classe dos assassinos. § *Casus accusandi.* VARR. O accusativo, 4° caso dos nomes latinos.

Fontes: Quicherat e Daveluy (1865, p.13); Saraiva (1927, p.14).

verbetes que descrevem o verbo *accuso*. Para começar, Saraiva mantém a fórmula completa do lema verbal, indicando cinco formas flexionadas para todos os verbos, ao passo que Quicherat reduz a três formas o lema de verbos com morfologia regular. Também indica a quantidade silábica por meio de diacríticos em todas as vogais, ao passo que Quicherat indica apenas na penúltima sílaba, e acrescenta sistematicamente uma informação gramatical a respeito da transitividade do verbo ('v. trans.') que é ausente no modelo.[112] A informação etimológica que vem em seguida, entre parênteses, é a mesma, porém com uma formulação textualizada no dicionário brasileiro, esquemática no francês. O registro de acepções é praticamente idêntico, com duas séries de equivalentes de tradução numeradas, acompanhadas de dois blocos de exemplos igualmente numerados e separados por '||' ou '§', respectivamente; o mesmo separador precede uma última expressão idiomática, não vinculada a nenhuma acepção ('casus accusandi'). Aqui, a diferença consiste em dois equivalentes a mais na segunda acepção e um comentário explicativo de natureza cultural para o significado da última frase.

O dicionário de Saraiva teve uma boa recepção da crítica, a julgar pela opinião de Sacramento Blake (1883, p.107):

dicionário compilado por seu tio e, após destacar a qualidade dos dicionários ingleses daquele século, observa: "D'autres pays, notamment le Portugal, se contentent de traduire mot pour mot le dictionnaire latin-français de Quicherat et Daveluy en mettant sur le titre qu'on a suivi le même plan" (Quicherat; Daveluy; Chatelain, 1892, p.vi).

112 Saraiva parece ter sido o primeiro lexicógrafo do latim em português a marcar sistematicamente a classe gramatical em todas as entradas, com substantivos divididos em próprios ('pr.') e apelativos ('app.') acompanhados da indicação de gênero ('m', 'f.' ou 'n.'), particípios divididos em presente ('part. pres.') e passado ('part. p.'), pronomes devidamente caracterizados ('pron. pess.', 'pron. demonstr.' etc.) e as demais trazendo apenas a sua classe ('adj.', 'adv.', 'conj.', 'prep.', 'interj.').

> é notavelmente superior ao *Magnum Lexicon*, tanto pelo avultado número de vocábulos, que neste não se encontram, como pela explicação das diversas accepções desses vocábulos, justificada com muitos exemplos tirados dos livros latinos e ainda pela clareza e rigor com que são definidos os vocábulos.

Foi reeditado pela Garnier até a nona edição, de 1927, aparentemente sem mudanças significativas; a partir de 1976, tal edição vem sendo reproduzida de modo fac-similar, em formato reduzido (com certo prejuízo à legibilidade), já alcançando a marca de 14 reimpressões.[113] Dos mesmos anos de 1970 data o parecer de Rabuske (1976), que ecoa a opinião de seu contemporâneo, Martins Terra, ao afirmar que o Saraiva é "um dicionário realmente bom e completo Latino-Português, para dirimir qualquer dificuldade de interpretação […], [o] melhor léxico no gênero que se tenha feito tanto em Portugal como no Brasil".

Não obstante, na primeira metade do século XX são publicados dois dicionários latim-português igualmente inspirados em Quicherat, porém com uma tendência não mais à exaustividade, e sim à condensação do conteúdo: o anônimo *Dicionário Latim-Português* da coleção "Diccionários do Povo" e o *Dicionário Latino Português* de Francisco Torrinha.

3.2.9 O *Dicionário Latim-Português* "do Povo"

Na passagem do século XIX para o século XX, consolida-se no ramo editorial um investimento em edições econômicas destinadas à expansão do comércio de livros para a camada urbana, assalariada e alfabetizada, que crescia diariamente, segundo uma estratégia baseada em baixo custo de produção e dinamização do consumo por meio do envio de livros para "qualquer canto do País" via correio (Macambira, 2010, pp.198–9). Segundo a autora, torna-se recorrente o uso das expressões 'do povo' e 'popular' nos títulos dos livros, termos que certamente não se referiam às camadas pobres ou de baixa renda, mas apenas expressavam o desejo de extrapolar as fronteiras econômicas e sociais (até então, a compra de livros era privilégio dos ricos). Nesse contexto, a coleção "Os Dicionários do povo", produzida em parceria pelas Livrarias Francisco Alves (Rio de Janeiro) e Bertrand (Lisboa) e autodescrita como provedora de "dicionários portáteis, econômicos, completos, indispensáveis em todas as famílias, escolas, bibliotecas, escritórios comerciais e repartições públicas", lança o *Diccionário Latim-Portuguez etymologico, prosodico e orthographico* em dois volumes tamanho in-folio (21x28cm) de aproximadamente 550 páginas cada, contemplando aproximadamente 60 mil entradas:[114] o número 6 contém as entradas de A-O; o número 7, de P-Z.

A página de rosto da edição consultada (Diccionario…, 1930?) não traz nenhuma informação sobre o dicionário além do título, contendo apenas os títulos dos demais números publicados na coleção e uma declaração de objetivos ("propaganda de instrução para portuguezes e brazileiros"), sem indicar autoria ou ano de publicação.[115] Em seguida, uma lista das abreviaturas

[113] O exemplar consultado, adquirido como livro novo, foi impresso em 2021.
[114] Valor impreciso, estimado com base na contagem manual de entradas em oito páginas aleatórias.
[115] A consulta às datas conhecidas de publicação dos demais números da coleção permitem deduzir limites tem-

Figura 16 – Página do *Diccionário Latim-Portuguez* da coleção "Do Povo".

Accrementum, i (*ákrêmen'tum*), *s. n.* Acremento, incremento, augmento.

Accresco, is, evi, etum, escere (*ákrés'kó*), *v. n.* Accrescer, augmentar-se. Sobrevir.

Accretio, onis (*ákré'ció*), *s. f.* Accrescimo, accrescimento, augmento.

Accretus, a um (*akré'tus*), *p. p.* de Accresco.

Accubans, antis (*á'kubans*), *part.* de Accubo. Accumbente, recostado.

Accubatio, Accubitio, onis (*ákubá'ció, ákubi'ció*), *s. f.* O estar encostado (á mesa) junto a outro. *Accubitio epularis amicorum.* Reunião de amigos, á mesa.

Accubitorius, a, um (*ákubitó'rius*), *adj.* Que serve para encosto (quando se está á mesa).

Accubitum, i (*áku'bitum*), *s. n.* **Accubitus, us** (*áku'bitus*), *s. m.* Accubito, especie de leito para recosto.

Accubo, as, ui, itum, are (*á'kubó*), *v. n.* Estar sentado junto d'outro (á mesa). Convizinhar, ser vizinho. *Accubare apud aliquem.* Comer em casa de alguem.

Accubuò (*áku'buó*), *adv.* Na posição de quem está deitado.

Accudo, is, udi, usum, ere (*áku'dó*), *v. a.* Bater moeda. Accrescentar.

Accumbo, is, ubui, ubitum, ere (*ákum'bó*), *v. n.* Deitar-se, extender-se. Tomar logar á mesa.

Accumulatè (*ákumulá'té*), *adv.* Accumuladamente, em cumulo.

Accumulatio, onis (*ákumulá'ció*), *s. f.* Accumulação, amontoamento.

Accumulator, oris (*ákumulá'tór*), *s. m.* Accumulador, amontoador.

Accumulo, as, are (*áku'muló*), *v. a.* Accumular, amontoar.

Accuratè (*ákurá'té*), *adv.* Accuradamente, esmeradamente.

Accuratio, onis (*ákurá'ció*), *s. f.* Cuidado, esmero, apuro.

Accuratus, a, um (*ákurá'tus*), *p. p.* de Accuro.

Accuro, as, are (*áku'ró*), *v. a.* Cuidar, pôr esmero em. *Prandium alicui accurare.* Preparar o jantar para alguem.

Accurro, is, urri, ursum, ere (*áku'rró*), *v. n.* Accorrer, correr para, affluir. Occorrer, representar-se, offerecer-se.

Accursus, us (*ákur'çus*), *s. m.* Concurso, affluencia.

Accusabilis, le, Accusandus, a, um (*ákuzá'bilis, ákuzan'dus*), *adj.* Accusavel, reprehensivel.

Accusans, antis (*áku'zans*), *adj.* Accusante, accusador.

Accusatio, onis (*ákuzá'ció*), *s. f.* Accusação. O discurso do accusador. *Accusationes Annibalis.* As accusações de que é objecto Annibal.

Accusativus, a, um (*ákuzáti'vus*), *adj.* Accusativo (caso grammatical).

Accusator, oris (*ákuzá'tór*), *s. m.* Accusador, o que accusa.

Accusatoriè (*ákuzátó'riè*), *adv.* Accusatoriamente.

Accusatorius, a, um (*ákuzátó'rius*), *adj.* Accusatorio, relativo á accusação.

Accusatrix, tricis (*ákúzá'trikch*), *s. f.* Accusadora.

Accusatus, a, um (*ákuzá'tus*), *p. p.* de Accuso.

Accusito, as, are (*áku'zitó*), *v. a.* Accusar com insistencia.

Accuso, as, are (*áku'zó*), *v. a.* Accusar. Reprehender, arguir. *Accusare quod...* Accusar de... *In quo te accuso.* O ponto sobre que te accuso.

Acenteta, orum (*ácenté'tá*), *s. n. pl.* Cristaes puros.

Aceo, es, ui, ere (*á'céó*), *v. n.* Estar azedo.

Acephalus, i (*ácé'fálus*), *s. m.* Acephalo.

Acer, eris (*á'cér*), *s. n.* Acer, bordo (arvore).

Acer, acris, acre (*á'cér*), *adj.* Acre, azedo. Forte, violento. Vivo, penetrante; agudo. Ligeiro, fogoso. *Acer splendor.* Muito esplendoroso. *Acer odor.* Cheiro activo. *Acres oculi.* Vista penetrante. *Acris memoria.* Memoria segura. *Acer in rebus agendis.* Cheio de actividade. *Acer equus.* Cavallo fogoso. *Successu acrior.* Animado pelo successo. *Acer rusticus.* O lavrador incansavel.

Aceratus, a, um (*ácerá'tus*), *adj.* V. Acerosus.

Acerbè (*ácér'bé*), *adv.* Acerbamente, asperamente.

Acerbitas, atis, Acerbitudo, inis (*ácér'bitás, ácérbitu'dó*), *s. f.* Acerbidade, aspereza. Angustia.

Acerbo, as, are (*ácér'bó*), *v. a.* Tornar acerbo, azedar. Exasperar. Exaggerar, aggravar. Angustiar.

Acerbum, i (*ácér'bum*), *s. n.* Amargor. Angustia.

Acerbus, a, um (*ácér'bus*), *adj.* Acerbo, amargo, aspero. Prematuro. Severo, cruel. Triste, funesto. Picante, pungente. Incompleto, imperfeito. *Acerba virgo.* Donzella que ainda não está em edade

Fonte: Diccionario... (1930?, p.13).

utilizadas precede um breve prefácio, assinado coletivamente ("os Editores"), em que os autores enfatizam a importância da obra, cuja função não se restringiria a auxiliar estudantes de latim, mas também subsidiar o estudo do português ("quem melhor conhece o latim, melhor saberá o português")[116] (Diccionario…, 1930?, p.3). Assim como Saraiva, afirmam ter se guiado pelo dicionário de Quicherat e Daveluy. Porém, diante da exigência de se manter o volume em proporções comedidas, imposta pela coleção ("sacrifícios importantes que a publicação tem obrigado") e da intenção de enriquecer o dicionário "com o maior numero de phrases para facilitar a traducção dos auctores latinos mais familiares nas escholas", adotam duas estratégias de redução do conteúdo: (a) não descrever particípios em estradas específicas ("diccionario não é grammatica"); (b) suprimir termos antiquados. A principal inovação é informar a pronúncia através de uma representação escrita que segue as regras do português, em vez da tradicional indicação da quantidade vocálica. A respeito dessa "pronúncia figurada", os editores acreditam ser esse sistema de indicação "mais profícuo porque é mais *prosódico*", na medida em que abrange também o som das consoantes. Reconhecendo as controvérsias sobre a "verdadeira pronuncia" do latim decorrentes de seu *status* de "lingua muda", os editores afirmam ter seguido as opiniões "mais autorizadas", sem especificar as fontes, e fornecem alguns detalhes sobre os padrões adotados na representação prosódica: "*s* por *ç*, quando não sôa como *z*; – *gua, gue, gui, guo* por *gha, ghe, ghi, gho*; *qua, que, qui, quo* por *ka, ke, ki, ko*; e *c*, no seu som natural, por *k*" (ibid.).

Em geral, o conteúdo do dicionário latim-português "do povo" é similar ao de Saraiva, embora pareça realmente se tratar de uma tradução renovada da mesma fonte. Para efeito de comparação, o verbete relativo ao lema *accuso* (cf. Figura 16) traz como lema apenas três formas, como Quicherat ('Accuso, as, are'), seguidas pela informação prosódica (*áku'zó*, com acentos tonais sobre as vogais e marcação de sílaba tônica) e uma indicação gramatical sobre o verbo ainda segundo a tradição antiga (*v.a.*, isto é, verbo ativo).[117] A explicação do significado mantém a divisão em duas acepções de Quicherat, mas exclui o delimitador e a numeração, e reproduz apenas dois exemplos traduzidos (contra onze do modelo), o que revela o nível de redução do conteúdo operada pelos editores.

Ao que parece, a coleção "Dicionários do Povo" permaneceu ativa até a década de 1930, com o dicionário latim-português chegando à 12ª edição (Diccionario…, 1930?) em formato de bolso (10x15x4cm). Nesse período, sofre a concorrência do *Novo Diccionario Latino Portuguez* de Francisco António de Souza (1916), publicado inicialmente pela Livraria Francisco Alves e

porais para a primeira edição dos dois volumes do dicionário latim-português. A coleção tem início em 1882, quando foram lançados um dicionário monolíngue do português e um dicionário bilíngue francês-português (nos. 1 e 2, respectivamente); o volume que precede o dicionário de latim (a saber, n. 5, português-inglês) tem como ano de publicação 1888. Logo, os volumes do dicionário latim-português (nos. 6 e 7) foram publicados pela primeira vez entre 1888 e 1910, que é a data da segunda edição (Verdelho; Silvestre, 2011, p.43)

[116] Uma crítica a essa e outras perspectivas do estudo do latim ancorado em interesses extrínsecos ao acesso à cultura escrita em latim se encontra em (Prado, 1992a)

[117] Também aqui a indicação da classe de palavras é sistemática, porém diversa da de Saraiva: substantivos são marcados por um 's.' acompanhado da indicação de gênero ('m.', 'f.', 'n.'); verbos, por 'v.' especificado como ativo, neutro ou depoente ('a.', 'n.', 'd.'); particípios passados, por ('p. p.'); e as demais, pela simples indicação de classe.

ampliado na década de 1950 por José Lello e Edgar Lello, que o mantiveram ativo até pelo menos 1992.[118]

3.2.10 O *Dicionário Latino-Português* de Francisco Torrinha

A década de 1930 marca, na história de Portugal, o início do Estado Novo. Em 1936, a Lei n.º 1.941, que estabeleceu as bases jurídicas da ação estatal em matéria de educação, ciência, cultura, juventude e desporto, instituiu a Junta Nacional de Educação (JNE) para concentrar as funções dos diversos institutos de incentivo à cultura do período republicano.[119] No ano seguinte, publica-se o *Dicionário Latino-Português*, subsidiado pela JNE e elaborado por Francisco Torrinha, "professor catedrático da extinta Faculdade de Letras da Universidade do Porto" (Torrinha, 1942, p. de rosto).

No prefácio, o autor esclarece ter recebido de seus superiores – entenda-se, a JNE – a incumbência de elaborar uma dupla de dicionários bilíngues latim-português e vice-versa, "destinados especialmente ao ensino do latim em estabelecimentos de ensino secundário", o que teria logrado fazer em menos de dois anos (Torrinha, 1942, p.VII). Para tanto, afirma ter utilizado dois dicionários como base, especialmente para a fraseologia: o de Lewis and Short (*A Latin Dictionary*, ed. 1927) – que consiste em uma revisão ampliada da tradução de Freund por Andrews, mencionada antes – e o de Quicherat, Daveluy e Chatelain (ed. 1922), "pelo rigor das citações que encerram e pela boa interpretação que lhes é atribuída" (ibid.). A necessidade de produzir uma obra compacta o levou a omitir toda indicação de referências abonadoras, bastando ao usuário saber que vocábulos e frases foram extraídas dos "melhores autores latinos, principalmente de César, Vergílio [sic], Horácio, Ovídio, T. Lívio, Fedro, Cornélio e poucos mais" (p.VIII). Afirma privilegiar "o significado do vocábulo e o sentido das frases", dispondo as acepções "segundo a ordem de sua evolução semântica" (ibid.); nesse caso, menciona outros dicionários como auxiliares, entre os quais o de Freund e o de Saraiva. Quanto às demais informações descritivas, comenta ter procurado fornecer os elementos necessários para a pronúncia correta, adotando uma estratégia econômica de marcar sistematicamente apenas a quantidade vocálica da penúltima sílaba dos polissílabos, exceto quando a vogal é breve mas a sílaba é longa por posição (p.IX). Também afirma ter registrado a "filiação" (entenda-se, a informação sobre a base lexical de um derivado), bem como a construção sintática de algumas palavras cujo conhecimento lhe pareceu "indispensável para uma mais fácil compreensão dos textos" (p.VIII).

A obra é de fato bastante compacta em comparação com as duas anteriores, não só nas dimensões (947 páginas de três colunas, formato 14x21cm) como no conteúdo, abrangendo aproximadamente 55 mil entradas.[120] Em seu esforço de condensação, Torrinha recorre a dois

[118] Data da edição mais recente encontrada na Estante Virtual (<https://www.estantevirtual.com.br/>). É possível que o lançamento do dicionário na década de 1910 tenha relação com o Decreto nº 8.660, de 5 de abril de 1911, que apresenta as linhas gerais que os programas de ensino deveriam atender.

[119] Pareceu pertinente observar que parte dessas funções são desempenhadas atualmente pelo Instituto Camões, herdeiro indireto da JNE (cf. <https://www.instituto-camoes.pt/sobre/sobre-nos/gestao-sobrenos/historia>).

[120] Valor impreciso, estimado com base na contagem manual de entradas em doze páginas aleatórias.

expedientes aparentemente imitados de Lewis and Short: (1) aglutinar em uma única entrada os lemas primitivos e lemas derivados, posicionando estes como subentradas (e.g. verbos e seus respectivos particípios e formas deverbais; adjetivos e seus respectivos advérbios); (2) reduzir a quantidade de formas do lema verbal para três (suprimindo a segunda pessoa e o infinitivo), indicando o paradigma flexional pelo número tradicionalmente atribuído a cada conjugação. Além disso, reduz a indicação gramatical ao mínimo, indicando apenas a transitividade dos verbos (e.g. 'tr.', 'i.') e o gênero dos substantivos (e.g. 'm.', 'f.'), deixando implícita a informação sobre a classe de palavras; esta é indicada para as demais ('adj.', 'adv.', 'conj.', etc.). Acepções múltiplas são numeradas, ocasionalmente acompanhadas de marcas de uso (e.g. 'absolutus, a, um […] 3. *Gram.* De sentido completo; indeclinável […]').

A Figura 17 reproduz a página do dicionário que contém o verbete para o lema *accuso*, para que se tenha uma ideia da técnica de condensação utilizada por Torrinha a partir da comparação com o verbete análogo de Saraiva. Como mencionado, as formas do lema são apenas três ('accūsō, āvī, ātum'), com a quantidade marcada apenas sobre as vogais longas; em seguida, o número '1' significa 'primeira conjugação', e a etimologia é fornecida entre colchetes, com uma glosa em latim ('ad- + causa := ad causam provocare'); segue-se a indicação de que se trata de um verbo transitivo ('tr.') e a explicação do significado, reduzida a uma única acepção formada por uma série de três equivalentes (os mesmos fornecidos na primeira acepção de Saraiva). Quanto à fraseologia, esta, com exceção da única frase independente de acepção (i.e, 'casus accusandi'), é completamente reduzida a indicações metalinguísticas sobre possibilidades de construção, expressa por meio de pronomes representando tipos de argumento sintático (e.g. 'aliquem alicuius rei', i.e. acusar 'alguém de alguma coisa', etc.).

O dicionário de Torrinha recebeu diversas reedições e reimpressões ao longo do século XX,[121] com boa circulação também no Brasil, particularmente até a terceira edição, de 1945. Nesse período, o mercado de dicionários bilíngues latim-português disponíveis para o uso escolar já incluía os seguintes títulos:

- o *Pequeno Dicionário Escolar Latino-Português*, de Henrique Koehler (1935, formato 15x19cm, com 343 páginas, republicado até 1960);
- o *Dicionário Latino Português* de Dirceu A. Victor Rodrigues (1943, 13x19cm, 702 p.);
- o *Dicionário latim-português* de Otavio A. P. de Queiroz (1943, publicado até 1961);
- o *Dicionário latino-portugues* de José Cretella Júnior e Geraldo de Ulhoa Cintra (1944, 1366 p.);
- o *Pequeno Dicionário Latino-Português* de Fernando de Azevedo (1944, formato 13x20, com 212 páginas, republicado até 1957);

[121] A data de publicação mais recente encontrada é 1994, em volume disponível no catálogo das bibliotecas da Universidade de São Paulo (<https://dedalus.usp.br/>)

Figura 17 – Página do *Dicionário Latino-Portugues* de Torrinha.



ACC / ACE

ção a. **2.** Ocorrer ou surgir repentinamente. **(Constr.:** *a) absol.; b) ac.* com *ad* ou *in).*

accursus, ūs [accurro], *m.* Corrida (para ou em direcção a).

accūsābilis, e [accuso], *adj.* Repreensível; censurável.

accūsātiō, ōnis [id.], *f.* **1.** Acusação. **2.** Discurso do acusador.

accūsātīvus, ī [accusatio], *m.* **1.** Acusativo. **2.** *Adj.: accusativus casus,* acusativo; *accusativa praepositio,* preposição que rege acusativo.

accūsātor, ōris [accuso], *m.* **1.** Acusador. **2.** Delator.

accūsātōriē [accusatorius], *adv.* Como acusador; como quem acusa; com paixão.

accūsātōrius, a, um [accuso], *adj.* Acusatório; de quem acusa.

accūsātrix, īcis [accuso], *f.* Acusadora.

accūsītō, ās, āre [accuso], *tr. freq.* Acusar.

accūsō, āvī, ātum, 1 [ad- + causa: = ad causam provocare], *tr.* Acusar; censurar; repreender. ‖ *casus accusandi:* o acusativo. [**Constr.:** *a) aliquem alicuius rei; b)* mais raras vezes, *aliquem aliquid* (só com *id, illud* e *quod); c) aliquem* e *abl.* regido de *de* ou sem *prep.*].

Acdestis, is, *m.* **1.** Acdeste (monstro nascido dum sonho de Júpiter). **2. Acdestinus,** a, um, *adj.* De Acdeste.

Acē, ēs, *f.* Cidade da Galileia (S. João de Acre).

acēdia, ae, *f.* Negligência; pesar; abatimento; tristeza.

acēdior, āris, ārī, *i. dep.* Ser negligente; sofrer com impaciência; encher-se de pesar.

Acelum, ī, *n.* Cidade da Venécia.

Acenippō = Acinippo.

acentētus, a, um, *adj.* **1.** Sem defeito; puro. **2. acentēta,** ōrum, *n. pl.* Cristais puros, sem defeitos.

acēō, ēre, (sem *sup.*), 2 [ăc- *ou* āc- «ser picante ou agudo»], *i.* **1.** Estar azêdo; ser ácido; azedar. **2.** Ser desagradável.

acephălus, ī, *adj.* **1.** Sem cabeça; sem chefe; acéfalo. **2. Acephălī,** ōrum: Acéfalos (seita de herejes).

1. **acer,** ĕris, *n.* Bôrdo; madeira de bôrdo.

2. **ācer,** cris, cre [ăc-; cf. *aceo*], *adj.* **1.** Agudo; pontiagudo. **2.** Picante (sabor); penetrante; acre. **3.** Enérgico; vivo; ardente. **4.** Violento; fogoso; terrível. **5.** Áspero; cruel; severo. **(Constr.:** *a)* com *abl.,* por vezes regido de *in; b)* com *gen.; c)* com *inf.).*

1. **acerātus,** a, um [acus, eris], *adj.* Misturado com palha.

2. **acerātus,** a, um, *adj.* Sem chifres.

acerba, *pl. de* acerbum, *empregado adv.* Com ferocidade; com dureza.

acerbātiō, ōnis [acerbo], *f.* = acerbitas.

acerbātus, a, um, *part. de* acerbo: tornado cruel.

acerbē [acerbus], *adv.* Duramente; asperamente; severamente; cruelmente.

acerbitās, ātis [id.], *f.* **1.** Agudeza; aspereza; agrura dos frutos verdes. **2.** Rigor; severidade; mau humor; violência. **3.** Angústia; aflição; pesar. **4.** Cólera; aborrecimento; ódio. **5.** Calamidade; desgraça; **6.** Espírito mordaz; graça picante. **7.** Infecção.

acerbitūdō, ĭnis [id.], *f.* = acerbitas.

acerbō, āvī, ātum, 1 [id.], *tr.* **1.** Tornar acre; tornar amargo. **2.** Agravar.

acerbōsitās, ātis, *f.* = acerbitas.

acerbum, ī [acerbus], *n.* **1.** Amargor; azedume. **2.** Calamidade; desgraça.

acerbus, a, um, *adj.* **1.** Azêdo; áspero ao paladar; verde (não maduro). **2.** Prematuro; imperfeito; que não está em idade de casar. **3.** Amargo; agudo; cruel; hostil. **4.** Molesto; penoso; incómodo.

acernĕus, a, um [1. acer], *adj.* = acernus.

acernia, ae, *f.* Nome de um peixe desconhecido.

acernus, a, um [1. acer], *adj.* De bôrdo.

acerōsus, a, um [acus, ĕris], *adj.* Cheio de palha; misturado com palha; grosseira (farinha).

acerra, ae, *f.* **1.** Acerra (caixa em que se guardava o incenso); incenso. **2.** Altar sôbre o qual se queimava incenso, perante um morto.

Acerrae, ārum, *f. pl.* **1.** Acerras, cid. da Campânia; cid. da Úmbria. **2. Acerrānī,** ōrum, *m. pl.* Habitantes de Acerras.

acersecōmēs, ae, *m.* Jovem querido; jovem.

acōrus, a, um, *adj.* Sem cera; que corre do cortiço (mel).

acervālis, e [acervus], *adj.* Em montão; amontoado. ‖ *argumentatio acervalis:* sorites.

acervātim [id.], *adv.* **1.** Em montão; em conjunto; conjuntamente. **2.** Em resumo; sumariamente.

acervātiō, ōnis [acervo], *f.* Amontoação; acumulação.

acervō, āvī, ātum, 1 [acervus], *tr.* Amontoar; acumular.

acervus, ī, *m.* **1.** Montão. **2.** Grande quantidade. **3.** Sorites.

acēscō, ăcŭī, (sem *sup.*), 3. [aceo], *i. inc.* Tornar-se azêdo.

Acesinēs, ae *ou* is, *m.* Acesina, rio da Índia.

Acesinus, ī, *m.* **1.** Acesino, rio da Crimeia. **2. Acesinus,** a, um, *adj.* Do Acesino.

acēsis, is, *f.* Bórax natural, proveniente duma árvore da Índia.

Acesta, ae *ou* **Acestē,** ēs, *f.* **1.** Segesta, cid. da Sicília. **2. Acestaeus,** a, um, *adj.* De Segesta. **3. Acestaei,** ōrum *ou* **Acestēnsēs,** ium, *m. pl.* Habitantes de Segesta.

Acestēs, ae, *m.* Acestes, rei da Sicília.

acētābŭlum, ī [acetum], *n.* **1.** Vinagreira; acetábulo. **2.** Medida correspondente a um quarto da hemina. **3.** Qualquer objecto com a forma duma vinagreira; trompa do polvo, cavidade em que encaixa um ósso, cálice das flores, copo de pelotiqueiro.

acētābŭlus, ī = acetabulum.

acētāria, ōrum [acetum], *n. pl.* Salada; legumes temperados com vinagre e azeite. **(Obs.** — Por *acetaria olera).*

— 10 —

Fonte: Torrinha (1942, p.10).

- o *Dicionário latino vernáculo* de Marques Leite e Novaes Jordão (1944, com 528 páginas, publicado até a 3ª ed. 1958);
- o *Dicionário latino português* de Nicolau Firmino (1945, com publicação até 1994).

Essa profusão de dicionários nas décadas de 1930 e, especialmente, de 1940 tem como pano de fundo duas tentativas de estabelecer parâmetros nacionais de organização curricular no Brasil: a Lei Francisco Campos (Decreto nº 19.890/1931) e a Lei de Capanema (Decreto no 4.244/1942). Em cada uma delas, o papel atribuído ao ensino de latim se altera, sofrendo ora com a perda de espaço devido à concorrência com as línguas modernas (principalmente o francês e o inglês), ora com a falta de profissionais qualificados para seu ensino (Santos Sobrinho, 2013). A primeira restringe a presença do latim no ensino médio mas constitui o sistema de universidades federais, com a inserção do ensino de latim nos cursos de Letras, Filosofia e Direito. A segunda resgata o espaço das línguas clássicas no secundário, consideradas como base para os estudos humanísticos modernos, uma concepção que se alinhava com a educação humanística europeia (Kaltner; Santos, 2024).

Na década seguinte, é a vez de o Estado brasileiro subsidiar a produção de um dicionário de latim, que se tornaria a principal referência para os estudantes brasileiros de latim a partir de então.[122]

3.2.11 O *Dicionário Escolar* de Ernesto Faria

A redemocratização que se seguiu a partir de 1945 deu ensejo a discussões voltadas para a erradicação do analfabetismo, a expansão da rede de ensino e a organização do ensino técnico e profissional; na década de 1950, a escola já recebia alunos de camadas sociais até então excluídas do sistema de ensino e, consequentemente, há a demanda por uma reestruturação do ensino secundário e mais investimentos na formação de professores; além disso, segundo Filgueiras (2013), parte da discussão incidia sobre o preço dos livros didáticos, que os tornava inacessíveis às camadas populares; a resposta veio em dois tempos. Primeiro, o Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), então dirigido por Anísio Teixeira, criou em 1952 a Campanha do Livro Didático e Material de Ensino (CALDEME), a fim de prestar assistência técnica aos professores por meio da elaboração de instrumentos para o trabalho docente, com distribuição a preços acessíveis visando diminuir o custo de ensino (Teixeira, 2024, p.171). Quatro anos mais tarde, já no governo Kubitschek, o Departamento Nacional de Educação (DNE) institui a Campanha Nacional de Material de Ensino (CNME), agora com o objetivo não só de melhorar a qualidade do material de ensino, mas também de "combater a elevação do custo do ensino,

[122] Os dados sobre os dicionários que circulavam no Brasil na primeira metade do século XX foram obtidos de forma indireta, com base nos anúncios de exemplares à venda no mercado de livros usados, em particular na loja *online* Estante Virtual (<https://www.estantevirtual.com.br>). Com exceção dos dicionários de Torrinha e de Faria, descrito a seguir, não foi possível até o momento ter acesso aos demais dicionários para verificar as informações apresentadas, bem como descrever suas características.

com o fornecimento de materiais didáticos a preço acessível para assistir os estudantes carentes" (Filgueiras, 2013, p.318).[123]

O programa inicial previa a edição de dois dicionários: o *Dicionário Escolar da Língua Portuguesa*, de Francisco da Silveira Bueno, e o *Dicionário Escolar Latino-Português*, de Ernesto de Faria. A escolha dos autores não é gratuita: ambos eram pessoas envolvidas nos debates educacionais das décadas de 1950 e 1960 e estavam vinculados a instituições que atuavam na formação docente (Filgueiras, 2013, p.330).[124] De fato, quando é incumbido de produzir um dicionário escolar latim-português, na década de 1950, Ernesto Faria já era um latinista consagrado, catedrático de Língua e Literatura Latina da Faculdade Nacional de Filosofia da Universidade do Brasil. Com quase três décadas de experiência como professor, era reconhecido por seu papel como renovador da didática das línguas clássicas, sendo considerado pelo linguista Eugenio Coseriu como um dos nomes da geração vencedora da batalha pela linguística científica (Valle, 2006).

Com uma obra consolidada a respeito do lugar do latim na educação contemporânea, em que buscava conciliar a tradição pedagógica das letras clássicas com o conhecimento das ciências modernas (Marques, 1947), sendo defensor fervoroso da pronúncia reconstituída ou restaurada, estabelecida com base nos estudos de fonética histórica (Freitas, 2006), também tinha experiência como lexicógrafo, tendo publicado em 1943 o *Vocabulário Latino-Português*, em que apresenta a "significação e a história das palavras, agrupadas por famílias" (Faria, 1943, p. de rosto). Nessa obra, parte da observação de que a aquisição de vocabulário constitui "a maior, senão talvez a única, dificuldade no aprendizado de latim" (Faria, 1943, p.viii-ix), donde o objetivo de "facilitar o conhecimento do léxico latino por meio de uma exposição sistemática desse léxico" (ibid.) por um método que consiste em agrupar as palavras por famílias, considerado não somente o mais racional, mas também o mais adaptado aos "novos programas de latim para o curso ginasial, […] que prescrevem para todas as séries" (ibid.) esse tipo de abordagem. Mesmo sendo uma obra para ser usada no ensino elementar, o autor ressalta ter recorrido "aos melhores trabalhos referentes à lexicografia latina, como sejam os dicionários de Freund, Gaffiot, Ernout-Meillet, Walde-Hofmann, Bréal-Bally, Lewes-Short [sic], Seccaroni, Saraiva, Torrinha […]" (ibid.) para sua confecção.

O *Dicionário Escolar Latino-Português*, por sua vez, é um dos primeiros dicionários produzidos e distribuídos por um órgão estatal no Brasil (Teixeira, 2024), destinado a suprir a necessidade de material didático por parte dos estudantes de latim. Num primeiro momento (as duas primeiras edições são, respectivamente, de 1955 e 1957), tratava-se de atender aos estu-

[123] Vale dizer que o projeto educacional da era Vargas também visara a produção, avaliação e distribuição de materiais didáticos, num contexto de investimentos na formação de uma consciência nacional que incluía "a produção de instrumentos linguísticos ou de saberes nacionais como a enciclopédia e o dicionário brasileiros" (Teixeira, 2024, p.162). Com esse objetivo foi criado o Instituto Cayrú em 1936, convertido em Instituto Nacional do Livro em 1937, que pode ser visto como órgãos estatais precursores do CNME.

[124] A série contou ainda com o *Dicionário Escolar Inglês-Português Português-Inglês*, do professor Oswaldo Serpa, e o *Dicionário Escolar Francês-Português Português-Francês*, do professor Roberto Alvin Corrêa (Filgueiras, 2013).

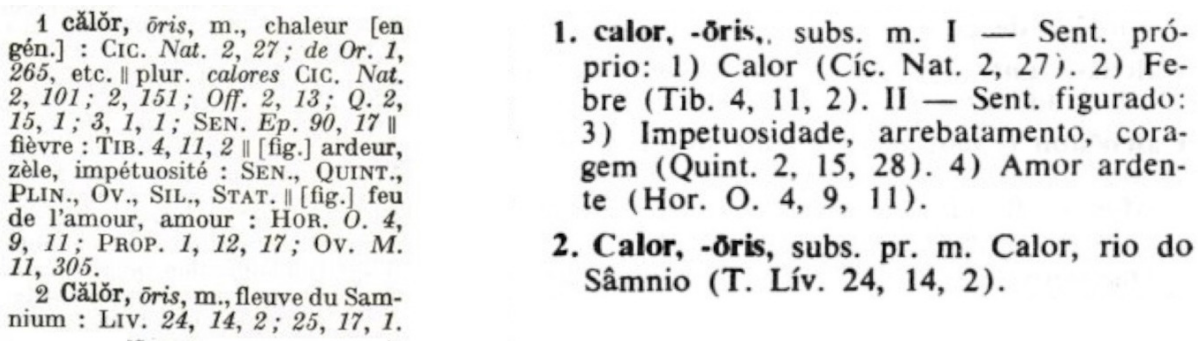
dantes dos cursos ginasial e médio (‘clássico’); após a aprovação da Lei de Diretrizes e Bases de 1961,[125] o público-alvo se restringe aos “estudantes dos cursos de grau médio”, para quem se proporciona “uma obra de alto padrão por preço acessível”, em conformidade com os objetivos da CNME, segundo o prefácio assinado por sua então diretora, Heloísa Araújo (Faria, 1962). Apesar dessa restrição, o dicionário de Faria alcançou a marca de 180 mil exemplares nas três primeiras edições (Filgueiras, 2013, p.320) e permanece sendo um dos principais dicionários de latim, tendo sua última reimpressão vindo à luz em data relativamente recente (pela Editora Garnier em 2020).

Na ausência de um prefácio do autor, pode-se supor que a compilação do *Dicionário Escolar* parte das mesmas fontes de informação declaradas no prefácio ao *Vocabulário*. Dentre a dúzia de dicionários mencionados, o *Dictionnaire Latin-Français* de Félix Gaffiot (1934) parece ter sido a principal. Essa escolha acompanha os movimentos do mercado editorial na França após a Primeira Guerra, onde o sucesso do grande dicionário ilustrado de Gaffiot, bem como de sua versão compacta (o *Abregé* de 1936), havia dominado o cenário dos estudos latinos, praticamente relegando os concorrentes ao esquecimento (Flobert, 2001, p.401). Reconhecendo-se tributário dos trabalhos lexicográficos de Freund, Georges e Lewis-Short, Gaffiot declara no prefácio que o *Dictionnaire* responde à incumbência, recebida dos editores, de compor “um dicionário latino similar ao dicionário grego de Bailly que contemplasse tanto as exigências escolares quanto as exigências científicas” (Gaffiot, 1934, s.p.). Para tanto, restringe a nomenclatura ao léxico da “latinidade em sentido geral” (i.e. da Lei das Doze Tábuas ao *Digesto* de Justiniano), com especial atenção à ”latinidade no sentido verdadeiro (isto é, a que vai de Plauto a Tácito)”, mas insere como entradas “todas as formas especiais que possam causar dificuldade no aluno” (e.g. formas verbais, comparativos e superlativos, etc.); já a descrição inclui significados gerais e particulares, expressões e construções acompanhados muitas vezes de exemplos referenciados e traduzidos, outras ao menos de uma referência para abonação. Para Flobert (2001), o modo preciso de referenciar os autores e traduzir os exemplos constituem inovações de Gaffiot em relação aos predecessores, bem como a tipografia arejada e a presença de ilustrações tiradas do *Dicionário das Antiguidades*.

Uma análise comparativa dos dois dicionários mostra o quanto Faria é tributário de Gaffiot, ao mesmo tempo que revela aspectos de seu trabalho de redução e adaptação do conteúdo. A Figura 18 mostra os verbetes relativos ao lema *calor* nos dois dicionários. As semelhanças e diferenças de microestrutura apontam para as características herdadas e adaptadas do modelo, respectivamente. Quanto ao lema, é comum a distinção numerada de homônimos, o uso de maiúsculas no lema para indicar nomes próprios e a indicação de quantidade silábica; porém, Faria expande sistematicamente a indicação de classe e gênero gramatical (‘subs. m.’, ‘subs. pr. m.’, bem como ‘adj.’, ‘adv.’, ‘v. tr.’) que em Gaffiot se resume apenas ao que é específico de subs-

[125] A Lei nº 4.024, de 24 de dezembro de 1961, revoga a Reforma Capanema e suprime o estudo do latim do currículo ginasial, tornando facultativa sua inclusão nas séries do ensino médio. A tristeza causada em Ernesto Faria pelo “duro golpe” que essa reforma representou para o estudo de latim costuma ser incluída entre as razões de sua morte, em 14 de março de 1962 (Valle, 2006, p.22).

Figura 18 – Colagem comparativa dos verbetes para *calor* nos dicionários de Gaffiot e Faria, respectivamente.



1 **călŏr,** *ōris*, m., chaleur [en gén.] : Cic. *Nat. 2, 27 ; de Or. 1, 265*, etc. ‖ plur. *calores* Cic. *Nat. 2, 101 ; 2, 151 ; Off. 2, 13 ; Q. 2, 15, 1 ; 3, 1, 1 ;* Sen. *Ep. 90, 17* ‖ fièvre : Tib. *4, 11, 2* ‖ [fig.] ardeur, zèle, impétuosité : Sen., Quint., Plin., Ov., Sil., Stat. ‖ [fig.] feu de l'amour, amour : Hor. *O. 4, 9, 11 ;* Prop. *1, 12, 17 ;* Ov. *M. 11, 305.*

2 **Călŏr,** *ōris*, m., fleuve du Samnium : Liv. *24, 14, 2 ; 25, 17, 1.*

1. calor, -ōris, subs. m. I — Sent. próprio: 1) Calor (Cíc. Nat. 2, 27). 2) Febre (Tib. 4, 11, 2). II — Sent. figurado: 3) Impetuosidade, arrebatamento, coragem (Quint. 2, 15, 28). 4) Amor ardente (Hor. O. 4, 9, 11).

2. Calor, -ōris, subs. pr. m. Calor, rio do Sâmnio (T. Lív. 24, 14, 2).

Fontes: Gaffiot (1934, p.247); Faria (1994, p.90).

tantivos (‘m.’) e verbos (‘tr.’); por outro lado, não é tão sistemático quanto Gaffiot na indicação de quantidade silábica. Quanto ao significado, Faria traduz as quatro acepções fornecidas por Gaffiot, incluindo uma reformulação expandida das marcas de caracterização do significado encontradas no modelo de forma abreviada: os delimitadores de sentido geral e figurado (‘en gén.’ e ‘fig.’, respectivamente) ganham seções em números romanos (‘I - Sent. próprio’ e ‘II - Sent. figurado’), envolvendo cada uma duas acepções, numeradas de 1 a 4; além disso, cada acepção recebe múltiplas referências bibliográficas em Gaffiot, não raro incluindo as citações acompanhadas de tradução, ao passo que Faria normalmente omite as citações e escolhe uma única referência para abonação do significado.

Faria também reduz substancialmente a nomenclatura, aproveitando pouco menos da metade das 82 mil entradas e subentradas de seu modelo.[126] Seu dicionário contempla um total aproximado de 30 mil itens lexicais descritos em pouco mais de 34 mil entradas, sem utilizar o recurso de subentrada; a diferença entre número de itens e de entradas se deve às numerosas entradas para formas flexionadas, principalmente os pretéritos verbais considerados menos óbvios. A maior parte das entradas descreve lemas univerbais: são apenas 395 entradas com lemas multipalavra e 30 entradas para afixos. Substantivos respondem por 50% das entradas, um terço das quais se referem a nomes próprios. Lemas adjetivos, incluindo particípios, contabilizam outros 30% das entradas, ao passo que os verbos constituem apenas 16% (ou 10%, caso se excluam as entradas para formas flexionadas como pretéritos e variantes). Por fim, assim como Torrinha, em muitos casos compensa a ausência de citações por indicação de construção, concentradas no final do verbete.

O dicionário de Faria permaneceu com publicação ativa mesmo após o estudo de latim perder espaço na educação brasileira, em decorrência das diversas reformas curriculares ocorridas no Brasil na segunda metade do século XX. Além de suas obras lexicográficas, voltadas para atender às necessidades dos estudantes, Faria também procuraria fornecer orientação e suporte aos docentes por meio de duas obras, uma de caráter linguístico (a *Gramática Superior*

[126] Cálculos feitos a partir da edição digital de Gréco (2016) e do texto digitalizado de Faria (1994) – este gentilmente compartilhado pela Profa. Dra. Giovanna Longo para este trabalho –, mediante aplicação parcial da estruturação de dados descrita no Capítulo 3.

da Língua Latina, de 1958), outra de caráter didático (*Introdução à didática do latim*, de 1959) (Kaltner; Santos, 2024).

Depois de Faria, talvez o único dicionário latim-português mais substancial seja o de Ferreira (1966), com suas mais de 1,2 mil páginas em formato médio (16x23cm), cuja edição revisada por António Rodrigues de Almeida continua ativa no catálogo da Editora Porto. Além desse, podem-se mencionar repertórios mais modestos, como os de Pena (1959), Busarello (1988), Garcia e Castro (2003), Silva e Montagner (2009), Botelho (2014), Rezende e Bianchet (2014) e Vieira e Hasegawa (2016). Aparentemente nenhum deles apresenta novidades no que diz respeito à descrição lexicográfica e, uma vez que não se encontram em domínio público, não podem contribuir como fonte para a base de dados de acesso aberto que se pretende construir. Em contrapartida, é o momento de considerar as novas formas de registrar e descrever o léxico proporcionadas pelas tecnologias computacionais no âmbito da lexicografia eletrônica.

3.3 LÉXICOS COMPUTACIONAIS

A despeito de seu caráter tradicional, o estudo das línguas clássicas foi pioneiro no campo das Humanidades a fazer uso da computação eletrônica para fins de pesquisa, já em meados do século XX. O primeiro projeto com esse tipo de abordagem foi o *Index Thomisticus*, uma empresa do jesuíta italiano Roberto Busa iniciada no final da década de 1940 que visava construir um índice completo das palavras presentes na obra em latim de Tomás de Aquino por meio do emprego de métodos computacionais, então incipientes (Hockey, 2004). Nas décadas seguintes, o tratamento estatístico de *corpora* permitiu a produção de dicionários frequenciais como os de Gardner (1971) e Delatte et al. (1981), assim como evidenciou que, no contexto computacional, dicionários tenderiam a se transformar em bases de dados lexicais (Sedelow, 1985).[127] Com a disseminação dos computadores pessoais na década de 1990, houve um incremento na criação de dicionários eletrônicos, isto é, ferramentas digitais de consulta lexical inicialmente distribuídas através de mídias físicas (como disquetes e CD-ROMs) e posteriormente disponibilizadas *online* através da Internet.[128]

O termo ‘lexicografia computacional’ pode se referir, portanto, a duas abordagens distintas (Hanks, 2022): (1) a transformação de dicionários impressos para consulta por meio de computadores; (2) o uso de técnicas computacionais para produzir novos dicionários. A primeira é mais comum e se apresenta em diferentes formatos e graus de legibilidade por máquina, da simples imagem digitalizada da página até a edição digital do conteúdo de forma estruturada. A segunda, mais recente, consiste em extrair dados lexicais diretamente do processamento de *corpora* anotados. No âmbito da lexicografia latina, apenas a primeira abordagem se encontra bem

[127] Exemplo dessa transformação é o *Latin Core Vocabulary* (Francese, 2020), um léxico computacional que parte do ranking de frequência de Delatte et al. (1981) para construir uma fonte de informação sobre formas canônicas, classe de palavras, domínio semântico, significado básico em diversas línguas dos mil lemas mais frequentes no latim clássico.

[128] Um retrospecto do desenvolvimento histórico da lexicografia eletrônica na segunda metade do século XX encontra-se em Tasovac (2020).

desenvolvida, cujos exemplos mais conhecidos serão apresentados a seguir; da segunda serão citados projetos promissores que ainda aguardam novos desenvolvimentos para se concretizarem como dicionários consultáveis.

3.3.1 Dicionários tradicionais em formato digital

A maior parte dos repertórios lexicais e dicionários tradicionais de latim descritos neste capítulo estão disponíveis para acesso e visualização em algum tipo de formato digital. Essa variedade de formatos depende do processo de digitalização utilizado – isto é, a "conversão de um sinal ou código analógico em sinal ou código digital" (Terras, 2012, p.47) –, que pode envolver desde a simples reprodução das páginas em formato de imagem (por meio de escaneamento, microfilme ou fotografia) até as edições digitais de textos (mediante anotação de conteúdo e inclusão de metadados), passando pelo formato de texto simples obtido por digitação manual ou reconhecimento automatizado de caracteres (OCR) a partir das imagens.

A digitalização recebeu um grande investimento a partir dos anos 1990, acompanhando o aumento do acesso à internet, do menor custo e melhor performance das tecnologias digitais e da maior percepção das possibilidades da tecnologia da informação (Terras, 2012). Grandes projetos de preservação digital de documentos resultaram na criação de repositórios e bibliotecas digitais, sendo as mais conhecidas o *Internet Archive*[129], o *Google Books*[130] e o *HathiTrust Digital Library*[131]; diversos dicionários de latim descritos neste capítulo podem ser encontrados em seus catálogos. Alguns itens específicos podem ser localizados em coleções especializadas, como a *Biblioteca Nacional Digital de Portugal*[132] para dicionários latim-português e os arquivos do projeto *Mirabile* (Lunardini, 2019; Pinelli, 2019)[133] para a extensa produção lexicográfica medieval. A maior parte desse material se encontra em formato imagem, normalmente reunida em um arquivo único do tipo Adobe PDF (Portable Document Format); logo, são limitados em termos de recursos para a consulta e localização de conteúdo, normalmente requerendo a passagem manual das páginas, como se faz no livro impresso. Um recurso eventualmente utilizado para agilizar a consulta é indexar as imagens das páginas pela primeira palavra que contêm ou pelas letras iniciais desta, como é feito em *Le Gaffiot numérisé*.[134]

Formatos de imagem são mais comuns devido à facilidade e baixo custo de produção; porém, são os materiais em formato de texto que possibilitam a exploração do conteúdo por ferramentas ordinárias de busca textual ou especializadas em processamento de *corpus*. Para Wooldridge (2004), essa possibilidade constitui o efeito mais radical do computador sobre a lexicografia, na medida em que agrega à quantidade limitada de formas de acesso à informação previstas pelo autor (baseadas em palavras-entrada) uma quantidade ilimitada de caminhos

[129] Cf. <https://archive.org/>.
[130] Cf. <https://books.google.com/>.
[131] Cf. <https://www.hathitrust.org/>.
[132] Cf. <https://bndigital.bnportugal.gov.pt/>
[133] Cf. <https://mirabileweb.it/content/lessici_ml_list>.
[134] Cf. <https://www.prima-elementa.fr/Gaffiot/>.

possíveis de serem estabelecidos pelo usuário (baseados em palavras-chave). Todavia, ainda que grande parte dos arquivos PDF mencionados acima incluam uma versão em texto digitalizado (obtido automaticamente por OCR), a falta de revisão ou controle de qualidade impede que sejam consideradas uma fonte de informação confiável.

Edições digitais acadêmicas, por outro lado, distinguem-se dos demais tipos de texto digital por sua capacidade de servir à pesquisa científica, na medida em que informam tanto a referência bibliográfica da fonte quanto os critérios e procedimentos editoriais adotados no processo de digitalização; além disso, o uso de tecnologias e licenças de acesso aberto propicia um maior alcance de público e um menor risco de obsolescência, devido à contínua possibilidade de compartilhamento, reuso e conversão dos dados em novos formatos (Nyhan, 2012). Na verdade, tais edições são elas próprias resultantes de pesquisa acadêmica e costumam estar vinculadas a projetos institucionais.[135] No caso dos repertórios e dicionários tradicionais de latim, podem-se citar:

- o projeto *Corpus Corporum* (Roelli, 2014), que reúne edições digitais de textos latinos provenientes de projetos específicos como o de Garcea, Plancq e Cinato (2010) para os gramáticos antigos ou o de Lunardini (2019), Pinelli (2019) para os repertórios lexicais medievais (como os de Isidoro e Hugúcio) ou renascentistas (como o de Erasmo);[136]
- o projeto *Lexicons of Early Modern English* (Lancashire, 2003), que contém um grande acervo de dicionários antigos relativos à língua inglesa, entre os quais se encontram os dicionários latim-inglês de Elyot, Cooper e Thomas;[137]
- o *Project for American and French Research on the Treasury of the French Language – ARTFL* (Morrissey, 1993), uma parceria do CNRS francês e da Universidade de Chicago, que traz em seu catálogo a edição digital do *Dictionarium Latinogallicum* de Estienne;[138]
- o projeto *CAMENA: Latin Texts of Early Modern Europe* (Schibel; Rydberg-Cox, 2006), que contém diversos dicionários de latim, com destaque para a edição do *Thesaurus* de Estienne por Gesner e o dicionário de latim medieval de Du Cange;[139]
- o projeto *Perseus Digital Library* (Crane, 2012), sediado na Tufts University (Boston, EUA), que inclui a edição digital do dicionário latim-inglês de Lewis and Short;[140]

[135] Embora em menor número, também são dignas de nota as iniciativas individuais de digitalização de dicionários de latim por acadêmicos como Diego A. Rodriguez Garcia, que provê o texto do *Lexicon* de Nebrija (cf. <https://github.com/Diegeau>) e Gérard Gréco, que liderou a revisão do texto digital do *Dictionnaire* de Gaffiot (cf. <https://gaffiot.fr/>) a partir das imagens digitalizadas por M. Okubo Katsuhiko; além disso, cabe uma menção ao catálogo especializado em dicionários de latim disponíveis em formato digital mantido por Nikita Murzintcev (cf. <https://latin-dict.github.io/>).

[136] Cf. <http://mlat.uzh.ch/>.

[137] Cf. <https://leme.library.utoronto.ca/>.

[138] Cf. <https://artfl-project.uchicago.edu/content/dictionarium-latinogallicum>

[139] Cf. <https://mateo.uni-mannheim.de/camenaref/gesner.html> e <https://mateo.uni-mannheim.de/camenaref/ducange.html>, respectivamente.

[140] Cf. <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059>.

- o projeto *Logeion* (Crane, 2012), sediado na University of Chicago (EUA), que integra em uma única plataforma de consulta uma série de dicionários monolíngues e bilíngues de latim, que fornecem ao seu modo explicações e equivalentes em inglês, francês, alemão, holandês e chinês;[141]

- o *Corpus Lexicográfico do Português* (CLP), um projeto de investigação da Universidade de Aveiro que teve por objeto "a inventariação, registo em base de dados e disponibilização dos textos do fundo lexicográfico e paralexicográfico do português" (Verdelho; Silvestre, 2007, p.207), fornecendo acesso aos textos digitais dos dicionários de Cardoso, Velez, Pereira, Fonseca, entre outros.

Parte desses projetos segue os padrões para edições digitais definidos pela Text Encoding Initiative (TEI) (Ide; Véronis, 1995), cuja grande influência no campo das Humanidades digitais se deve justamente ao fato de fomentar a utilização, pelos projetos de digitalização, de metadados padronizados para tornar legível por máquina a informação veiculada nos textos e sobre os textos, incluindo especificações sobre os critérios e procedimentos adotados (Nyhan, 2012). Para um participante da história da digitalização de dicionários desde os primórdios da tecnologia de fichas perfuradas como Wooldridge (2004), a anotação de metadados no padrão TEI tem como vantagem não apenas a exibição correta do texto na tela, mas também a possibilidade de pesquisar e referenciar o texto por campos e categorias de conteúdo, na medida em que permite agregar anotações e distinções semânticas. É essa capacidade de referenciação específica que propicia, por exemplo, integrar em uma única plataforma de consulta diversos dicionários de latim (como é feito pelo projeto *Logeion*, por exemplo) ou disponibilizar a consulta ao dicionário como ferramenta integrada de ambientes ou assistentes de leitura (como nos casos da *Perseus Digital Library* e do aplicativo *Alpheios*[142], respectivamente). Com isso, o resgate da função fundamental do dicionário (isto é, de servir à compreensão de textos), preconizado por Tarp (2019), encontra um modo de concretização.

Nos últimos anos, tem se consolidado um novo padrão para interconectar recursos linguísticos computacionais, desenvolvido em torno do conceito de Linked Data (Berners-Lee; Lassila, 2001). Alguns dicionários tradicionais tiveram seu conteúdo modelizado segundo esse novo paradigma e conectados a uma plataforma de integração que permite a pesquisa simultânea de dicionários e atestações em *corpora*. Nesse caso, o conteúdo dos dicionários impressos tornam-se parte de um grande repertório lexical que, embora seja em parte construído sobre o legado da lexicografia tradicional, pode ser considerado uma inovação da lexicografia computacional, merecendo ser tratado em um tópico à parte.

[141] Cf. <https://logeion.uchicago.edu/about>.
[142] Cf. <https://alpheios.net/>

3.3.2 Uma base de conhecimento interligado para o Latim

O século XXI foi palco de importantes inovações na arquitetura da rede mundial de computadores conhecida como Internet. Uma das mais importantes foi a chamada Web 2.0 ou Web Semântica, que transfere o foco da conexão entre documentos digitais para a conexão entre os dados e informações contidos no interior desses documentos (Berners-Lee, 2007), a fim de promover interatividade entre a informação produzida para consumo humano e a produzida para processamento por máquina. Nesse novo modelo de dados interligados (*Linked Data*), cada conceito pertence a uma ontologia e é identificado por um nome que corresponde ao seu endereço eletrônico (URL), cujas partes identificam simultaneamente o responsável pelo provedor de serviços e o documento em que a informação (ou definição) se encontra, fornecendo ao usuário não só a informação que ele procura mas também uma espécie de "respaldo social" (*social backing*) para essa informação, composto por aspectos como proveniência, confiabilidade, persistência, direitos e possibilidades de uso.

A associação entre esse modelo e a cultura do acesso aberto gera o paradigma *Linked Open Data* (LOD) (Berners-Lee, 2006), que recomenda que os dados publicados na Web tenham preferencialmente como características: (1) estarem disponíveis sob licença aberta (e.g. Creative Commons); (2) serem estruturados e legíveis por máquina (e.g. uma planilha, não uma imagem ou fotografia); (3) não utilizarem formatos patenteados de armazenamento (e.g. texto separado por caracteres, e não arquivos do Microsoft Word ou Excel); (4) utilizarem formatos abertos e padronizados pelo consórcio W3C para descrição dos dados (e.g. RDF e SPARQL);[143] (5) identificarem os dados de acordo com uma ontologia, isto é, um vocabulário especializado para representação do conhecimento (e.g. OntoLex-lemon para dados sobre o léxico das línguas naturais, cf. adiante). Um segundo conjunto de recomendações é constituído pelos princípios FAIR, sigla formada pelos termos em inglês correspondentes à capacidade de um dado ou documento ser encontrável (*findability*), acessível (*accessibility*), interoperável (*interoperability*) e reusável (*reusability*) (Wilkinson et al., 2016).

O padrão Linked Open Data de estruturar e interligar dados foi utilizado pelo projeto *LiLa–Linking Latin* para construir uma base de conhecimento linguístico sobre o latim, a LiLa Knowledge Base (LiLaKB) (Passarotti et al., 2020). Nessa base de dados, recursos computacionais textuais e lexicais são descritos por meio de um vocabulário comum de representação do conhecimento, o que permite um alto grau de interoperabilidade entre eles, ampliando as formas

[143] The World Wide Web Consortium (W3C) é um consórcio internacional de interesse público e sem fins lucrativos formado por diversas organizações e aberto ao público e voltado para o desenvolvimento dos padrões utilizados na Web (cf. <https://www.w3.org/about/>). RDF é a sigla para Resource Description Framework, um padrão da W3C para representar, descrever e compartilhar dados cuja estrutura permite a caracterização de recursos na Web por meio de triplas, isto é, expressões compostas por sujeito-predicado-objeto em que o sujeito é normalmente um recurso digital representado por um identificador exclusivo de recurso (URI), o predicado é uma relação ou propriedade atribuída ao sujeito, definida por uma ontologia e também identificada por URI, e o objeto é o valor atribuído à propriedade, que pode ser tanto um recurso digital identificado por URI ou uma descrição na forma de texto (cf. <https://www.w3.org/RDF/>). Já o SPARQL Protocol and RDF Query Language (SPARQL) é a linguagem padrão de consulta ao dados armazenados no formato RDF (cf. <https://www.w3.org/TR/rdf-sparql-query/>).

Tabela 1 – Amostra de itens lexicais com suas propriedades no LiLa Lemma Bank.

id	wr	pos	fl-cat	gen	num	grad
lilaLemma:102758	facio	verb	v5r			
lilaLemma:51124	facio	noun	n3	feminine		
lilaLemma:51125	faciola	noun	n1	feminine		
lilaLemma:51126	faciolus	noun	n2	masculine		
lilaLemma:51128	facionatus	adjective	n6	*		
lilaLemma:51133	faciom, facium	noun	n2n	neuter		
lilaLemma:51139	factionarii	noun	n2	masculine	pt	
lilaLemma:130507	ultimus, ultumus	adjective	n6s	*		Abs

Legenda: **id** (identificador); **wr** (formas gráficas); **pos** (classe de palavras); **fl-cat** (padrão flexional); **gen** (gênero). O prefixo 'lilaLemma:' representa a parte inicial do URI ('<http://lila-erc.eu/data/id/lemma/>').
Fonte: <https://github.com/CIRCSE/LiLa_Lemma-Bank>

de explorar a informação que portam de modo integrado. Na base dessa interoperabilidade está o fato de que *corpora* e *lexica* têm em comum a possibilidade de serem lematizados. Logo, a LiLaKB se estrutura em torno de uma coleção de lemas que funciona como uma grande nomenclatura, à qual se interconectam tanto as ocorrências presentes nos *corpora* quanto as entradas para cada lema nos recursos lexicais. Em outras palavras, a LiLaKB é como um grande dicionário construído sobre um esquema modular, em que a diversidade de recursos interconectados fornece diferentes tipos de informação lexical.

Como se disse, a nomenclatura é constituída pelo *LiLa Lemma Bank*, atualmente com quase 170 mil lemas latinos. Formado inicialmente pela combinação das nomenclaturas de três dicionários de referência para o latim – a saber, o *Onomasticon* de Forcellini para nomes próprios, DuCange para o léxico medieval e o *Oxford Latin Dictionary* para o vocabulário clássico (Passarotti et al., 2020) –, esse banco de dados tem sido progressivamente aumentado pela inclusão contínua de itens lexicais na medida em que novos recursos lexicais e textuais são conectados à plataforma. A cada lema são atribuídos um identificador unívoco (URI) e uma série de propriedades formais. Algumas delas são obrigatórias para todos os itens (como a forma gráfica, a classe de palavras e o padrão flexional), outras são específicas de certa classe (como o gênero dos substantivos), outras ainda são eventuais (como o número, no caso de substantivos lematizados no plural, ou o grau, no caso de adjetivos lematizados no comparativo ou superlativo). A Tabela 1 oferece uma amostra dos dados.

Uma lista adicional é constituída pelos chamados 'hipolemas', que representam palavras derivadas que, ocasionalmente, podem ser utilizadas como lemas nas anotações de *corpus*. São consideradas hipolemas as formas participiais dos verbos (e.g. *facio* > *faciens, factus, facturus*) e as formas adverbiais derivadas de adjetivos (e.g. *facilis* > *facile, faciliter*). Além da relação entre lemas e hipolemas, alguns lemas se encontram relacionados entre si como variantes morfológicas, quando os itens lexicais possuem formas canônicas alternativas ou então participam de mais de um paradigma flexional (e.g. o substantivo *domus* possui duas entradas vinculadas

Tabela 2 – Amostra de dados extraídos do PrinParLat para o lema verbal *facio*.

lexeme	cell	orth_form	analysed_orth_form
f0048	prs.act.inf	facere	fac+ere
f0048	fut.act.ind.3.sg	faciet	fac+iet
f0048	prf.act.ind.1.sg	fecei	fece+i
f0048	prf.act.ind.1.sg	feci	fec+i
f0048	prf.pass.ptcp.nom.n.sg	factum	fact+um
f0048	fut.act.ptcp.nom.n.sg	facturum	factur+um

Fonte: <https://github.com/CIRCSE/PrinParLat/blob/main/forms.csv>

como variantes na nomenclatura, uma da segunda declinação, outra da quarta).

Além disso, em torno de 45 mil lemas da coleção recebem informação sobre etimologia e formação de palavras, provida pelo recurso *Word Formation Latin (WFL) lexicon* (Litta; Passarotti; Mambrini, 2019). Esse recurso associa os lemas a seus elementos formadores (bases lexicais, prefixos e sufixos) e define a sequência de transformações envolvidas no processo de derivação. Nessa estrutura, as bases lexicais servem como um ponto de conexão entre palavras pertencentes a uma mesma família derivacional.

Já o *PrinParLat* (Pellegrini et al., 2023) é o módulo que informa sobre as demais formas canônicas dos lemas verbais além da principal, fornecendo diferentes radicais que participam da morfologia flexional dos verbos latinos. Em outras palavras, esse recurso documenta a flexão verbal latina por meio de "partes principais", ou seja, o conjunto de formas a partir das quais todo o paradigma flexional e derivacional pode ser inferido. Embora o conceito não seja novo, já que corresponde em grande parte ao conjunto de formas que constituem a entrada dos dicionários tradicionais de latim, o recurso tem a vantagem de apresentar formas completas (ie. não fragmentadas) e contemplar todas as variantes gráficas de cada lema. A Tabela 2 oferece uma amostra dos dados provenientes deste recurso.

Quanto à informação sobre o significado, a LiLaKB incorpora o sistema de conceitos e sinônimos da Latin WordNet (Franzini et al., 2019). As chamadas WordNets são recursos léxico-semânticos em que itens pertencentes às classes de palavras com significado lexical (substantivos, verbos, adjetivos e advérbios) são agrupados em conjuntos de sinônimos (*synsets*) associados a conceitos demarcados por um identificador e descritos por uma definição (Boschetti, 2019), de modo que cada *synset* representa o conjunto de formas lexicalizadas de um conceito. A iniciativa original, desenvolvida na Universidade de Princeton (Miller, 1995; Fellbaum, 1998), consiste em uma base de dados lexicais da Língua Inglesa, estruturados manualmente em forma de rede 'para representar palavras e conceitos como um sistema relacional […] aparentemente compatível com o modo como os falantes organizam seu léxico mental" (Fellbaum, 1998, p.7). Sua inovação consiste principalmente na variedade de relações semânticas que estabelece entre palavras e conceitos, baseadas em similaridade e contraste. Duas delas, *sinonímia* e *antonímia*, são relações simétricas que se estabelecem entre palavras de qualquer das classes descritas

(verbo, substantivo, adjetivo e advérbio). Outras se aplicam apenas a substantivos: *hiponímia* e *hiperonímia* são relações transitivas entre *synsets*, ao passo que *meronímia* e *holonímia* são relações semânticas complexas que qualificam os modos de um substantivo designar a parte de um todo (a saber, como componente, como substância ou como membro). Outras duas se aplicam exclusivamente a verbos: a *troponímia* se assemelha à hiponímia, no sentido de que o significado de certos verbos acrescentam um componente específico (normalmente indicando modo) ao significado de um verbo tido como genérico (e.g. 'caminhar' e 'correr' em relação a 'mover-se'); e, por fim, a relação de *implicação* vincula verbos como 'casar' e 'divorciar-se', em que a realização do segundo implica a realização prévia do primeiro. Além disso, a cada *synset* é atribuído um domínio semântico (e.g. 'animal', 'artefato', 'sentimento' no caso de substantivos; 'movimento', 'percepção', 'mudança' no caso de verbos)[144] e, particularmente no caso dos verbos, um *frame* valencial com argumentos representados por formas pronominais (e.g. 'Somebody —-s something', 'Something —-s to somebody' e outras 33 possibilidades).

A WordNet de Princeton (PWN) motivou a criação de bases estruturadas semelhantes para outras línguas, sendo os projetos mais conhecidos a EuroWordNet (Vossen, 2002) e a MultiWordNet (Pianta; Bentivogli; Girardi, 2002). Cada uma representa um modelo diferente de construção e alinhamento com os *synsets* da WordNet original. Como explica Boschetti (2019, p.326), o modelo utilizado pela EuroWordNet consiste em construir redes léxico-semânticas específicas para cada língua e, num segundo momento, estabelecer as correspondências entre elas. Já o modelo adotado pela MultiWordNet consiste em construir *synsets* para as novas línguas por meio da tradução da WordNet original, feita geralmente com o auxílio de dicionários bilíngues e multilíngues, preservando assim toda a rede de relações pré-estabelecidas para o inglês. Embora esse segundo método se preste a críticas, na medida em que condiciona a organização e descrição do léxico ao recorte do mundo em conceitos, palavras e significados específicos da Língua Inglesa[145], ele tem a vantagem de atender à necessidade crescente de interoperabilidade entre as diversas *wordnets* em pesquisas e aplicações multilíngues (Piasecki et al., 2013). No nosso caso, essa correspondência entre *synsets* é utilizada para obter informação contrastiva entre o latim e o português, através das respectivas *wordnets* desenvolvidas para essas línguas.

Segundo Franzini et al. (2019), a Latin WordNet (LWN) foi desenvolvida inicialmente na

[144] Tecnicamente, as etiquetas que designam domínios semânticos na WordNet de Princeton (PWN) são formadas pela classe de palavras a que se associam seguida da especificação semântica, sendo 25 referentes a substantivos ('noun.act', 'noun.animal', 'noun.artifact', 'noun.attribute', 'noun.body', 'noun.cognition', 'noun.communication', 'noun.event', 'noun.feeling', 'noun.food', 'noun.group', 'noun.location', 'noun.motive', 'noun.object', 'noun.person', 'noun.phenomenon', 'noun.plant', 'noun.possession', 'noun.process', 'noun.quantity', 'noun.relation', 'noun.shape', 'noun.state', 'noun.substance', 'noun.time') e 15 referentes a verbos ('verb.body', 'verb.change', 'verb.cognition', 'verb.communication', 'verb.competition', 'verb.consumption', 'verb.contact', 'verb.creation', 'verb.emotion', 'verb.motion', 'verb.perception', 'verb.possession', 'verb.social', 'verb.stative', 'verb.weather'). Cf. <https://wordnet.princeton.edu/download>.

[145] Uma crítica se encontra em Steurs et al. (2020, p.15), que reconhecem o fato de ainda não haver uma resposta sobre como o conteúdo específico de dicionários científicos e suas descrições detalhadas de significado e uso para línguas específicas poderiam ser completamente vinculados e integrados em uma infraestrutura de conhecimento multilíngue"; os autores indicam como solução ideal buscar "conceitos e *frames* semânticos universais e começar a construção de bases de conhecimento multilíngue a partir deles".

Universidade de Verona por Minozzi em 2004[146] por meio de técnicas de expansão da MultiWordNet, que consistem na tradução automática para o latim dos pares alinhados da WordNet em inglês e italiano, com o auxílio de dicionários bilíngues. Vale dizer que, ao contrário de outras propostas de *wordnet*, a LWN não traduz conceitos para o latim, mas apenas associa a estrutura conceitual da PWN ao léxico latino por meio dos identificadores de *synset*.[147] Para sua inclusão na LiLaKB, o trabalho de Minozzi foi revisado e atualizado; o mapeamento entre os *synsets* da LWN e os identificadores do *LiLa Lemma Bank* associa aproximadamente 15 mil *synsets* a 8540 lemas.[148] A Tabela 3 apresenta uma amostra do conteúdo da Latin WordNet associada à LiLaKB.[149]

Por fim, a informação sintática é fornecida pelo *Latin Vallex*, um recurso que descreve a valência de aproximadamente 1,3 mil itens, especialmente verbos.

A primeira versão desse léxico valencial foi produzida a partir da anotação semântica e pragmática de trechos de dois *treebanks* de textos latinos, a saber, o *Index Tomisticus treebank* e o *Latin Digital Treebank* (Passarotti; Saavedra; Onambele, 2016). Sua estrutura, baseada no estilo de anotação utilizado nas árvores de dependência de Praga (PDT), se caracteriza por construir as camadas de anotação semântica e pragmática sobre a camada sintática, tendo como marco teórico a descrição gerativa funcional (Functional Generative Description ou FGD). A informação valencial é associada às acepções individuais de cada item lexical e assume a forma de papéis semânticos (chamados *functors* na FGD) usados para expressar tipos de relação entre lemas e seus complementos, que podem ser: Agente (ACT), Paciente (PAT), Destinatário (ADDR), Efeito (EFF) e Origem (ORIG), Lugar(partida) (DIR1) e Lugar(destino) (DIR3). A valência prototípica do verbo *uenio* (‘chegar’), por exemplo, pode ser descrita como formada por três argumentos, expressos na forma [ACT, DIR1, DIR3].

Uma versão ampliada desse léxico foi desenvolvida para ser integrada à LiLaKB a partir do alinhamento entre a informação valencial, os *synsets* da PWN e os lemas do *LiLa Lemma Bank* (Mambrini et al., 2021b). Ao contrário da primeira versão, que era baseada nos contextos extraídos de um *corpus* anotado, o Latin Vallex 2.0 estabelece os *frames* de forma intuitiva, com base nos diversos *synsets* associados a cada item, o que diminui o grau de confiança dos

[146] Cf. Minozzi (2017).

[147] A mesma estratégia foi utilizada para produzir *wordnets* da Língua Portuguesa, como a OpenWordNet-PT (Paiva; Rademaker; Melo, 2012), que utiliza os identificadores de *synsets* da WordNet original e está disponível para *download* em formato estruturado sob licença aberta.

[148] O *dataset* completo está disponível em <https://github.com/CIRCSE/latinWordnet-revision>.

[149] A despeito da clareza com que a polissemia é tratada pela estrutura da WordNet, com definições precisas atribuídas a cada significado associado ao lema, o método indireto de identificação de significados pode gerar um número aparentemente excessivo de acepções. Isso ocorre normalmente com os verbos, aos quais se atribuem dezenas de *synsets*, cuja pertinência parece depender de contextos de uso específicos, comprováveis apenas com base em pesquisas de *corpus*. Verbos de alta frequência como *agere* (31 *synsets*), *dare* (41 *synsets*) e *ferre* (54 *synsets*) são os exemplos mais evidentes. Não obstante, o conjunto de dados fornecidos para cada *synset* permite explorar novas perspectivas de descrição do léxico, seja fornecendo um identificador unificado de acepções entre diferentes dicionários, seja fornecendo um parâmetro para a descrição sintático-semântica. McGillivray (2014), por exemplo, aplica as categorias semânticas da WordNet na detecção de padrões de preferência selecional para verbos.

Tabela 3 – Amostra de dados extraídos da versão da Latin WordNet publicada na LiLaKB para o lema verbal *video*.

lemma	lila-id	synset-id	definition
uideo	lilaLemma:130163	wn30:02198234-v	appear to one's own mind or opinion
uideo	lilaLemma:130163	wn30:00720808-v	realize beforehand
uideo	lilaLemma:130163	wn30:00693780-v	keep in mind or convey as a conviction or view
uideo	lilaLemma:130163	wn30:02488488-v	go to see for a social visit
uideo	lilaLemma:130163	wn30:00598954-v	get to know or become aware of, usually accidentally
uideo	lilaLemma:130163	wn30:02128873-v	perceive or be contemporaneous with
uideo	lilaLemma:130163	wn30:02151966-v	be vigilant, be on the lookout or be careful
uideo	lilaLemma:130163	wn30:02134672-v	seem to be true, probable, or apparent
uideo	lilaLemma:130163	wn30:02129289-v	perceive by sight or have the power to perceive by sight
uideo	lilaLemma:130163	wn30:02133435-v	give a certain impression or have a certain outward aspect
uideo	lilaLemma:130163	wn30:01635432-v	imagine; conceive of; see in one's mind
uideo	lilaLemma:130163	wn30:00623151-v	make sense of; assign a meaning to
uideo	lilaLemma:130163	wn30:02156225-v	observe as if with an eye
uideo	lilaLemma:130163	wn30:00689950-v	look on as or consider
uideo	lilaLemma:130163	wn30:00920336-v	find out, learn, or determine with certainty, usually by making an inquiry or other effort

Fonte: <https://github.com/CIRCSE/latinWordnet-revision/blob/master/LiLa_LatinWordnet.csv>

Tabela 4 – Amostra de dados extraídos do LatinVallex 2.0 para o lema verbal *sum*.

lemma	lila-id	synset-id	argumentos
sum	lilaLemma:126689	00344643-v	ACT
sum	lilaLemma:126689	02005948-v	ACT,LOC
sum	lilaLemma:126689	02072849-v	ACT
sum	lilaLemma:126689	02301680-v	ACT,ADDR
sum	lilaLemma:126689	02603699-v	ACT
sum	lilaLemma:126689	02614181-v	ACT
sum	lilaLemma:126689	02649830-v	ACT,LOC
sum	lilaLemma:126689	02664769-v	ACT

Fonte: <https://github.com/CIRCSE/Latin_Vallex2.0/blob/master/export_LV.tsv>

dados.[150] A Tabela 4 fornece uma amostra dessa associação para o lema verbal *sum*.

Nesse complexo sistema de dados interconectados, também há espaço para o conteúdo de dicionários tradicionais, incluídos na LiLaKB por meio de sua modelização segundo os padrões

[150] A documentação do recurso, disponível em <https://github.com/CIRCSE/Latin_Vallex2.0> não deixa claro até que ponto tais *frames* são versões recodificadas dos *frames* baseados em formas pronominais da Princeton WordNet.

do *Lexicon Model for Ontologies* ou, mais simplesmente, OntoLex-lemon (Cimiano; McCrae; Buitelaar, 2016). A experiência pioneira de Mambrini et al. (2021a), que aplicou o modelo sobre o conteúdo extraído da versão TEI do dicionário latim-inglês de Lewis and Short, serviu de referência para a modelização de diversos outros dicionários latinos, entre os quais estão três dicionários latim-português – a saber, Cardoso (1570), Velez (1744), Fonseca (1798) –, cujo conteúdo foi extraído dos textos digitais disponíveis no CLP e convertidos em dados estruturados (Dezotti; Passarotti; Mambrini, 2024).

3.3.3 Léxicos dinâmicos extraídos de *corpora* anotados

A abundância de dados interconectados e a exaustividade da lista de lemas que caracterizam a LiLaKB são, em parte, frutos da incorporação dos dicionários da tradição impressa. Para Bamman e Crane (2009), as edições digitais desses dicionários significam, para a lexicografia computacional, o que os incunábulos da tradição manuscrita significaram nos primórdios da tecnologia de impressão, isto é, uma mera transposição de suporte material, sem grande avanço no conteúdo da informação. As propostas de léxicos inovadores, produzidos exclusivamente pela extração automática de dados de *corpora*, existem mas esbarram na carência de *corpora* anotados em quantidade suficiente para uma compilação abrangente.

Um exemplo é o *Syntacticus*, uma plataforma de consulta aos *treebanks* do *PROIEL* (Haug; Jøhndal, 2008), um projeto interessado na anotação sintática de textos escritos em línguas antigas do tronco indo-europeu, incluindo o latim. Uma das funcionalidades da plataforma é a compilação de um repertório de itens lexicais presentes no *corpus* associados a informações sobre a morfologia, os padrões sintáticos e a cronologia das ocorrências. O caráter inovador da informação obtida diretamente dos textos contrasta com o caráter limitado da nomenclatura, restrita aos lemas que ocorrem nos poucos textos incluídos no *corpus*.[151]

A mesma dificuldade é obstáculo para a realização de um projeto arrojado de dicionário proposto por Bamman e Crane (2009), que envolve a produção de um léxico dinâmico que combine informação sobre valência e preferências selecionais extraídas de *treebanks* com informação sobre equivalentes de significado extraídas de *corpora* paralelos com tradução alinhada. Esse modelo de dicionário eletrônico inovador, totalmente produzido a partir de recursos computacionais e, portanto, independente de informações fornecidas por dicionários tradicionais, depende todavia de uma quantidade significativa de *corpora* sintaticamente anotados e com tradução alinhada, infelizmente ainda indisponível.

3.4 CONCLUSÃO DO CAPÍTULO

A lexicografia latina se desenvolveu, no decorrer de dois milênios, sobre a compilação e reorganização do conhecimento sobre o léxico latino acumulado. Em cada época há um conjunto

[151] A saber, os *Commentarii belli Gallici* de César, as *Epistulae ad Atticum* e o *De officiis* de Cícero, partes da *Vulgata* de Jerônimo de Stridon, a *Opus agriculturae* de Paládio e o anônimo *Peregrinatio Aetheriae*.

de interesses que move a produção de dicionários. Na Antiguidade, repertoriavam-se os desvios da língua literária em relação à língua corrente, quer se tratasse de empregos do passado, quer de formulações de natureza retórica. Na Idade Média, buscava-se sistematizar o léxico enquanto rede de formas primitivas e derivadas, ao mesmo tempo que se compilavam grandes coleções de glosas para o significado das palavras de uma língua histórica, conhecida apenas por meio de aprendizado formal. No Renascimento, o objetivo de aproximar o latim falado como língua franca do latim falado no tempo dos autores clássicos, bem como fornecer explicação para as palavras encontradas nos textos literários, cada vez mais numerosos, leva à produção tanto de repertórios exaustivos monolíngues para fins acadêmicos (como os *thesauri*), quanto de repertórios seletivos bilíngues para fins pedagógicos (como os dicionários escolares). Na Modernidade, ganha relevo a leitura e interpretação correta dos autores clássicos segundo os métodos da filologia, que fornece novas edições dos textos assim como critérios para a definição de uma história dos significados. A época contemporânea, por sua vez, assiste à consolidação das tecnologias digitais como veículos da informação, com dicionários tradicionais digitalizados e léxicos produzidos por processamento computacional formando uma rede de conexões entre conhecimento acumulado e conhecimento novo.

A lexicografia latino-portuguesa impressa, em particular, manteve um diálogo produtivo com a produção europeia dominante. Na primeira fase, humanista, a importação de modelos se traduz em experiências autênticas de compilação, seja por meio de uma combinação de fontes secundárias (como em Cardoso e, de forma quase exaustiva, em Pereira), seja mediante compilações mais autênticas (como a de Velez a partir do conteúdo da gramática ou de Roboredo a partir das frases da *Iannua*). Na segunda, pós-reformas pombalinas, observa-se uma tendência de os lexicógrafos elegerem um único modelo estrangeiro como base e produzirem dicionários a partir de uma tradução de seu conteúdo, eventualmente adaptado e abreviado; assim se posicionam os dicionários de Fonseca e Cabral em relação ao italiano Pasini, o de Saraiva e o de Faria em relação aos franceses Quicherat e Gaffiot, respectivamente.

Em suma, os contextos exercem sua influência na prática lexicográfica, orientando os compiladores quanto à seleção da nomenclatura e do conteúdo explicativo a ser fornecido. Logo, cada dicionário ou léxico apresenta uma combinação específica de conteúdos que justifica sua utilização como fonte de conhecimento sobre o léxico latino, ao passo que a combinação deles pode resultar em uma descrição mais completa desse léxico.

No caso de um dicionário bilíngue latim-português em formato digital, de cuja produção este trabalho se ocupa, trata-se principalmente de congregar conjuntos de dados de diferentes fontes. De um lado, a LiLaKB pode fornecer a nomenclatura exaustiva e sistematicamente descrita quanto às propriedades gramaticais e lexicais do LiLa Lemma Bank, as formas canônicas do PrinParLat e a relação entre lema e *synset*, que permite conectar latim e português à rede de relações lexicais e conceituais estabelecidas pela Princeton WordNet. De outro, os dicionários latim-português da tradição impressa fornecem especialmente informação contrastiva autorizada, seja por meio das séries de equivalentes de tradução para o lema que caracterizam as

acepções, seja por meio de um repertório valioso de frases e expressões idiomáticas traduzidas para o português, ainda inexistente como recurso computacional.

Do ponto de vista técnico, porém, a viabilidade dessa integração depende da forma como tais informações são organizadas e armazenadas em uma base de dados estruturada, em formato legível por máquina. Os métodos utilizados para construir essa base de dados, bem como para produzir uma ferramenta de consulta a esses dados, constituem o assunto dos próximos capítulos.

4 COMPILAÇÃO DE UMA BASE DE DADOS

Nos capítulos precedentes, foram apresentados um projeto de dicionário digital baseado em uma combinação de fontes primárias e secundárias, bem como um levantamento amplo destas últimas no interior da tradição da lexicografia bilíngue latim-português. Os próximos dois capítulos apresentam, respectivamente, a metodologia utilizada para construir uma base de dados legível por máquina a partir de uma seleção dessas fontes heterogêneas e a metodologia utilizada para produzir uma interface para a ferramenta de consulta lexicográfica, que corresponde ao dicionário digital propriamente dito.

Como se disse anteriormente, o trabalho de seleção e conversão de um conjunto de dados legíveis por máquina em informação organizada para consulta humana é auxiliado por um *Dictionary Writing System* (DWS), uma ferramenta computacional que automatiza o processo de compilação e garante uniformidade e consistência na produção de cada verbete, permitindo ao lexicógrafo dedicar mais tempo à coleta, análise e classificação de dados linguísticos (Abel, 2012, p.83). O propósito de produzir um dicionário digital a partir de múltiplas fontes e acessível pela Web dentro de um contexto de conhecimento aberto exige um DWS de uso gratuito e versátil, que permita a utilização de diferentes bases de dados e favoreça a elaboração de uma interface de consulta adequada a essa diversidade. Até o momento, o aplicativo que parece melhor atender a tais necessidades é o *lexicographR* (Lugli, 2024b), uma ferramenta gratuita de edição e publicação de dicionários que tem como principal característica o fato de propiciar ao lexicógrafo controle total sobre o processo produtivo, atuando desde a determinação do *corpus* e da base lexical até o planejamento e produção dos elementos que devem constituir os verbetes, sejam eles elementos textuais ou visualizações gráficas. De fato, o aplicativo provê meios para utilizar uma base de dados lexicográficos pré-existente, enriquecer essa base com dados extraídos de *corpora*, produzir dados estatísticos e criar visualizações a partir desses dados e, por fim, compilar os dados em uma interface de consulta pronta para publicação na Web. Uma das vantagens desse sistema é sua grande flexibilidade, que se traduz na possibilidade de experimentar funcionalidades e soluções na forma de protótipos e, consequentemente, adaptar rapidamente o dicionário ao *feedback* obtido junto aos usuários, trazendo assim agilidade ao processo de desenvolvimento (Lugli, 2024a).[1]

O *lexicographR* possui uma interface gráfica escrita em linguagem de programação R (R Core Team, 2023).[2] As diversas funcionalidades se distribuem como etapas de um fluxo de processamento que se inicia com a importação dos dados e termina com a criação da interface

[1] Para o conceito de agilidade no desenvolvimento de *softwares*, cf. Ambler (2002), Conboy e Fitzgerald (2004). A rigidez do processo de publicação e a dependência de uma equipe de desenvolvimento para implementar mudanças nas bases de dados ou no layout são desvantagens elencadas por Lugli (2024a) no uso de outras plataformas mais conhecidas como o Lexonomy (Měchura, 2017) e o Sketch Engine (Kilgarriff et al., 2014).

[2] Neste trabalho, todos os aplicativos e *scripts* em linguagem R mencionados foram desenvolvidos e executados utilizando-se o ambiente integrado de desenvolvimento Rstudio (Posit Team, 2023).

de consulta (ver Figura 19).

O aplicativo requer a importação de dados estruturados, em formato de planilha, que podem ser de dois tipos: dados lexicais, indexados por lemas; e *corpora* de textos anotados ou, no mínimo, lematizados. Uma vez que a lematização é usada como elo de integração das diversas fontes de dados, uma planilha especial de dados lexicais deve fornecer ao programa a base do dicionário, isto é, a lista de lemas que serão utilizados como indexadores para a extração de dados do *corpus*, como nomenclatura para a compilação dos verbetes e como (macro)estrutura de acesso para a interface de consulta. Em outras palavras, o aplicativo organiza os dados em uma base de dados relacional (*relational database*),[3] em que os valores de uma coluna comum (no caso, a coluna 'lemma') servem de chave de ligação entre as diversas planilhas de dados.

As seções a seguir descrevem a preparação da base do dicionário e das demais planilhas para sua devida importação pelo *lexicographR*. Tendo em vista as especificidades quanto ao tipo de pré-processamento necessário, a descrição foi dividida conforme o tipo de fonte de que provêm os dados, isto é, conforme os dados tenham sido obtidos de léxicos computacionais, de dicionários tradicionais ou de *corpora* de textos.

4.1 DADOS DE LÉXICOS COMPUTACIONAIS

A planilha de dados lexicais que define a base do dicionário tem como requisito mínimo conter a lista de lemas, representados por valores únicos (i.e. sem repetição), de preferência ocupando uma coluna específica de um arquivo em formato planilha; na prática, cada linha da planilha corresponde a um único lema, declarado na coluna 'lemma', sem redundâncias ou ambiguidades. Assim, caso se decida incluir dados descritivos nessa mesma planilha, estes devem ser constituídos de valores igualmente únicos, para que não ocorra a duplicação de valores na coluna principal.

A princípio não há limites para o tamanho da nomenclatura, isto é, para a quantidade de lemas a ser incluída na macroestrutura do dicionário. Todavia, a abrangência da nomenclatura pode ser determinada pela capacidade do equipamento utilizado para produzir o dicionário – no caso, um computador de uso pessoal que se mostrou capaz de completar o processamento de dados para um conjunto de aproximadamente mil lemas.[4] Dada a importância do critério da frequência para a definição da nomenclatura, elegem-se para compor a nomenclatura deste protótipo do **eLatim** os mil lemas mais frequentes no latim clássico, com base na lista fornecida pelo *Latin Core Vocabulary* (Francese, 2020).[5]

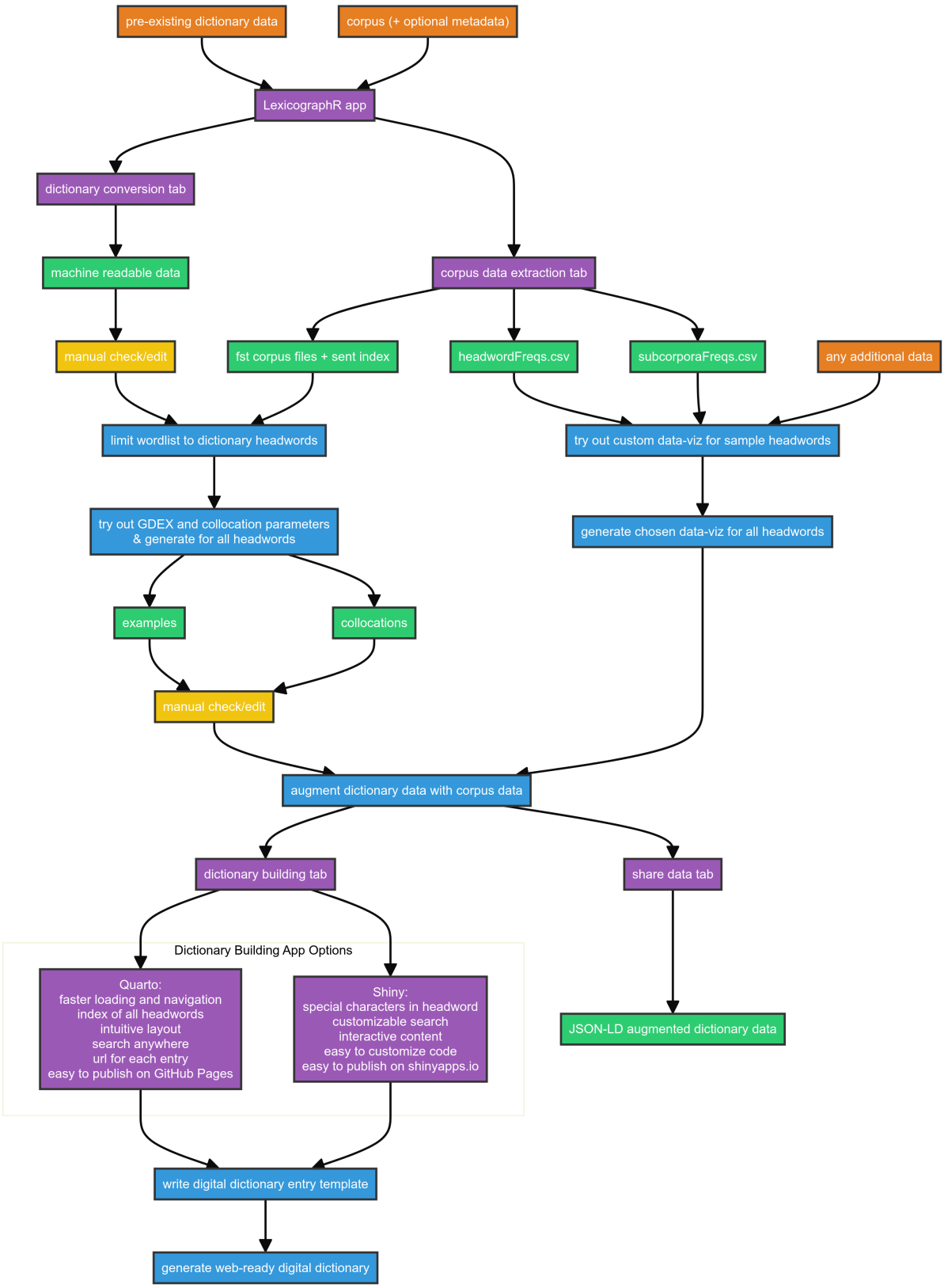
O *Latin Core Vocabulary* fornece, além da lista de lemas, informação sobre classe de palavras e paradigma flexional, domínio semântico e posição no ranking de frequência de cada lema. Todavia, a forma como lida com a homografia entre lemas – fazendo a distinção através de

[3] Para o conceito de base de dados relacional, cf. Baker, Hardie e McEnery (2006)

[4] Trata-se de um MacBook Air (13-inch, 2017), com processador 1,8 GHz Dual-Core Intel Core i5, memória de 8 GB 1600 MHz DDR3, rodando o sistema operacional macOS Monterey, v.12.7.6 (21H1320).

[5] Disponível em <https://dcc.dickinson.edu/latin-core-list1>.

Figura 19 – Fluxograma de processamento de dados do aplicativo *lexicographR*.



Fonte: Lugli (2024b).

informações suplementares à forma principal, seja pelas demais formas canônicas (e.g. "mundus -ī" vs. "mundus -a -um"), seja pela designação de classe de palavras (e.g. "ne […] enclitic" vs. "ne […] conjunction") – não é adequada à utilização da forma principal como indexador, na medida em que produz ambiguidade. Esse problema é resolvido tipicamente adicionando-se um algarismo ao final dos lemas ambíguos menos frequentes, de modo que o lema *mundus* designe o substantivo (350º no ranking de frequência), ao passo que *mundus2* designe o adjetivo (997º no mesmo ranking); pelo mesmo critério se distinguem os pares *adeo* e *adeo2*, *aduersus* e *aduersus2*, *eo* e *eo2*, *hic* e *hic2*, LATUS e *latus2*, *liber* e *liber2*, *licet* e *licet2*, *mundus* e *mundus2*, *ne* e *ne2*, *os* e *os2*.

A intenção de utilizar os diversos recursos computacionais disponíveis na base de conhecimento do projeto Linking Latin (LiLa) requer o mapeamento e alinhamento prévio do *Latin Core Vocabulary* com os identificadores do *LiLa Lemma Bank*. Esse trabalho é feito por meio de um processo semi-automatizado de pareamento de formas (*string matching*), que compara a lista de lemas do primeiro com o catálogo completo do segundo;[6] para tanto, utiliza-se um programa em linguagem R escrito especialmente para esse fim. A fim de aumentar a precisão e reduzir o número de resultados ambíguos ou falso-positivos, os lemas do *Latin Core Vocabulary* são normalizados e recebem etiquetas gramaticais segundo os padrões do *LiLa Lemma Bank*; na prática, a padronização da ortografia consiste basicamente em conversão para minúsculas, substituição de caracteres (*j* por *i*, *v* por *u*), supressão de diacríticos e expansão de abreviaturas, ao passo que as etiquetas gramaticais dizem respeito principalmente à classe de palavras e ao paradigma flexional. Os resultados são classificados em pares únicos (1:1), pares ambíguos (1:N) e itens sem pareamento (1:0); os dois últimos são pareados manualmente.

Isso resulta em uma planilha contendo 1.010 lemas acompanhados de dados descritivos obtidos pela combinação dos dois recursos, traduzidos para o português; o identificador único do lema e as informações sobre classe de palavras, paradigma flexional e propriedades gramaticais como gênero dos substantivos, bem como variantes ortográficas e lexicais, são extraídos do *LiLa Lemma Bank*; a posição no ranking de frequência e a informação sobre grupo semântico são extraídas do *Latin Core Vocabulary*. A Tabela 5 oferece uma amostra simplificada da planilha base do dicionário.

Além da base do dicionário, outros dados lexicais obtidos de léxicos computacionais podem ser fornecidos por planilhas específicas. Três recursos foram explorados para a compilação do protótipo do **eLatim**: o conjunto de formas verbais do *PrinParLat* (Pellegrini et al., 2023); as formas canônicas com marcação prosódica do *Lexicala* (Kernerman, 2018); e a rede de conceitos da *Latin WordNet Revision* (Litta Modignani; Pedonese; Franzini, 2023).

No caso do *PrinParLat*, os dados originais são selecionados e reunidos em uma única planilha, contendo o lema e as formas básicas da conjugação acompanhadas de descrição morfológica; uma amostra é fornecida pela Tabela 6.

[6] Disponível em <https://github.com/CIRCSE/LiLa_Lemma-Bank>.

Tabela 5 – Amostra da planilha de dados que serve de base para a compilação do dicionário.

	lemma	lila_id	classe	flexao	genero	grupoSemantico
16	si	lilaLemma:124676	conjunção subordinativa	indeclinável		
17	sui	lilaLemma:131255	pronome	pronominal		
18	omnis	lilaLemma:114954	determinante	2ª classe (tema em -i-)		Medidas
19	nec	lilaLemma:113512	conjunção coordenativa	indeclinável		
20	sed	lilaLemma:123994	conjunção coordenativa	indeclinável		
21	ab	lilaLemma:86824	preposição	indeclinável		Lugar
22	ipse	lilaLemma:108945	determinante	1ª classe, pronominal (tema em -a/o-)		
23	possum	lilaLemma:118569	verbo	irregular		Fazer, Criar
24	aut	lilaLemma:90827	conjunção coordenativa	indeclinável		
25	magnus	lilaLemma:111319	adjetivo	1ª classe (tema em -a/o-)		Medidas
26	ex	lilaLemma:101716	preposição	indeclinável		Lugar
27	suus	lilaLemma:127165	determinante	1ª classe (tema em -a/o-)		
28	do	lilaLemma:99970	verbo	1ª conjugação (tema em -a-)		Sociedade
29	qua	lilaIpoLemma:110471	advérbio	indeclinável		
30	per	lilaLemma:116569	preposição	indeclinável		Lugar
31	uideo	lilaLemma:130163	verbo	2ª conjugação (tema em -e-)		Visão
32	facio	lilaLemma:102758	verbo	5ª conjugação (tema em -ĭ- breve)		Fazer, Criar
33	dico	lilaLemma:99302	verbo	3ª conjugação (sem vogal temática)		Discurso, Oratória
34	iam	lilaLemma:106025	advérbio	indeclinável		Tempo
35	atque	lilaLemma:90490	conjunção coordenativa	indeclinável		
36	alius	lilaLemma:88481	determinante	1ª classe (tema em -a/o-)		
37	res	lilaLemma:121868	substantivo	5ª declinação (tema em -e-)	feminino	Vida e Existência
38	habeo	lilaLemma:105044	verbo	2ª conjugação (tema em -e-)		
39	animus	lilaLemma:89094	substantivo	2ª declinação (tema em -o-)	masculino	Vida e Existência
40	meus	lilaLemma:112262	determinante	1ª classe (tema em -a/o-)		
41	deus	lilaLemma:99132	substantivo	2ª declinação (tema em -o-)	masculino	Religião
42	multus	lilaLemma:112998	determinante	1ª classe (tema em -a/o-)		Medidas
43	tuus	lilaLemma:129328	determinante	1ª classe (tema em -a/o-)		
44	fero	lilaLemma:103126	verbo	irregular		Verbo de transporte

Fonte: elaboração própria.

Tabela 6 – Amostra de dados adaptados do *PrinParLat* para o lema verbal *facio*.

lemma	forma	descrição gramatical
facio	facio	1ª pessoa do singular do presente do indicativo na voz ativa
facio	facere	infinitivo presente ativo
facio	faciet	3ª pessoa do singular do futuro do indicativo na voz ativa
facio	feci, fecei	1ª pessoa do singular do pretérito perfeito do indicativo na voz ativa
facio	factum	particípio passado no nominativo singular do gênero neutro
facio	facturum	particípio futuro no nominativo singular do gênero neutro

Fonte: elaboração própria.

O *Lexicala*, por sua vez, é um dicionário latim-francês de acesso aberto modelizado como *Linked Data* e vinculado à LiLaKB que possui, entre os elementos de composição de seus verbetes, as formas canônicas com a indicação de quantidade silábica. Estas são extraídas da LiLaKB por meio de um terminal de acesso em linguagem SPARQL[7], utilizando-se um *script* elaborado pelo autor; desse modo, obtêm-se para o lema *habeo*, por exemplo, o seguinte conjunto de formas: "hăbĕō hăbēs hăbēre hăbŭī hăbĭtum".

Por fim, o mapeamento entre os identificadores do *LiLa Lemma Bank* e os identificadores da Princeton WordNet fornecidos pela *Latin WordNet Revision* possibilita a combinação de dados obtidos de diferentes versões dessa rede semântica, dentre as quais duas se destacam por oferecerem conteúdo em língua portuguesa: a OpenWordNet-PT (Paiva; Rademaker; Melo, 2012) e a Portuguese Unified Lexical Ontology (PULO) (Simões; Guinovart; Almeida, 2016). Com isso, pode-se elaborar uma planilha de dados contendo definições e equivalentes em português, obtidas pelo cruzamento de dados das duas *wordnets* mencionadas, acrescidos de informação sobre domínio semântico, obtida a partir da tradução das etiquetas da Princeton Wordnet.[8] A Tabela 7 apresenta uma amostra dessa planilha, com dados referentes ao lema *uideo*.

Essa planilha serve de base para a produção de um tipo de gráfico associativo hierárquico e dinâmico, denominado *collapsibleTree* (Khan, 2023), que permite a visualização das sugestões de equivalentes de tradução agrupados por domínio semântico. A Figura 20 traz uma amostra para o mesmo lema *uideo*.

4.2 DADOS DE DICIONÁRIOS TRADICIONAIS

A conversão de conteúdo textual produzido originalmente na forma impressa (i.e. para leitura humana) em dados estruturados (i.e. legíveis por máquina) é um processo fundamental para que a informação possa ser buscada e processada por algoritmos (Hanks, 2022, p.482–3). Ainda que o artigo lexicográfico seja um tipo de texto com grau de estruturação relativamente alto – no sentido de que é normalmente composto de um conjunto regular de categorias de informa-

[7] Cf. <https://lila-erc.eu/sparql/>.

[8] Dados disponíveis em <https://github.com/own-pt> e <http://wordnet.pt/download> e <https://wordnet.princeton.edu/download>, respectivamente.

Tabela 7 – Amostra da planilha com dados extraídos de uma combinação de WordNets para o lema *video*.

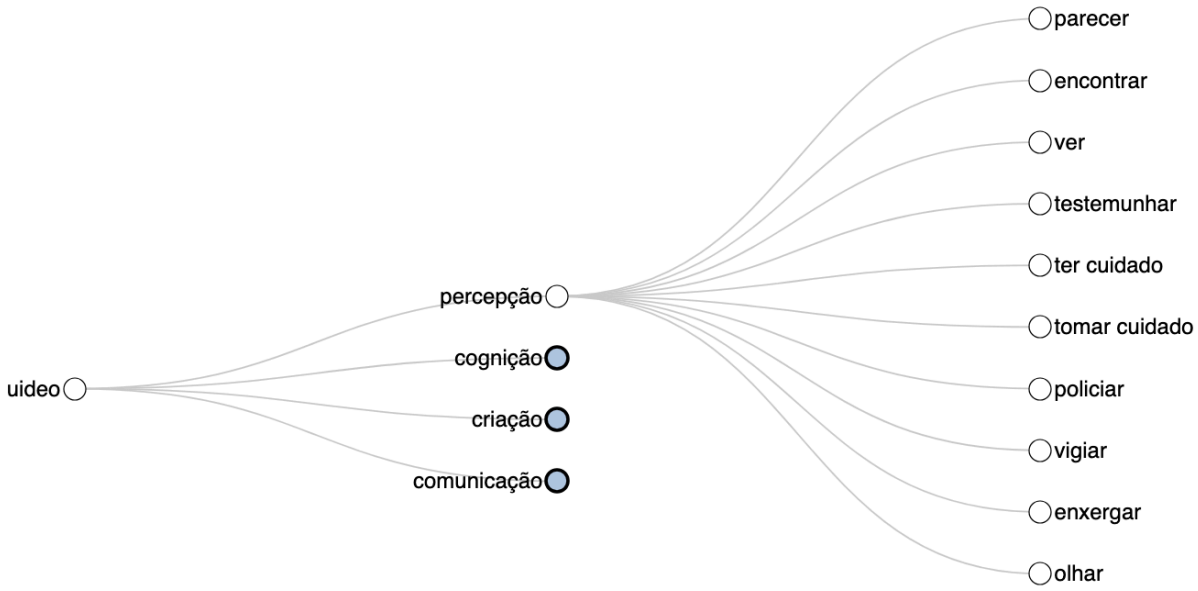
lemma	domínio	definição	equivalentes
uideo	cognição	perceber antecipadamente	antecipar, antever, prever
uideo	cognição	ter em mente ou transmitir, como convicção ou opinião	considerar, julgar, manter
uideo	cognição	conhecer ou tomar conhecimento de, geralmente acidentalmente	achar, descobrir, informar-se, inteirar, ouvir, saber, ver
uideo	cognição	fazer sentido; atribuir um significado à	compreender, entender, construir, interpretar, ver
uideo	cognição	observar ou considerar	considerar, reputar, estimar
uideo	comunicação	descobrir, aprender, ou determinar com certeza, geralmente, depois de fazer um inquérito ou outro esforço	aprender, averiguar, constatar, descobrir, ver, verificar, determinar
uideo	criação	imaginar; conceber; ver na imaginacao	afigurar, envisionar, figurar, ver, visualizar
uideo	percepção	parecer para sua própria mente ou opinião	parecer
uideo	percepção	perceber ou ser contemporâneo com	encontrar, ver, testemunhar
uideo	percepção	ser vigilante, estar atento ou cuidando	ter cuidado, tomar cuidado, policiar, vigiar
uideo	percepção	parecer ser verdade, provável ou aparente	parecer
uideo	percepção	perceber pela visão, ou têm o poder de perceber pela visão	enxergar, olhar, ver
uideo	percepção	dar uma certa impressão ou ter um certo aspecto exterior	parecer

Fonte: elaboração própria.

ção graficamente identificadas –, o panorama histórico apresentado no Capítulo 2 mostra que, quando se observa de uma perspectiva diacrônica a microestrutura dos dicionários que formam a tradição lexicográfica do latim, o que se tem é, sobretudo, diversidade. Isso representa um desafio para o processo de estruturação de dados, particularmente quando se pretende automatizar o processamento e, posteriormente, a recuperação dos dados de forma integrada e em larga escala, impondo assim a necessidade de se encontrar uma estrutura genérica o suficiente para acomodar em um esquema único a variedade de informações apresentadas em cada dicionário da tradição.

A estratégia utilizada para obter esse esquema consiste em um pré-processamento do texto em três etapas, que dizem respeito, respectivamente: (1) à segmentação do verbete em partes constituintes; (2) à representação da informação contida nos verbetes, segundo um modelo in-

Figura 20 – Gráfico dinâmico do tipo *collapsibleTree* mostrando equivalentes de tradução sugeridos pela WordNet para o lema *uideo*, agrupados por domínio semântico.



Fonte: elaboração própria.

terpretativo; (3) à identificação unívoca dos lemas de cada entrada lexical.

Antes de passar aos detalhes metodológicos de cada etapa, vale dizer que tais procedimentos foram aplicados ao conteúdo completo de quatro dicionários, selecionados com base em dois critérios: ter seu conteúdo em domínio público e estar digitalizado em formato texto. São eles: os dicionários de Cardoso (1570), Velez (1744), Fonseca (1798), Faria (1994); acrescenta-se ainda um conjunto de onze entradas extraidas de Saraiva (1927) relativas aos verbos latinos mais frequentes.[9]

Desse conjunto de fontes provém o material usado para a exemplificação dos processos descritos a seguir, cujo resultado é a criação de uma base de dados especializada em dicionários latim-português intitulada **latinolusitanicum-db**.[10]

4.2.1 Princípio de segmentação do verbete

Na lexicografia contemporânea, o padrão de organização da informação em um verbete parece bem estabelecido (Hartmann, 2001, p.65); o verbete pode ser dividido em duas partes: a primeira à esquerda no texto, com informações gerais sobre a forma do lema (e.g. ortografia, pronúncia, classe gramatical, padrão flexional e etimologia); e a segunda à direita, com informação sobre um ou mais significados, geralmente expressos por meio de definições e equivalentes,

[9] O texto base dos três primeiros recursos foi extraído da versão digital publicada pelo *Corpus Lexicográfico do Português* (Verdelho; Silvestre, 2002); o quarto foi gentilmente cedido pela professora Giovanna Longo (UNESP/Araraquara), a quem o autor agradece imensamente pela colaboração; o último é fruto de elaboração própria, obtido por digitalização manual.)

[10] Os arquivos-fonte e os *scripts* de processamento encontram-se em: <https://github.com/lucascdz/latinolusitanicum-db>.

eventualmente acompanhados de informações complementares como domínio semântico, sinônimos e antônimos, padrões sintáticos, colocações e exemplos, marcas de uso e, mais raramente, ilustrações (Welker, 2004). Todavia, essa estrutura convencional foi obra de séculos de desenvolvimento da lexicografia moderna a partir da herança medieval, como mostra Merrilees (1996) em sua descrição diacrônica da crescente complexidade de microestrutura observada nos grandes dicionários medievais de Latim.

Merrilees (1996) parte da estrutura mínima do verbete herdada dos glossários, composta por dois elementos: a palavra-entrada (lema) e a explicação semântica (definição); a essa estrutura se agregam, ocasionalmente, informações suplementares variadas, dispostas junto aos elementos principais. Por conveniência de análise, o autor considera toda informação que não é nem lemática nem definicional como metalinguagem e estabelece três posições relativas na microestrutura em que essa informação pode ocorrer: a posição pós-lemática, quando a informação suplementar ocorre entre o lema e a definição; a posição pós-definicional, quando a informação suplementar ocorre após a definição; e a posição marginal, quando a informação suplementar é colocada na margem do manuscrito. A partir desse esquema, observa que certos tipos de informação metalinguística vão sendo gradualmente associados a cada uma dessas posições na evolução da lexicografia do período medieval ao período moderno, com um predomínio de informações formais na posição pós-lemática e informações semânticas na posição pós-definicional. Porém, esses padrões estavam longe de se fixar no momento em que os dicionários começaram a ser impressos, de modo que a informação metalinguística permaneceu por um bom tempo sem posição determinada na microestrutura do verbete.[11]

A análise baseada na posição relativa de cada elemento (portanto, independente de conteúdo) fornece um princípio eficaz para se realizar a segmentação do conteúdo dos verbetes de diferentes dicionários. Esse processo de segmentação foi realizado de modo semi-automático, usando sequências de caracteres como delimitadores, especialmente aquelas utilizadas no texto como marcadores estruturais (e.g. marcas de pontuação, desinências formais, termos e expressões metalinguísticas). Na ausência de uma lista prévia das possibilidades de utilização de caracteres gráficos com tais funções pelas fontes, sua identificação ocorreu gradualmente durante o processo, mediante checagem manual a cada passo da segmentação. Ao final, obtém-se uma tabela em que cada linha corresponde a uma entrada lexical, com as informações agrupadas em colunas correspondentes às posições do modelo de Merrilees (excluída a posição marginal, própria de textos manuscritos). A Tabela 8 traz uma amostra do resultado, com os verbetes dos quatro dicionários processados relativos ao lema *iter* (‘caminho’).

Se, por um lado, o esquema é suficiente para representar verbetes simples como CARDOSO1570.ITER e VELEZ1744.ITER, por outro, deixa a desejar quando se trata de verbetes complexos como FARIA1994.ITER, em que a informação semântica se divide em múltiplas acepções, cada uma

[11] Essa característica é evidente em dicionários compilados durante o Renascimento, como mostra Merrilees (1988) quando analisa o *Dictionarius* de Firmin Le Ver (1440); o mesmo fazem Wooldridge (2002) quando analisa o *Dictionarium latinogallicum* de Robert Estienne (1538) e Stein (2014) quando analisa o *Dictionary of syr Thomas Eliot knyght* (1538).

Tabela 8 – Amostra de verbetes segmentados segundo a análise posicional de Merrilees (1996).

ID	LEMA	POS-LEMA	DEFINIÇÃO	PÓS-DEFINIÇÃO
CARDOSO1570.ITER	Iter, itineris		Ho caminho.	
VELEZ1744.ITER	Iter, itineris		o caminho.	
FONSECA1798.ITER	Iter, itineris	n. Cic.	Jornada, viagem. Caminho, estrada.	
FARIA1994.ITER	iter, itinĕris	subs. n.	I – Sent. próprio: 1) Percurso, caminho percorrido, marcha, viagem: *iter facere* (Cés. B. Gal. 1, 7. 3) "marchar". Daí: 2) Estrada, caminho, passagem (Cíc. At. 14, 10, 1). – II – Sent. figurado: 3) Via, meio, maneira (Cíc. De ar. 2. 234). 4) Curso: *iter amoris nostri* (Cíc. At. 4. 2. 1) "O curso de nossa afeição". 5) Direito de passagem (Cés. B. Gal. 1. 8, 3).	Obs.: Nom. arc. *itiner* (Plaut. Merc. 913); (Lucr. 6. 339). Gen. *iteris* (Ac. Tr. 627). Abl. *itere* (Lucr. 5. 653).

Fonte: elaboração própria.

contendo sua própria definição acompanhada de metalinguagem (e.g. marcadores de derivação semântica como "Sent. próprio", "Daí" e "Sent. figurado", além da numeração das acepções). Essa complexidade exige um modelo de representação de dados próprio para lexicografia, que permita estruturar os componentes convencionais de cada verbete (formas, acepções e exemplos lexicográficos) em planilhas distintas, no interior das quais o princípio de segmentação posicional pode ser aplicado com maior eficácia.

4.2.2 Representação da informação lexical

Segundo Andrade e Neves (2017), o processo informacional tem por objetivo o acesso e uso da informação por usuários de um centro de informações (isto é, qualquer local que ofereça informações úteis aos seus públicos). Das atividades e procedimentos técnicos que o constituem – quais sejam, descrição, análise, representação e classificação da informação –, a representação é aquela que se desenvolve com fins de recuperação da informação, pois "objetiva identificar de que trata o documento, retirar os principais conceitos e transformá-los em uma linguagem artificial passível de recuperação manual ou informatizada" (Andrade; Neves, 2017, p.103). As autoras destacam o caráter investigativo, e não somente objetivo ou descritivo, da tarefa, na medida em que envolve uma análise conceitual prévia, cujo resultado é traduzido, isto é, convertido

em determinado conjunto de termos de indexação. Para Maimone, Silveira e Tálamo (2011), a representação da informação é atividade propulsora de novos cenários intelectuais, uma vez que participa e promove a recuperação e assimilação, por parte dos usuários de arquivos, do conhecimento acumulado pelas sociedades humanas e registrado na forma de documentos sobre suportes diversos.

No campo das humanidades digitais, o potencial que a representação da informação tem de contribuir com o desenvolvimento científico depende fundamentalmente da existência de padrões comuns de criação, anotação e compartilhamento de recursos. Um conceito-chave nesse sentido é o de *interoperabilidade*, isto é, a "dimensão da capacidade de diferentes sistemas, organizações e/ou indivíduos para trabalhar em conjunto em busca de um objetivo comum" (Ide; Pustejovsky, 2010, s.p.).[12] Dois tipos de interoperabilidade se distinguem no âmbito dos sistemas computacionais: a sintática e a semântica. A interoperabilidade sintática se baseia em especificações de formatos de dados, protocolos de comunicação e coisas do gênero para garantir a intercomunicação e a troca de dados; os sistemas envolvidos podem processar a informação compartilhada, mas não há como garantir que a interpretação seja a mesma. Já a interoperabilidade semântica permite que dois sistemas sejam capazes de interpretar automaticamente a informação compartilhada de maneira significativa e precisa, mediante o uso de um modelo comum de referência para a troca de informações; nesse caso, o conteúdo das requisições de troca de informação são definidos de forma unívoca, isto é, a informação enviada corresponde à informação interpretada. Na origem desse tipo de modelo está a definição de um vocabulário comum para os termos de indexação.

Atualmente, dois modelos de representação de dados se destacam no âmbito da lexicografia: (1) as diretrizes estabelecidas Text Encoding Initiative (TEI) para a codificação de dicionários; e (2) o modelo OntoLex-lemon de representação da informação lexical em ontologias.[13]

As diretrizes da TEI são voltadas para a anotação *in-line* de texto digitalizado e tem servido especialmente como padrão para a retro-digitalização de dicionários impressos (Tiberius et al., 2021)[14]. Elas fornecem um modelo de anotação extremamente flexível, com diversas possibilidades de codificação para os mesmos elementos; entretanto, essa mesma característica pode significar um obstáculo para a interoperabilidade (Costa; Salgado; Almeida, 2021). Uma versão simplificada dessas diretrizes foi desenvolvida pelo projeto *TEI Lex-0* (Romary; Tasovac, 2018), em que são implementadas restrições à forma de descrever certos elementos e relações hierár-

[12] No original: "Broadly speaking, interoperability can be defined as a measure of the degree to which diverse systems, organizations, and/or individuals are able to work together to achieve a common goal."

[13] Um terceiro modelo tem sido desenvolvido pelo projeto ELEXIS visando a cooperação e a troca de informações entre a comunidade de pesquisa em Lexicografia. Tiberius et al. (2021) apresenta os resultados de uma pesquisa de campo com especialistas em lexicografia quanto aos elementos básicos e as definições desse vocabulário controlado de descritores lexicográficos.

[14] Segundo Kallas et al. (2019), "retro-digitalização é o processo de converter um dicionário impresso em um formato digital legível por computador, processo esse que envolve não apenas escaneamento e reconhecimento de caracteres, mas também codificação e enriquecimento dos dados". No original: "[retrodigitization is] the process of converting a dictionary published in paper into a digital, computer-readable format, which involves not only scanning and OCRing but also data encoding and enrichment".

quicas previstas pela TEI original a fim de promover interoperabilidade entre recursos lexicais heterogêneos; nesse caso, o caráter altamente restritivo do *TEI Lex-0* tem como consequência uma perda de flexibilidade que cria dificuldades para sua harmonização com o princípio geral de segmentação apresentado acima.

O *Lexicon Model for Ontologies* ou, simplesmente, OntoLex-lemon (Cimiano; McCrae; Buitelaar, 2016), por sua vez, é um modelo criado pelo grupo de trabalho Ontology Lexica (OntoLex) no âmbito do World Wide Web Consortium (W3C); tal modelo apresenta um grau de flexibilidade e de interoperabilidade semântica suficiente para se adaptar aos diferentes conjuntos de dados extraídos dos dicionários latim-português selecionados. Segundo Tiberius et al. (2021), esse modelo conceitual, originalmente desenvolvido para a representação de informação lexical em ontologias,[15] tornou-se o principal padrão para representação de informação lexical tanto em RDF quanto como *Linked Data*. O propósito do modelo é prover meios de associar informação linguística às ontologias, isto é, agregar informações sobre como os elementos do vocabulário de uma determinada ontologia (indivíduos, classes e propriedades) são lexicalizados em uma dada língua natural. Para tanto, seu vocabulário contempla os principais elementos necessários para construir um léxico computacional, além de permitir a combinação de elementos de outras ontologias; com isso, pode-se construir uma estrutura personalizada e adaptada aos dados a serem representados e, ao mesmo tempo, preservar um nível satisfatório de interoperabilidade.

O modelo OntoLex-lemon é formado por diversos módulos, que recobrem diferentes aspectos da informação lexical. O módulo básico, denominado *Ontology-lexicon interface module* (ou, simplesmente, *ontolex*), fornece instrumentos para representar a estrutura básica de uma entrada lexical. Para a representação da informação lexical extraída dos dicionários tradicionais de latim, foram selecionadas três classes de objetos: LexicalEntry, LexicalSense e Form.

A classe LexicalEntry representa uma unidade de análise do léxico, definida como "um conjunto de formas gramaticalmente relacionadas e um conjunto de significados associados a todas essas formas"; [16] vale dizer que o termo "formas gramaticalmente relacionadas" se refere àquelas que pertencem a uma mesma classe de palavras, seguem um mesmo paradigma flexional e possuem uma mesma etimologia. Os indivíduos da classe LexicalEntry são, portanto, entradas correspondentes a itens lexicais individuais, que podem ser ulteriormente especificados por meio de três subclasses: palavra (Word), afixo (Affix) ou expressão multipalavra (MultiwordExpression).

[15] Ontologia aqui se refere a um documentos escritos no padrão OWL (Web Ontology Language), uma linguagem criada no âmbito do consórcio W3C para representar conhecimento rico e complexo sobre coisas, grupos de coisas e relações entre coisas. É uma linguagem baseada na lógica computacional, que visa a exploração do conhecimento por computadores, sendo parte integrante das tecnologias da W3C voltadas para a Web Semântica, junto com RDF, RDFS, SPARQL, etc. (cf. <https://www.w3.org/2001/sw/wiki/OWL>. Entre as vantagens do modelo baseado em ontologias está o fato de favorecer a integração de dicionários científicos à infraestrutura da Web Semântica, permitindo que aplicações de PLN voltadas para mineração de textos (*text mining*) e processamento de dados possam usar esse conhecimento lexicográfico profundo (*in-depth*) (Steurs et al., 2020).

[16] No original: "A lexical entry represents a unit of analysis of the lexicon that consists of a set of forms that are grammatically related and a set of base meanings that are associated with all of these forms." (cf. <http://www.w3.org/ns/lemon/ontolex#LexicalEntry>).

A classe Form serve para representar cada uma das formas associadas a um certo item lexical.[17] Cada indivíduo da classe Form é, portanto, uma forma gramatical específica, cuja representação escrita é designada pelo valor da propriedade *writtenRepresentation*.[18]

A classe LexicalSense, por sua vez, representa individualmente o significado linguístico de um item lexical em referência a certo elemento de uma ontologia.[19] Em outras palavras, trata-se da forma lexicalizada que determinado conceito recebe em determinada língua, usualmente representada pela propriedade *definition*, emprestada à ontologia SKOS para prover uma definição em formato de texto.[20] A classe LexicalSense também pode abrigar eventuais propriedades adicionais, relativas a condições específicas sob as quais a associação do significado com o item lexical é considerada válida (e.g. contexto, registro, domínio, especificações formais como número, tempo verbal, etc.).

A associação entre os indivíduos das três classes selecionadas, por fim, é realizada através de propriedades específicas: a propriedade *lexicalForm* conecta uma entrada lexical (i.e. um indivíduo da classe LexicalEntry) às suas respectivas formas associadas (i.e. indivíduos da classe Form), ao passo que a propriedade *sense* conecta as entradas às suas respectivas acepções (i.e. indivíduos da classe LexicalSense). A Figura 21 ilustra a estrutura de relações do módulo *ontolex*.

A base formada por LexicalEntry, Form e LexicalSense é suficiente para modelar a estrutura básica de descrição do léxico, já que espelha, grosso modo, os tipos de informação que se encontram tradicionalmente nas obras lexicográficas, a saber (Hartmann, 2001, p.60–1): (1) uma palavra-entrada (lema), acompanhada de uma ou mais formas canônicas; (2) um comentário formal, que descreve o item lexical em termos de ortografia, pronúncia, morfologia e etimologia; e (3) um comentário semântico, com informações sobre o significado e o uso do item em contextos específicos. Todavia, outras estruturas frequentemente encontradas em textos lexicográficos – como sub-entradas representando homônimos ou palavras derivadas, acepções agrupadas hierarquicamente em diversos níveis ou exemplos de uso acompanhados de tradução em dicionários bilíngues – não são contempladas pelo módulo *ontolex*. Por isso, o modelo dispõe de um módulo específico para representar tais estruturas, denominado *OntoLex Lemon Lexicography Module* ou, simplesmente *lexicog* (Bosque-Gil; Gracia, 2019).

O módulo *lexicog* se organiza essencialmente sobre a classe LexicographicComponent, definida como "um elemento estrutural que representa as (sub-)estruturas de verbetes lexicográficos que fornecem informação sobre entradas, sub-entradas e acepções".[21] A classe Entry

[17] Na definição original: "A form represents one grammatical realization of a lexical entry." (cf. <http://www.w3.org/ns/lemon/ontolex#Form>)

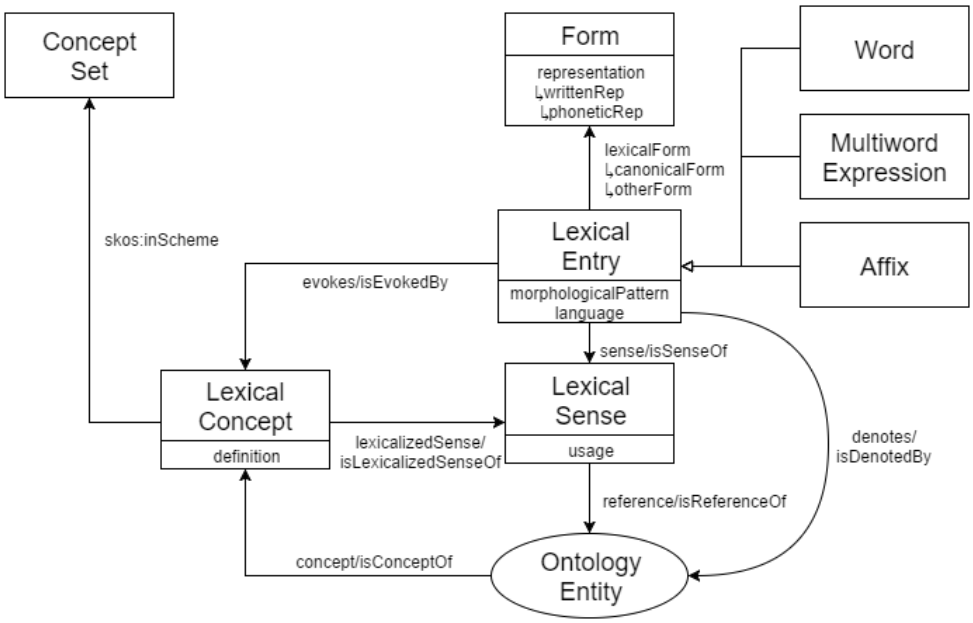
[18] Cf. <http://www.w3.org/ns/lemon/ontolex#writtenRep>.

[19] No original: "A lexical sense represents the lexical meaning of a lexical entry when interpreted as referring to the corresponding ontology element." (cf. <http://www.w3.org/ns/lemon/ontolex#LexicalSense>). Note-se que o modelo prevê a associação do léxico com uma ontologia de conceitos, através da classe LexicalConcept, tarefa que não foi contemplada neste trabalho.

[20] Cf. <http://www.w3.org/2004/02/skos/core#definition>.

[21] No original: "A lexicographic component is a structural element that represents the (sub-)structures of lexico-

Figura 21 – Diagrama de relações entre elementos do módulo *ontolex*.



Fonte: Cimiano, McCrae e Buitelaar (2016).

constitui um tipo especial de componente estrutural, usado para representar a principal estrutura de acesso do dicionário, isto é, a palavra-entrada, ponto inicial do verbete. Em todo caso, o fato de serem elementos estruturais significa que os indivíduos dessas classes não portam informação linguística, mas apenas serve para representar o modo de estruturação de entradas e sub-entradas específico de cada dicionário; a vinculação entre os componentes estruturais e a descrição lexical representada pelos elementos do módulo *ontolex* é feita por meio da propriedade *describes*. Além disso, os indivíduos da classe LexicographicComponent podem ser hierarquizados através da propriedade *subComponent*, permitindo representar as relações de entrada e sub-entrada ou subdivisões no âmbito das acepções. Por fim, o módulo *lexicog* também inclui a classe UsageExample,[22] que fornece meios de representar exemplos de uso constituídos por mais de uma instância textual; na prática, isso significa que um mesmo exemplo pode ser representado em diferentes "versões", como é o caso dos exemplos acompanhados de tradução típicos de dicionários bilíngues. Indivíduos da classe UsageExample se vinculam diretamente às acepções (representadas pela classe LexicalSense do módulo *ontolex*) através da propriedade *usageExample*.[23]

A aplicação combinada dos módulos *ontolex* e *lexicog* aos dados pré-estruturados, resultantes da segmentação dos verbetes apresentada anteriormente, permite produzir um conjunto de quatro tabelas relacionais, cada uma com o conteúdo pertinente a uma classe do modelo: do módulo *ontolex*, as classes LexicalEntry, Form e LexicalSense; do módulo *lexicog*, a classe

graphic articles providing information about entries, senses or sub-entries." (cf. <http://www.w3.org/ns/lemon/lexicog#LexicographicComponent>

22 Cf. <http://www.w3.org/ns/lemon/lexicog#UsageExample>.

23 Cf. <http://www.w3.org/ns/lemon/lexicog#usageExample>.

UsageExample.[24] Em cada tabela, as linhas correspondem aos indivíduos da respectiva classe e as colunas, às propriedades que lhe são atribuídas. Entre elas, duas propriedades são emprestadas de outros sistemas de organização do conhecimento[25] – *lexinfo:note* e *skos:note*[26] –, sendo utilizadas para representar a informação metalinguística suplementar de forma genérica. Assim, se na planilha das entradas lexicais esse tipo de informação se identifica com as posições pós-lemática e pós-definicional do esquema de segmentação inicial, nas demais planilhas elege-se uma propriedade central em relação à qual a metalinguagem se localiza antes ou depois: no caso das Formas, essa propriedade é a grafia; no caso das Acepções, a definição; e no caso dos Exemplos, a tradução.

A Tabela 9 traz uma amostra conjunta das quatro planilhas em que se distribui o conteúdo do verbete para o lema nominal *iter* em Faria (1994). Observe-se que, na primeira delas, referente à entrada lexical, as formas do lema e as diversas acepções são substituídas por identificadores que servem de chaves de relação para o conteúdo individualizado nas planilhas específicas. Assim, as duas formas canônicas que constituem o lema (‘iter, itinĕris’), por exemplo, são representadas como indivíduos da classe Form, a que se atribuem identificadores (ITER-FORM1 e ITER-FORM2, respectivamente) e se descrevem suas grafias (*writtenRep*); os mesmos identificadores são utilizados, na planilha das Entradas lexicais, como valores da propriedade *lexicalForm*, servindo como chaves de relação entre a entrada lexical e suas formas canônicas. O mesmo vale para as cinco acepções da referida entrada (ITER-SENSE1, ITER-SENSE2, etc.) e para os dois exemplos que acompanham duas dessas acepções (ITER-EXEMP1, ITER-EXEMP2).

Esse modo de estruturar e representar dados lexicográficos, baseado no modelo OntoLex e no princípio de segmentação posicional, alcança um nível de generalização suficiente para ser virtualmente aplicável a qualquer dicionário da tradição. Para os fins deste trabalho, ele foi aplicado ao conteúdo dos cinco dicionários latim-português selecionados como fontes de dados. A Tabela 10 traz resultados quantitativos da estruturação, em termos de indivíduos por classe OntoLex em cada dicionário.

Ao mesmo tempo, o método constitui um passo importante para a publicação dos dados em formato *Linked Open Data* (LOD), aumentando o alcance e o potencial de reuso desses dados. O segundo passo é vincular esses dados com uma base de conhecimento externa, como a LiLa

[24] As classes Entry e LexicographicComponent, embora sejam importantes para a representação completa do dicionário impresso, não influem na estruturação dos dados indexados por lema, requerida pelo aplicativo DWS utilizado (*lexicographR*), não sendo, portanto, contempladas nesta exposição.

[25] Sistema de organização do conhecimento (*Knowledge Organization System*, abreviado KOS), também conhecido como sistema conceitual ou esquema conceitual, é um termo genérico usado para designar um conjunto seletivo de conceitos acompanhados de relações semânticas selecionadas, que têm entre suas aplicações o enriquecimento semântico de recursos computacionais, a fim de facilitar o gerenciamento e a recuperação do conhecimento e da informação (Hjørland, 2016). Entre os padrões mais conhecidos estão o *Resource Description Framework* (RDF), o *Resource Description Framework Schema* (RDFS), o *Simple Knowledge Organization System* (SKOS), o Dublin Core Metadata Initiative (DCMI), o LexInfo (voltado para anotação linguística de recursos lexicais), entre outros. O padrão SKOS, em particular, tem sua utilidade em lexicografia analisada por Costa, Salgado e Almeida (2021).

[26] Cf. <http://www.lexinfo.net/ontology/3.0/lexinfo#note> e <http://www.w3.org/2004/02/skos/core#note>, respectivamente.

Tabela 9 – Amostra de verbete segmentado em tabelas relacionais.

Entradas lexicais				
LexicalEntry	**lexicalForm**	**lexinfo:note**	**sense**	**skos:note**
FARIA1994.ITER	ITER-FORM1 , ITER-FORM2	subs. n.	ITER-SENSE1 , ITER-SENSE2 , ITER-SENSE3 , ITER-SENSE4 , ITER-SENSE5	Obs.: Nom. arc. *itiner* (Plaut. Merc. 913); (Lucr. 6. 339). Gen. *iteris* (Ac. Tr. 627). Abl. *itere* (Lucr. 5. 653).

Formas			
Form	**lexinfo:note**	**writtenRep**	**skos:note**
ITER-FORM1		iter	
ITER-FORM2		itinĕris	

Acepções				
LexicalSense	**lexinfo:note**	**skos:definition**	**skos:note**	**usageExample**
ITER-SENSE1	Sent. próprio	Percurso, caminho percorrido, marcha, viagem		ITER-EXEMP1
ITER-SENSE2	Daí	Estrada, caminho, passagem	Cíc.At.14.10.1	
ITER-SENSE3	Sent. figurado	Via, meio, maneira	Cíc.De ar.2.234	
ITER-SENSE4		Curso		ITER-EXEMP2
ITER-SENSE5		Direito de passagem	Cés.B.Gal.1.8.3	

Exemplos				
UsageExample	**rdfs:value@la**	**lexinfo:note**	**rdfs:value@pt**	**skos:note**
ITER-EXEMP1	iter facere	Cés.B.Gal.1.7.3	marchar	
ITER-EXEMP2	iter amoris nostri	Cíc.At.4.2.1	O curso de nossa afeição	

Fonte: elaboração própria.

Tabela 10 – Número de indivíduos por classe do modelo OntoLex em cada dicionário.

dicionário	**LexicalEntries**	**Forms**	**LexicalSenses**	**UsageExamples**
Cardoso (1570)	27.751	53.925	26.843	76
Velez (1744)	4.829	9.339	5.060	2.420
Fonseca (1798)	32.951	89.115	38.186	3.626
Saraiva (1927)	12	58	266	1.235
Faria (1994)	38.209	–	47.137	–

NB: As contagens acima são precisas e resultam da estruturação completa dos dicionários, com exceção de Saraiva (1927). Fonte: elaboração própria.

Knowledge Base. Isso passa pela identificação dos lemas de acordo com o catálogo do LiLa Lemma Bank, que provê ao mesmo tempo o indexador comum de lemas para o protótipo de dicionário construído neste trabalho.

4.2.3 Identificação dos lemas

A perspectiva de acesso integrado ao conteúdo estruturado dos diversos dicionários latim-português requer uma estratégia de indexação homogênea e unívoca de cada lema. À primeira vista, bastaria parear as formas canônicas principais de cada dicionário para vincular os verbetes correspondentes. Porém, o uso dessa estratégia simples encontraria problemas de eficiência, seja pela existência de lemas homógrafos (e.g. DICO ‘dizer’ e DICO ‘dedicar’), seja pelas variações ortográficas produzidas pela história do uso e descrição do latim (e.g. *caelum*/*coelum*, ‘céu’).

Uma solução mais consistente é a utilização de identificadores unívocos, como aqueles estabelecidos pelo LiLa Lemma Bank, de forma a assegurar a coerência entre as diferentes fontes de dados no que tange à descrição lexical. Para tanto, é preciso mapear as formas canônicas de entrada dos dicionários com as formas correspondentes no referido banco de lemas.

O pareamento entre os lemas é feito através de um processo de *string matching*. A fim de reduzir o número de resultados ambíguos ou falso-positivos, os lemas dos dicionários são normalizados e etiquetados com categorias gramaticais. A normalização consiste em: conversão para minúsculas, substituição de caracteres (*j* por *i*, *v* por *u*), supressão de diacríticos e expansão de abreviaturas e ligaduras (e.g. *cõpositè* normalizada como *composite*, *Pompeij* normalizada como *pompeii*). A categorização gramatical é deduzida do conjunto de formas canônicas que tradicionalmente constitui o lema dos dicionários de latim (e.g. ‘Grammatica,ae’ é convertido em ‘grammatica.NOUN.n1’, isto é, substantivo da primeira declinação; ‘Do, das’ é convertido em ‘do.VERB.v1r’, isto é, verbo regular da primeira conjugação).

O processo de pareamento é feito de modo semi-automatizado, através de um programa escrito em linguagem R que utiliza as bibliotecas *tidyverse* (Wickham et al., 2019) e *stringdist* (Van der Loo, 2014) e adota como estratégia uma abordagem progressiva em três etapas:

1. primeiro, a sequência completa formada pelo lema e as etiquetas gramaticais é comparada

Tabela 11 – Resultados quantitativos do pareamento de lemas entre os dicionários e o LiLa Lemma Bank.

	Cardoso (1570)	Velez (1744)	Fonseca (1798)	Faria (1994)
lemas totais	27.752	4.723	32.723	37.904
pareamento único (1:1)	18.467	4.093	28.546	33.600
pareamento ambíguo (1:N)	893	368	1.096	1.912
sem pareamento (1:0)	8.392	262	3.081	2.392
1:0 com um candidato	789	151	1.903	823
1:0 com muitos candidatos	1.153	104	1.088	1.185
1:0 sem candidatos	6.450	7	90	384
novos lemas	1.467	140	1.886	909
novas grafias	656	60	706	679
erros tipográficos	153	30	489	435

Fonte: elaboração própria.

com a base de dados do projeto LiLa; os resultados são classificados em pares únicos (1:1), pares ambíguos (1:N) ou sem pareamento (1:0);

2. em seguida, uma segunda rodada de comparação é realizada sobre os não pareados (1:N) na etapa anterior, desta vez utilizando, como sequência de caracteres, apenas o lema; os resultados obtidos são classificados da mesma forma que no passo anterior;

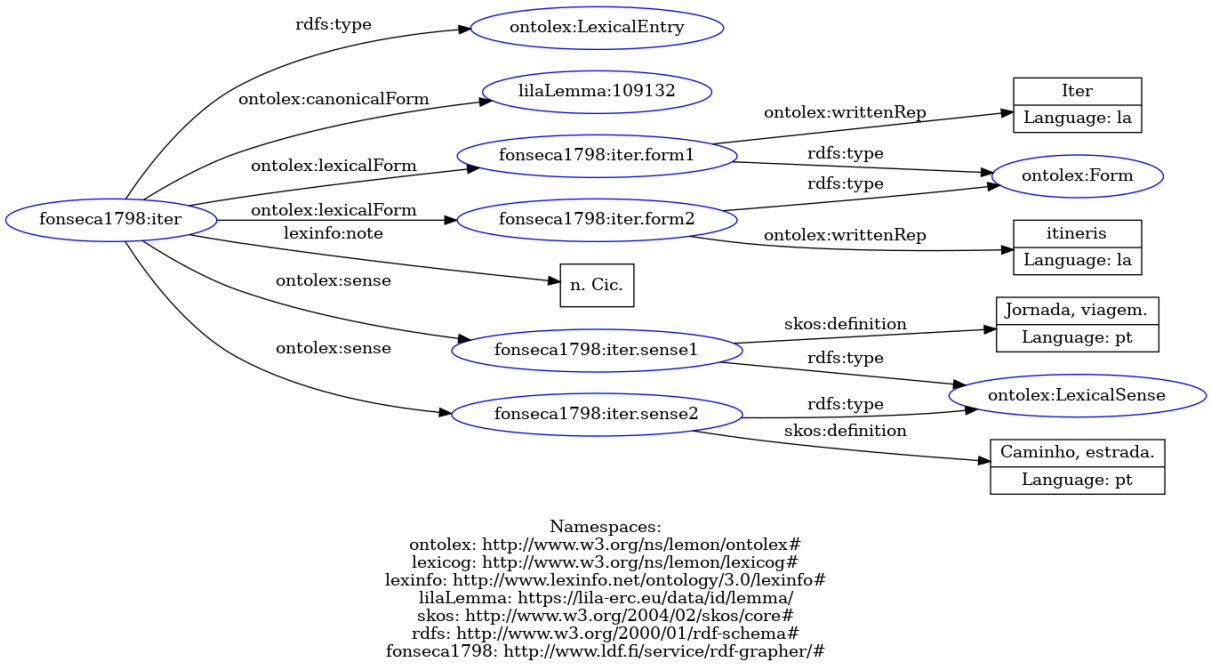
3. por fim, os lemas não pareados (1:N) remanescentes têm sua distância (*edit distance*) calculada em relação aos itens da base LiLa, o que permite fazer um levantamento de possíveis candidatos para o pareamento.

O trabalho de desambiguação dos pares ambíguos (1:N) e dos candidatos ao pareamento (1:0) é feito manualmente. Embora demande um tempo considerável, esse tipo de verificação dos resultados 1:0 gerou benefícios para as duas bases de dados envolvidas: de um lado, a informação obtida do LiLa Lemma Bank permitiu eliminar erros tipográficos e identificar subentradas não detectadas no processo de estruturação dos dicionários; de outro, novas entradas foram incluídas no Lila Lemma Bank com base na informação fornecida pelos dicionários. Um panorama quantitativo dos resultados do processo de pareamento de quatro dicionários – quais sejam, Cardoso (1570), Velez (1744), Fonseca (1798), Faria (1994) – podem ser vistos na Tabela 11.

O identificador de cada lema é acrescentado à base de dados como um valor de uma propriedade *canonicalForm*, a qual se associa às entradas lexicais no modelo OntoLex. Além de garantir a coesão interna entre os dados obtidos das diferentes fontes que compõem a base de dados do **eLatim**, esse procedimento de identificação também torna possível a publicação dos dados estruturados dos dicionários de forma vinculada à estrutura da LiLaKB, segundo o padrão LOD.[27] A Figura 22 traz uma visualização gráfica das relações entre os elementos de cada tabela

[27] A publicação como LOD passa pela codificação dos dados em formato Turtle RDF (Terse RDF Triple Language,

Figura 22 – Visualização gráfica da representação LOD do verbete *iter* em Fonseca (1798).



Legenda: Balões ovais indicam indivíduos, setas legendadas indicam relações/propriedades e caixas retangulares indicam valores textuais. Fonte: Elaboração própria por meio do aplicativo *RDF Grapher* (<https://www.ldf.fi/service/rdf-grapher>).

estabelecidas pelas diferentes propriedades, utilizando como amostra a entrada para o lema *iter* em Fonseca (1798)).

4.2.4 Formatação para publicação

Se a estrutura de tabelas relacionais é um formato adequado para o compartilhamento e processamento de dados, para visualização humana e compreensão da informação é certamente pouco amigável. Não obstante, na medida em que os dados estejam estruturados, i.e. separados e classificados de forma homogênea, é possível criar um roteiro de transformação para recompilar os dados distribuídos na planilha em um texto formatado para o consumo humano da informação. No caso do *lexicographR*, utiliza-se a linguagem HTML,[28] que é linguagem de marcação padrão utilizada para criar páginas Web.

Um *script* desenvolvido em linguagem R é usado para dispor os dados segundo as convenções usuais da lexicografia, porém de modo mais arejado, sem o adensamento típico do verbete impresso. O plano do verbete prevê três seções, que dizem respeito a: (1) formas canônicas e informações adicionais sobre o lema; (2) acepções e informações adicionais sobre o significado; (3) expressões multipalavra e exemplos lexicográficos. Na primeira, as formas canônicas são extraídas da planilha 'Formas', reunidas em uma única linha e destacadas em negrito para compor a cabeça do verbete; a informação metalinguística sobre o lema, extraída da planilha 'Entradas

<https://www.w3.org/TR/turtle/>). Até o momento, três dicionários tiveram seus dados completos publicados como LOD na plataforma LiLa, a saber, Cardoso (1570), Velez (1744), Fonseca (1798); o quarto dicionário (Faria, 1994) está em vias de publicação. Cf. <https://lila-erc.eu/data-page/>.

[28] Sigla para Hyper Text Markup Language, cf. <http://www.w3schools.com/html/>

Tabela 12 – Amostra de verbetes compilados e codificados em HTML para utilização no aplicativo *lexicographR*.

lemma	fonte	verbete compilado (HTML)
item	Faria1994	item <small>adv.</small> <small></small> <p> <b
iter	Faria1994	iter itinĕris subs. n. <small
iterum	Faria1994	itĕrum <small>adv.</small> <small></small> <p> <
iter	Faria1994	itinĕris <small>gen. de iter.</small> <small></small
eo	Faria1994	itum supino de eo.
eo	Faria1994	ītur <small>pass. impess. de eo:</small> <small></
iaceo	Faria1994	jacĕō -ēs -ēre jacŭī. iacĕō
iacio	Faria1994	jacĭō -is -ĕre jēcī. jáctum iac
iaceo	Faria1994	jacŭī <small>perf. de jacĕo.</small> <small></small

Fonte: elaboração própria.

Figura 23 – Amostra de verbete no modo de exibição, a partir da formatação HTML.

iter itinĕris
subs. n.
Obs.: Nom. arc. *itiner* Plaut. Merc. 913; Lucr. 6.339. Gen. *iteris* Ac. Tr. 627.
Abl. *itere* Lucr. 5.653.

1 Sent. próprio Percurso, caminho percorrido, marcha, viagem
2 Daí Estrada, caminho, passagem Cíc. At. 14.10.1
3 Sent. figurado Via, meio, maneira Cíc. De ar. 2.234
4 Sent. figurado Curso
5 Sent. figurado Direito de passagem Cés. B. Gal. 1.8.3

1 **iter facere** "marchar" Cés. B. Gal. 1.7.3
4 **iter amoris nostri** "O curso de nossa afeição" Cíc. At. 4.2.1

Fonte: elaboração própria.

lexicais', é posicionada logo abaixo, em letra menor. Na segunda, as acepções são compostas a partir da planilha correspondente, cada uma ocupando uma linha de texto, com o conteúdo da definição formatado em letra com tamanho regular ladeado pelos comentários adicionais em letra menor, respeitando a posição original; o mesmo não vale para a numeração que acompanha as acepções, acrescentada e padronizada como uma sequência de algarismos arábicos, independente do sistema usado originalmente no dicionário impresso. Na terceira seção, por fim, os exemplos são agrupados conforme a numeração das acepções a que correspondem, com texto latino destacado em negrito, tradução em português em letra normal e comentários adicionais em letra menor. A Tabela 12 traz uma amostra da planilha de dados codificados em HTML para utilização pelo aplicativo *lexicographR*; a forma como o conteúdo é exibido ao usuário é vista na Figura 23.

4.3 DADOS EXTRAÍDOS DE *CORPUS*

É consenso dizer que a linguística de *corpus* revolucionou a prática lexicográfica. Granger (2012) coloca o uso de *corpora* – i.e. "coleções eletrônicas de dados linguísticos autênticos" (Granger, 2012, p.3) – como a primeira de seis inovações resultantes do uso das tecnologias computacionais em lexicografia, ao lado da maior qualidade e quantidade de dados disponíveis, da maior variedade e eficiência das estruturas de acesso e das possibilidades de personalização, integração com outras ferramentas e colaboração do usuário, as quais têm permitido que os dicionários eletrônicos sejam mais do que a forma digitalizada dos produtos da tradição impressa. A autora destaca que, além do uso como matéria-prima para lexicógrafos produzirem verbetes enriquecidos, os próprios *corpora* podem ser disponibilizados como parte integrante dos dicionários eletrônicos, de modo que os usuários tenham a oportunidade de acessá-los e explorá-los por si mesmos. Para Hanks (2012a), a lexicografia eletrônica enseja novas possibilidades que anteriormente eram vedadas à lexicografia tradicional, como novos tipos de evidência, novas formas de descrição e novas maneiras de organizar as evidências, ligadas às possibilidades de explorar bases de dados estruturadas e os vínculos entre hipertextos. Nesse contexto, o principal impacto do uso de *corpus* para o lexicógrafo é "a capacidade de obter evidências suficientes para fazer generalizações com um nível razoável de segurança" (Hanks, 2012a, p.63), tanto para uma melhor descrição gramatical (i.e. das valências e construções em que as palavras são normalmente usadas) quanto para identificar e medir colocações lexicais (i.e. as preferências que as palavras têm de se apresentar na companhia de determinadas outras palavras). No âmbito da lexicografia bilíngue, em especial, o autor ressalta a vantagem do uso de *corpus* na identificação de padrões fraseológicos de equivalência entre línguas (quer dizer, frases que não podem ser traduzidas 'palavra-por-palavra'), provendo um meio de produzir conhecimento contrastivo de modo sistemático baseado em um grande número de evidências e não mais em leitura seletiva e introspecção.

Três aspectos condicionam o uso de *corpus* em lexicografia (Heid, 2008): (1) a seleção do material linguístico que será usado como fonte de evidência (mediante critérios quantitativos e qualitativos); (2) a seleção dos tipos de dados linguísticos que serão extraídos para inclusão no dicionário; e (3) a ferramenta utilizada para a manipulação de dados de *corpus* no trabalho lexicográfico. Esses aspectos são tratados nos tópicos a seguir, começando pelas questões envolvidas na compilação de *corpora* em geral e de *corpora* paralelos em especial, passando pela descrição do processo e dos resultados da compilação de um *corpus* paralelo de textos latim-português, e finalizando com os tipos de dados extraídos desse *corpus* segundo as possibilidades oferecidas pela ferramenta *lexicographR*.

4.3.1 Fundamentos para a compilação de *corpora*

A compilação de um *corpus* pressupõe uma fase de planejamento que envolve especificações relacionadas a duas áreas (Atkins; Clear; Ostler, 1992). Do ponto de vista linguístico, é

preciso definir no mínimo o tipo de *corpus* que se quer construir e, se possível, os tipos de texto, as variedades de linguagem e o tamanho das amostras a serem incluídas. Do ponto de vista administrativo, é preciso definir as estratégias de amostragem utilizadas (de base estatística), bem como verificar direitos de uso do material selecionado, as formas de captura e armazenamento dos dados e, no caso da atividade lexicográfica, o modo de vincular o *corpus* à base de dados lexicais. Esse conjunto de variáveis faz do planejamento e compilação de um *corpus* uma atividade complexa, quando não "propensa ao paradoxo, onde até mesmo uma decisão aparentemente simples pode ter várias ramificações" (Hunston, 2008, p.154).

Em se tratando de uso de *corpus* em lexicografia, a escolha do tipo de *corpus* a ser utilizado deve estar em conformidade com o plano do dicionário (Heid, 2008). No caso de dicionários bilíngues, como o **eLatim**, a literatura especializada recomenda o uso de um *corpus* paralelo, por ser "insuperável quando se trata de descobrir correspondências de tradução" (Aijmer, 2008). Um *corpus* paralelo consiste em um conjunto de textos-fonte acompanhados de tradução em outras línguas (Teubert, 2001). Quando conta com alinhamento entre texto e tradução (isto é, a associação explícita entre unidades de tradução e seus equivalentes) ao menos no nível do período (*sentence*), um *corpus* paralelo pode fornecer ao lexicógrafo dados sintagmáticos preciosos sobre o uso das palavras em ambas as línguas (Heid, 2008). Caso esteja disponível para consulta pelo usuário, um *corpus* paralelo pode funcionar como um dicionário bilíngue, em que se observam diretamente as correspondências entre itens nas duas línguas, com a vantagem de prover um inventário mais amplo de significados e, sobretudo, mais informações contextuais (Aijmer, 2008; Teubert, 2001).

Quanto à seleção de textos, há uma série questões a se considerar, essencialmente relacionadas com a necessidade de conferir validade e confiabilidade à pesquisa feita com base no *corpus* (Kennedy, 1998). Se o objetivo de se constituir um *corpus* é possibilitar generalizações a respeito de uma determinada língua ou variante, fatores como autenticidade, representatividade e balanceamento são tidos como requisitos básicos, assim como o tamanho, isto é, a quantidade de textos necessária para contemplar tais requisitos.

Autenticidade diz respeito ao fato de que o material linguístico que compõe o *corpus* foi produzido em situações naturais de comunicação para fins de comunicação e expressão, e não com a finalidade de ser objeto de análise linguística ou mesmo com a consciência de que poderia ser usado para tal fim (Aluísio; Almeida, 2006; Stefanowitsch, 2020).

Representatividade é a relação entre o *corpus* e a língua que ele pretende representar (Hunston, 2008). Em um artigo fundamental sobre a questão, Biber (1993) define representatividade em termos gerais, como a medida em que uma amostra encerra o espectro de variabilidade em uma população, de modo que possa ser utilizada como base para se fazer generalizações acerca dessa população. Para o autor, no caso de *corpora* de textos, a variabilidade pode ser considerada de duas perspectivas: uma situacional, relacionada com os tipos de texto (também chamados de gêneros ou registros); outra linguística, relacionada com a distribuição de fatos de língua no interior (e através) dos textos. Embora os gêneros (ou registros) não sejam definidos com base em

características linguísticas,[29] haveria, segundo ele, uma certa estabilidade quanto às propriedades linguísticas características de um mesmo gênero, bem como, consequentemente, diferenças linguísticas significativas entre os gêneros. Assim, uma condição da representatividade linguística é o *corpus* conter toda a variedade de gêneros textuais em que uma língua se apresenta. Além disso, há a questão da proporção em que cada gênero deve ser representado; uma abordagem simplesmente estatística da questão seria propor que as amostras fossem incluídas no *corpus* na mesma proporção em que ocorrem na realidade; porém, a decisão de incluir ou excluir textos em um *corpus* para pesquisa linguística também depende da importância relativa de determinados gêneros e tipos de textos, relacionada a fatores como o nível de influência cultural (acadêmica e/ou literária), o alcance e interesse do público-leitor e o nível de tipicidade do texto em relação ao gênero (Kennedy, 1998).

O balanceamento, por sua vez, refere-se ao equilíbrio interno do *corpus*, isto é, à proporção de textos entre os diferentes subconjuntos que formam o *corpus* (Hunston, 2008). Segundo a autora, a principal vantagem de um *corpus* balanceado é que cada subconjunto (subcorpus) é grande o suficiente para permitir que se façam comparações quanto à presença ou ausência de certa característica sem que o resultado seja devido à maior ou menor representatividade entre os subconjuntos.[30] Todavia, é possível que o ideal do balanceamento (equilíbrio entre subcorpora) pode entrar em conflito com o da representatividade (cada subcorpus proporcional à sua importância). Para Atkins, Clear e Ostler (1992), esse controle só pode ser realizado após a compilação do *corpus* e sua utilização, de modo que seria excessivamente limitante esperar que um *corpus* seja cientificamente balanceado para que se possa fazer uso dele.

Quanto ao tamanho do *corpus*, por fim, não há uma resposta precisa para a quantidade de palavras ou textos necessária para garantir uma descrição adequada (Stefanowitsch, 2020). Para esse autor, idealmente o *corpus* deveria ser grande o suficiente para conter amostras igualmente numerosas de cada estrutura gramatical ou item lexical de uma língua (isto é, a máxima representatividade no sentido da distribuição linguística). Sendo este um objetivo difícil de controlar, pode-se dizer que quanto mais dados há a disposição, mais é possível apreender, especialmente em se tratando de itens com baixa frequência (Tiberius et al., 2022). Além do tamanho total do *corpus*, Kennedy (1998) discute questões associadas como a quantidade de amostras por categoria de texto (i.e., os gêneros que definem os subcorpora) e, por conseguinte, a quantidade de texto por amostra. Diante da possibilidade de se utilizar documentos completos ou amostras de documentos, o autor observa que ambas podem gerar distorções na representatividade do *corpus*: no primeiro caso, o uso de documentos completos tem como consequência uma menor variedade de textos e autores; no segundo, o uso de amostras pode gerar omissões de certas pro-

[29] Na verdade, pode-se dizer com McEnery e Wilson (2001, p.79) que toda estratificação é "um ato de interpretação por parte do compilador, uma vez que se baseia em modos particulares de dividir a linguagem em entidades (e.g. gêneros) que não são naturalmente inerentes a ela".

[30] Em caso de subcorpora de dimensões diferentes, um efeito similar pode ser obtido pela normalização da frequência de um item, isto é, o cálculo da frequência relativa do item por uma mesma quantidade total de palavras (Hunston, 2008; Tiberius et al., 2022)

priedades linguísticas, especialmente itens de baixa frequência. Em todo caso, o autor estima que um *corpus* com tamanho entre meio milhão e um milhão de palavras seja suficiente para uma análise confiável de muitos processos sintáticos e itens de vocabulário mais frequentes.

Em suma, as dificuldades teóricas inerentes ao projeto de compilação de um *corpus* podem se referir seja à definição de um universo de textos que deve ser representado pelas amostras, seja à definição de categorias de textos que devem orientar o balanceamento da amostragem, seja à própria quantidade de textos necessária para se ter um *corpus* representativo e/ou balanceado.[31] Tais dificuldades são acompanhadas por outras de ordem prática, como limitações do software utilizado para processar os textos, restrições relativas a direitos de uso e propriedade e, principalmente, a disponibilidade de textos em formatos adequados à sua utilização imediata na compilação de um *corpus* eletrônico (Hunston, 2008). No caso de um *corpus* paralelo, o trabalho é reconhecidamente mais complexo (Aijmer, 2008; Heid, 2008), uma vez que tais dificuldades incidem tanto sobre os textos da língua-fonte quanto nas respectivas traduções na língua-alvo e incluem, adicionalmente, a necessidade de alinhamento entre os textos correspondentes. Por isso, como afirma Hunston (2008, p.156–7), muitas vezes "o *corpus* vai depender mais do que está disponível gratuitamente e em formato facilmente conversível do que de qualquer outro critério", de modo que "todo *corpus* resulta de uma conciliação entre o que é desejável (i.e. o que o pesquisador planejou) e o que é possível".[32]

Diante da complexidade das questões apresentadas, a saída, segundo Hunston (2008, p.162–4) é "ser transparente em relação a como o *corpus* foi planejado e o que ele contém, para permitir ao usuário avaliar o grau de representatividade [e balanceamento desse conteúdo] […] e ajustar as hipóteses de acordo".[33] Com essa recomendação em mente, descrevem-se a seguir as decisões e os resultados da compilação do *corpus* paralelo latim-português utilizado neste trabalho.

4.3.2 A compilação de um *corpus* paralelo latim-português

Assim como no caso dos dicionários tradicionais, o principal critério para a inclusão de textos no *corpus* paralelo aqui descrito foi a disponibilidade de textos em formato eletrônico compatível com o projeto do **eLatim**. Isso implica, de um lado, em textos latinos lematizados, de preferência alinhados com o Lila Lemma Bank (para sua imediata compatibilidade com a estrutura de indexação) e em formato CoNLL-U[34] (para sua imediata compatibilidade com a

[31] Esse cenário faz com que alguns autores considerem a compilação de um *corpus* verdadeiramente representativo como uma tarefa praticamente (e talvez mesmo teoricamente) impossível – cf., por exemplo, Heid (2008, p.134) e Stefanowitsch (2020, p.35).

[32] No original: "Design of the corpus will depend more on what is freely available in an easily-converted format than on other criteria. […] All corpora are a compromise between what is desirable, that is, what the corpus designer has planned, and what is possible."

[33] No original: "[Among the] possible responses [is] to acknowledge the problems but to do the best that is possible in the circumstances and to be transparent about how the corpus has been designed and what is in it. This allows the degree of representativeness to be assessed by the corpus user. […] an imbalance in a corpus does not matter so long as it is known and hypotheses can be adjusted accordingly." Cf. também Atkins, Clear e Ostler (1992, p.14): "Knowing that your corpus is unbalanced is what counts".

[34] Cf. <https://universaldependencies.org/format.html>.

ferramenta de gerenciamento de dados utilizada, o *lexicographR*). De outro, implica em textos em português (quer dizer, traduções) devidamente segmentados em períodos (*sentences*) compatíveis com a segmentação dos textos latinos.

O total de textos latinos em domínio público disponíveis na Web para reuso constitui um universo proporcional ao longo período de utilização do latim como língua de cultura na Europa ocidental. Um catálogo abrangente[35] contabiliza mais de 3.200 autores e mais de 5.300 obras escritas em latim entre os séculos III a.C. e XX d.C. Boa parte desse material está disponível em formato eletrônico em diversos repositórios, sendo o maior deles o *Corpus corporum*, que consiste em um conjunto de textos, lematizados segundo um padrão próprio, que somam aproximadamente 215 milhões de palavras latinas.[36] Outro repositório importante, sobretudo por conter textos lematizados e formatados segundo os padrões preferenciais mencionados acima, é a LiLaKB, com textos de diversas épocas cujas dimensões atingem a marca de quase 15 milhões de palavras latinas.[37] Do conjunto destaca-se o *LASLA corpus* (Fantoli et al., 2022), um conjunto de textos lematizados e morfologicamente anotados contendo quase dois milhões de *tokens* distribuídos por 131 obras de 21 autores, em sua maioria representantes do período clássico da Literatura Latina.

No que diz respeito a *corpora* de textos latinos sintaticamente anotados (*treebanks*), os números são bem mais modestos. O repositório do projeto Universal Dependencies (UD)[38] (Nivre et al., 2020) contabiliza um total de aproximadamente 1 milhão de *tokens* em seis *corpora* em formato CoNLL-U, quase todos resultantes de processos de conversão automática dos padrões de anotação específicos de cada projeto para o esquema UD. A única exceção é o *UD_Latin-CIRCSE*[39], um *treebank* anotado originalmente segundo o padrão UD que contém textos de Sêneca e Tácito, contabilizando pouco mais de 18 mil *tokens*. Os demais estão listados abaixo:

- *UD_Latin–Perseus*, que corresponde a uma parte do *Latin Dependency Treebank* (Bamman; Crane, 2007) – projeto de anotação sintática de textos em latim e grego clássicos vinculado à Perseus Digital Library – convertida para o esquema UD, contendo aproximadamente 29 mil *tokens*, em sua maioria pertencentes a textos representantes do latim clássico (pouco mais de 27 mil), com uma pequena parcela (2 mil *tokens*) representando o latim tardio da *Vulgata*, a Bíblia oficial em latim;

- *UD_Latin–PROIEL*, constituído por uma parte do *PROIEL Treebank* (Haug; Jøhndal, 2008) – *corpus* de textos morfológica e sintaticamente anotados de diversas línguas indo-europeias – convertida para os padrões da UD, contendo aproximadamente 200 mil *tokens*,

[35] Cf. <https://catalog.digitallatin.org/>.
[36] Cf. <https://mlat.uzh.ch/home>.
[37] Dados obtidos através do terminal SPARQL da plataforma LiLa (<https://lila-erc.eu/sparql/>) por meio de *script* elaborado pelo autor.
[38] Cf. <https://universaldependencies.org/la/>
[39] Cf. <https://github.com/CIRCSE/UD_Latin-CIRCSE>.

com mais da metade representando o latim tardio da *Vulgata* (110 mil) e de um tratado agrícola de Paládio (10 mil), outra parte representando o latim de César e Cícero (80 mil);

- *UD_Latin–LLCT*, resultado da conversão do *Late Latin Charter Treebank* (Korkiakangas, 2021), um *corpus* com aproximadamente 240 mil *tokens* de documentos jurídicos representantes do latim medieval (séculos VIII–IX);
- *UD_Latin–ITTB*, resultado da conversão do *Index Thomisticus Treebank* (Cecchini et al., 2018), contendo 450 mil *tokens* representantes do latim medieval presente nas obras de Tomás de Aquino (século XIII);
- *UD_Latin–UDante* (Cecchini et al., 2020), um resultado da conversão e anotação das obras em latim de Dante Alighieri (século XIV), com aproximadamente 55 mil *tokens*.

Quanto às traduções de textos latinos em língua portuguesa, a situação é menos favorável. A despeito de essa atividade ser praticada há séculos e das muitas pesquisas em estudos da tradução dos clássicos no Brasil, apenas recentemente se produziu um primeiro mapeamento de autores e obras latinas traduzidas para o português brasileiro (Fernandes, 2017), que utilizou como fonte o levantamento de itens bibliográficos relativos à literatura latina presentes nos acervos das bibliotecas brasileiras realizado por Tuffani (2006). Ambos certamente podem servir como ponto de partida para a elaboração de um *corpus* de traduções, mas não respondem às necessidades imediatas deste trabalho, quais sejam, textos em formato eletrônico e, de preferência, segmentados pelo menos no nível dos períodos sintáticos (*sentences*). Diante disso, ganha importância o conjunto de traduções acadêmicas de textos latinos variados feitas José Dejalma Dezotti[40] e gentilmente cedidas para a realização deste trabalho. Esse *corpus* encerra textos de diversos autores representantes do latim clássico e pós-clássico,[41] com datação entre os séculos I a.C. a II d.C., contabilizando quase 300 mil palavras latinas traduzidas.

É, portanto, da intersecção entre os textos latinos e as traduções disponíveis que nascem dois protótipos de *corpora* paralelos latim-português com tradução alinhada no nível do período. O primeiro, denominado **translatinumCor-pt**, é constituído de textos latinos lematizados,[42] destinados a serem processados pelo aplicativo *lexicographR* para a extração de dados, a produção de visualizações e a compilação dos verbetes do **eLatim**. O segundo, denominado **translatinumCor-ptree**, contém textos latinos lematizados e sintaticamente anotados, destinados a servir de fonte de dados para uma ferramenta de observação direta de padrões sintag-

[40] Professor aposentado do Departamento de Linguística, Literatura e Letras Clássicas da Faculdade de Ciências e Letras da Universidade Estadual Paulista (câmpus de Araraquara).

[41] Segundo a periodização tradicional, encontrada por exemplo em Cuzzolin e Haverling (2009). Uma perspectiva semelhante, que inclui uma breve apreciação das características de cada "fase" da língua com base na oposição entre língua intelectual e língua afetiva, encontra-se em Marmorale e Bartolomeu Júnior (1974, p.13–21).

[42] No caso dos *corpora* não incluídos na LiLaKB, como o *UD_Latin-Perseus* e o *UD_Latin-Proiel*, a lematização original foi harmonizada utilizando-se os identificadores do *LiLa Lemma Bank*, através do mesmo processo de mapeamento utilizado para lematizar os dicionários tradicionais, descrito acima.

máticos e fraseológicos, denominada **eLatim-vallex**, cuja construção é abordada no próximo capítulo.[43]

Uma vez que a compilação de ambos os *corpora* se fundamenta sobretudo na disponibilidade de textos em condições de reuso, cabe ao menos descrever seu conteúdo e observar o quanto (e em que proporções) eles são representativos das variedades de latim conhecidas. A esse respeito, aliás, é preciso reconhecer que a taxonomia de classificação de textos não resulta de um procedimento objetivo, mas sim de um ato de interpretação (Baker; Hardie; McEnery, 2006); assim, cabe ao compilador descrever com precisão na documentação do *corpus* o modo como tal classificação é concebida. Neste trabalho, definem-se dez gêneros textuais, divididos simetricamente em dois conjuntos de cinco subgêneros segundo a distinção básica entre poesia e prosa – o critério mais simples relacionado à variação linguística segundo (Hunston, 2008, p.155) –, combinado com especificadores derivados das distinções tradicionais dos gêneros poéticos (lírico, épico, dramático, didático e satírico) e prosaicos (historiográfico, oratório, retórico, filosófico, erudito, epistolográfico, ficcional).[44] Assim, os dez gêneros utilizados para classificar os textos dos *corpora* paralelos compilados neste trabalho são os seguintes: (1) poesia lírica (inclui odes, elegias, epigramas, éclogas); (2) poesia cômica (i.e. dramas do tipo comédia); (3) poesia trágica (i.e. dramas do tipo tragédia); (4) poesia épica (i.e. epopeias do tipo heroico); (5) poesia didática (inclui a chamada épica didática, bem como as sátiras e as fábulas); (6) prosa filosófica (inclui textos filosóficos de todo tipo, das discussões teóricas e dialogadas aos tratados sobre atividades práticas); (7) prosa oratória (i.e. textos elaborados como discursos de tipo epidítico, deliberativo ou judiciário); (8) prosa epistolográfica (i.e. cartas); (9) prosa histórica (inclui textos de história geral, bem como monografias e biografias); e (10) prosa ficcional (i.e. textos narrativos de caráter não historiográfico).

A Tabela 13 traz um sumário das obras que compõem o *corpus*. Cada obra é armazenada em um único arquivo em formato CoNLL-U, cujo nome (‘filename’) é composto por sua forma abreviada de citação – baseada na lista de autores e obras do *Oxford Latin dictionary* (Glare, 1968) – somada à indicação do *corpus*-fonte do texto latino (‘Lasla’, ‘Perseus’ ou ‘Proiel’). A cada arquivo se associa um conjunto de metadados que inclui informações como nome do autor, título da obra, gênero, período de composição (século), contagem de sentenças e palavras e percentual da obra recoberto pela amostra. Tais metadados servem para a extração de frequência normalizada e informações sobre dispersão lexical (ver adiante), bem como permitem avaliar a representatividade do *corpus* em relação à variação linguística do latim tanto de uma perspectiva diacrônica (por séculos) quanto da perspectiva literária (por gêneros).

[43] Os arquivos-fonte e *scripts* de processamento de ambos os *corpora* se encontram em <https://github.com/lucascdz/translatinumCor-pt>.

[44] Segundo a divisão encontrada em Cardoso (1989).

Tabela 13 – Relação das obras que compõem o protótipo do TranslatinumCor-pt.

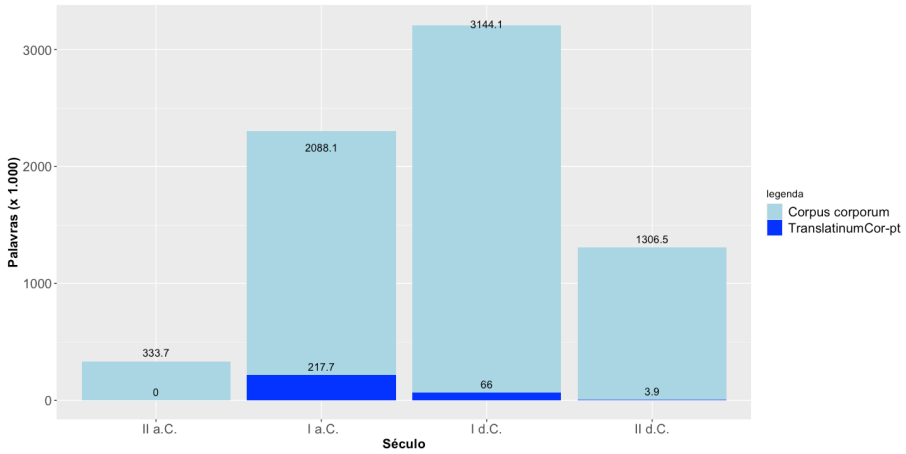
nome do arquivo	autor	obra	gênero	século	sentences	tokens	amostra (%)
CaesGal__Lasla	Caesar	Bellum Gallicum	Prosa histórica	I a.C.	1043	27033	58%
CicAmici__Lasla	Cicero	De Amicitia	Prosa filosófica	I a.C.	453	9562	100%
CicArch__Lasla	Cicero	Pro Archia	Prosa oratória	I a.C.	111	3197	100%
CicAtt__Proiel	Cicero	Epistulae ad Atticum	Prosa epistolográfica	I a.C.	3306	37739	92%
CicBalb__Lasla	Cicero	Pro Balbo	Prosa oratória	I a.C.	2	41	1%
CicCael__Lasla	Cicero	Pro Caelio	Prosa oratória	I a.C.	408	8692	100%
CicCatil__Lasla	Cicero	In Catilinam	Prosa oratória	I a.C.	520	12899	100%
CicDeiot__Lasla	Cicero	Pro Rege Deiotario	Prosa oratória	I a.C.	195	4020	100%
CicDom__Lasla	Cicero	De Domo Sua	Prosa oratória	I a.C.	37	1108	7%
CicFont__Lasla	Cicero	Pro M Fonteio	Prosa oratória	I a.C.	2	14	< 1%
CicLig__Lasla	Cicero	Pro Ligario	Prosa oratória	I a.C.	189	3335	100%
CicMan__Lasla	Cicero	De Imperio Cn Pompei	Prosa oratória	I a.C.	237	6822	100%
CicMarc__Lasla	Cicero	Pro Marcello	Prosa oratória	I a.C.	117	2855	100%
CicMil__Lasla	Cicero	Pro Milone	Prosa oratória	I a.C.	480	10712	100%
CicMur__Lasla	Cicero	Pro Murena	Prosa oratória	I a.C.	207	4847	45%
CicOff__Proiel	Cicero	De officiis	Prosa filosófica	I a.C.	46	863	9%
CicPhil__Lasla	Cicero	Philippica Oratio	Prosa oratória	I a.C.	1304	24033	44%
CicRabPerd__Lasla	Cicero	Pro C Rabirio Perduellionis Reo	Prosa oratória	I a.C.	133	3624	100%
CicSen__Lasla	Cicero	De Senectute	Prosa filosófica	I a.C.	350	8438	100%
CicVer1__Lasla	Cicero	In Verrem Actio Prima	Prosa oratória	I a.C.	158	4818	100%
CicVer2__Lasla	Cicero	In Verrem Actio Secunda	Prosa oratória	I a.C.	700	14509	15%
HorCarm__Lasla	Horatius	Carmina	Poesia lírica	I a.C.	197	4160	30%

Tabela 13 – Relação das obras que compõem o protótipo do TranslatinumCor-pt (cont.).

nome do arquivo	autor	obra	gênero	século	sentences	tokens	amostra (%)
HorEpod__Lasla	Horatius	Epodi	Poesia lírica	I a.C.	3	66	2%
HorS__Lasla	Horatius	Sermones	Poesia lírica	I a.C.	381	7284	49%
OvMed__Lasla	Ovidius	Medicamina Faciei Femineae	Poesia lírica	I a.C.	42	640	100%
OvMet__Perseus	Ovidius	Metamorphoseon	Poesia didática	I d.C.	317	5202	100%
Phaed__Perseus	Phaedrus	Fabulae	Poesia didática	I d.C.	582	6587	100%
SalCat__Lasla	Sallustius	Catilina	Prosa histórica	I a.C.	492	10917	100%
SenAg__Lasla	Seneca	Agamemnon	Poesia trágica	I d.C.	409	5714	100%
SenEp__Lasla	Seneca	Ad Lucilium Epistulae Morales	Prosa filosófica	I d.C.	2135	29064	24%
SenHelv__Lasla	Seneca	Ad Helviam Matrem De Consolatione	Prosa filosófica	I d.C.	302	6835	100%
SenMarc__Lasla	Seneca	Ad Marciam De Consolatione	Prosa filosófica	I d.C.	6	311	4%
SenOed__Lasla	Seneca	Oedipus	Poesia trágica	I d.C.	401	6098	100%
SenOt__Lasla	Seneca	De Otio	Prosa filosófica	I d.C.	82	1993	100%
SenProv__Lasla	Seneca	De Providentia	Prosa filosófica	I d.C.	221	4154	100%
SuetAug__Perseus	Suetonius	Divus Augustus	Prosa histórica	II d.C.	38	907	11%
TacHist__Lasla	Tacitus	Historiae	Prosa histórica	II d.C.	126	2945	5%
VergA__Lasla	Vergilius	Aeneis	Poesia épica	I a.C.	226	5454	8%
VulgMarc__Proiel	Hieronymus	Evangelium secundum Marcum	Prosa ficcional	IV d.C.	509	4451	43%
TOTAL					**16467**	**291943**	

Fonte: elaboração própria.

Figura 24 – Gráfico da representatividade do TranslatinumCor-pt quanto à variação diacrônica por século (em milhares de *tokens*).



Fonte: elaboração própria.

Do ponto de vista cronológico, a Tabela 14 traz uma contagem aproximada de palavras dos quatro *corpora* mencionados acima – a saber, (1) o Corpus Corporum ('CorpCorp') representando o conjunto geral de textos latinos de todas as épocas, (2) a LiLaKB representando a parte dos textos lematizadas segundo o LiLa Lemma Bank, (3) o conjunto de *treebanks* no padrão UD ('UD_Latin') e (4) o TranslatinumCor-pt ('transCor-pt') – agrupados em períodos de cinco séculos; desse modo, percebe-se que a quase totalidade dos textos incluídos no translatinumCor-pt representa textos escritos entre os séculos III a.C. e II d.C., que coincide com um período de uso do latim conhecido tradicionalmente como Antiguidade Clássica. A Figura 24, por sua vez, ilustra tanto a proporção externa (i.e. em relação ao total de textos) quanto a proporção interna ao próprio *corpus* dos textos abarcados pelo translatinumCor-pt em cada século do período mencionado acima; desse modo, fica evidente que o século mais bem representado no *corpus* é o século I a.C.

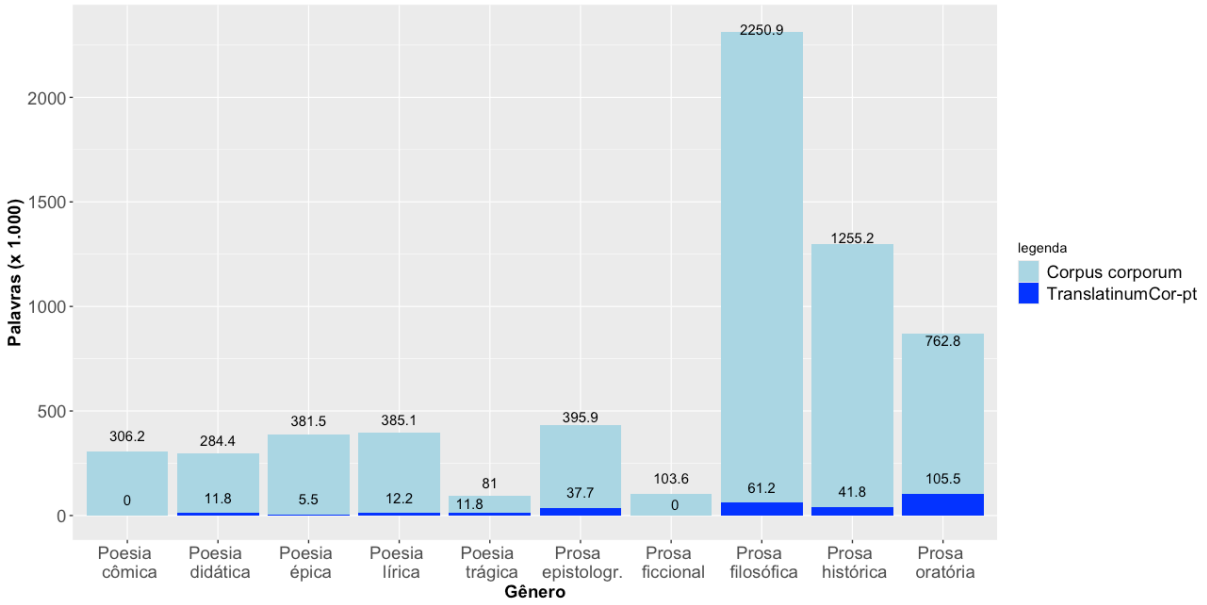
A representatividade quanto aos gêneros é sintetizada pela Figura 25. percebe-se que quase todos os gêneros cultivados na Antiguidade Clássica são contemplados pelo *corpus* paralelo, sendo exceções apenas a poesia cômica e a prosa ficcional. Por outro lado, também a representação se mostra desigual, tanto externamente quanto internamente. No primeiro caso, se as

Tabela 14 – Tamanho dos *corpora* consultados, delimitados por período (em milhões de *tokens*).

Período	CorpCorp	LiLaKB	UD_Latin	transCor-pt
séc. III a.C. a II d.C.	6,94	1,87	0,13	0,29
séc. III d.C. a VII d.C.	36,53	8,36	0,12	< 0,01
séc. VIII a XII d.C.	51,95	1,02	0,24	0
séc. XIII a XVII d.C.	99,48	0,48	0,51	0
séc. XVIII a XXI d.C.	2,64	0,09	0	0

Fonte: elaboração própria.

Figura 25 – Gráfico da representatividade do TranslatinumCor-pt quanto à variação por gênero textual (em milhares de *tokens*).



Fonte: elaboração própria.

amostras de poesia trágica e de prosa oratória chegam a quase 15% de cobertura do total de textos do gênero produzidos na Antiguidade Clássica (13,8% e 14,6%, respectivamente), as amostras de poesia épica e de prosa filosófica não chegam aos 3% (1,4% e 2,8%, respectivamente).[45] No segundo, os dois extremos são a prosa oratória (com 105 mil *tokens*) e a poesia épica (com apenas 5,5 mil).

Resta detalhar a composição do TranslatinumCor-ptree, resultante da intersecção entre os textos traduzidos e os *corpora* sintaticamente anotados no padrão UD. Dada a possibilidade de haver diferenças de lematização entre os diferentes treebanks, todos os lemas dos textos utilizados foram mapeados pelos identificadores do LiLa Lemma Bank, para garantir uma indexação homogênea do *corpus* resultante. A Tabela 15 traz um sumário das obras e das respectivas quantidades de sentenças e *tokens* que compõem o referido *corpus*. Fica evidente que é um *corpus* bastante limitado, com textos de sete autores, dos quais a maioria é de prosadores (Cícero, César, Salústio, Suetônio e Jerônimo), com apenas dois poetas (Ovídio e Fedro). Também da perspectiva cronológica, o século I a.C. é muito mais bem representado do que os demais, respondendo por 86% dos *tokens*.

[45] A propósito, as porcentagens de cobertura dos demais gêneros são: poesia didática (4,1%), poesia lírica (3,2%), prosa epistolográfica (9,6%) e prosa histórica (3,3%).

Tabela 15 – Relação das obras que compõem o protótipo do TranslatinumCor-ptree.

nome do arquivo	autor	obra	gênero	século	sentences	tokens	amostra (%)
CicAtt__Proiel	Cicero	Epistulae ad Atticum	Prosa epistolográfica	I a.C.	3306	37739	92%
CicOff__Proiel	Cicero	De officiis	Prosa filosófica	I a.C.	46	863	9%
CaesGal__Proiel	Caesar	Commentarii belli Gallici	Prosa histórica	I a.C.	1386	26244	96%
SalCat__Perseus	Sallustius	Catilina	Prosa histórica	I a.C.	336	4999	38%
CicCatil__Perseus	Cicero	In Catilinam	Prosa oratória	I a.C.	137	1897	29%
OvMet__Perseus	Ovidius	Metamorphoseon	Poesia didática	I d.C.	183	2649	51%
Phaed__Perseus	Phaedrus	Fabulae	Poesia didática	I d.C.	389	4116	62%
SuetAug__Perseus	Suetonius	Divus Augustus	Prosa histórica	II d.C.	15	318	4%
VulgMarc__Proiel	Hieronymus	Evangelium secundum Marcum	Prosa ficcional	IV d.C.	509	4451	43%
TOTAL					**6307**	**83276**	

Fonte: elaboração própria.

4.3.3 Extração e visualização de dados: frequência, colocações, exemplos

O uso de *corpus* em lexicografia pode se dar por três abordagens, segundo Tiberius et al. (2022). Na abordagem orientada por *corpus* (*corpus-driven*), os dados extraídos são determinantes de todos os elementos do verbete, como definição da nomenclatura, distinção de acepções e seleção de exemplos. Na abordagem fundamentada em *corpus* (*corpus-based*), este tem um papel importante, porém algumas decisões não dependem dos dados extraídos (e.g. a nomenclatura pode ser definida por uma lista à parte, independente de sua ocorrência no *corpus*). Na abordagem incrementada por *corpus* (*corpus-informed*), por sua vez, o uso dos dados extraídos é apenas pontual (e.g. dicionários existentes podem ter seus exemplos lexicográficos enriquecidos por sentenças coletadas em um *corpus*).

Heid (2008, p.138–9) distingue as duas primeiras abordagens em termos de projeção apriorística de categorizações linguísticas sobre o *corpus*. Assim, a abordagem *corpus-driven* evitaria esse tipo de projeção (e consequentemente os muitos tipos de pre-processamento automático de texto) para confiar exclusivamente (ou principalmente) nos dados distribucionais obtidos estatisticamente, acreditando que as propriedades combinatórias, semânticas e sintáticas podem ser inferidas intelectualmente pela observação dos dados. A abordagem *corpus-based*, por sua vez, adota o pre-processamento computacional do *corpus* (incluindo os possíveis erros de classificação que ele pode conter) e utiliza hipóteses de linguística descritiva como fundamento para a busca, extração e classificação automática dos dados obtidos. Nesse sentido, Teubert (2001) chama atenção para o fato de que, dos três procedimentos envolvidos na pesquisa de *corpus* – a saber, identificação dos dados (por categorização), correlação entre os dados (por cálculo estatístico) e interpretação dos dados –, os dois primeiros são automatizáveis, mas o terceiro não; de fato, este pressupõe a intencionalidade humana: "toda interpretação é um ato que envolve consciência e, portanto, não pode ser transferida para um algoritmo" (Teubert, 2001, p.129).[46]

Neste trabalho, pode-se afirmar que o uso de dados de *corpus* oscila entre uma abordagem *corpus-based* – na medida em que parte de textos lematizados e, em menor escala, sintaticamente anotados – e *corpus-informed* – na medida em que enriquece o verbete com dados extraídos de *corpus*, porém sem correlacioná-los com as acepções e outras distinções sintáticas, semânticas e pragmáticas sugeridas pelos dicionários e bases lexicais. Com efeito, os dados extraídos de *corpus* apresentados nos verbetes do **eLatim** resultam exclusivamente dos procedimentos computacionais proporcionados pelo aplicativo *lexicographR*, não sendo acompanhados de nenhuma forma de interpretação dos dados. Com isso, fica a cargo do usuário fazer as devidas associações com o contexto de seu problema de consulta, bem como utilizar os dados apresentados como indícios para uma eventual pesquisa de *corpus* propriamente dita.

[46] No original: "Whilst the first two steps should be done automatically as much as possible, the last step requires human intentionality, as any interpretation is an act involving consciousness and, therefore, not transmutable into an algorithmic procedure. This is the main difference between corpus linguistics and computational linguistics, which reduces language to a set of procedures.". Para o autor, aliás, é na ausência ou presença da etapa interpretativa da pesquisa que residiria a diferença, respectivamente, entre a linguística computacional e a linguística de corpus.

Tabela 16 – Frequência absoluta e normalizada (por século) de lemas iniciados por FACI–.

	I.a.C.		I.d.C.		II.d.C.		IV.d.C.	
lema	Freq	Norm	Freq	Norm	Freq	Norm	Freq	Norm
facies	12	0,56	22	3,37	0	0,00	1	2,25
facile	129	6,00	17	2,60	2	5,24	0	0,00
facilis	27	1,26	23	3,52	1	2,62	1	2,25
facilitas	8	0,37	3	0,46	1	2,62	0	0,00
facinorosus	6	0,28	0	0,00	0	0,00	0	0,00
facinus	77	3,58	5	0,76	0	0,00	0	0,00
facio	1037	48,27	273	41,76	7	18,32	40	89,87

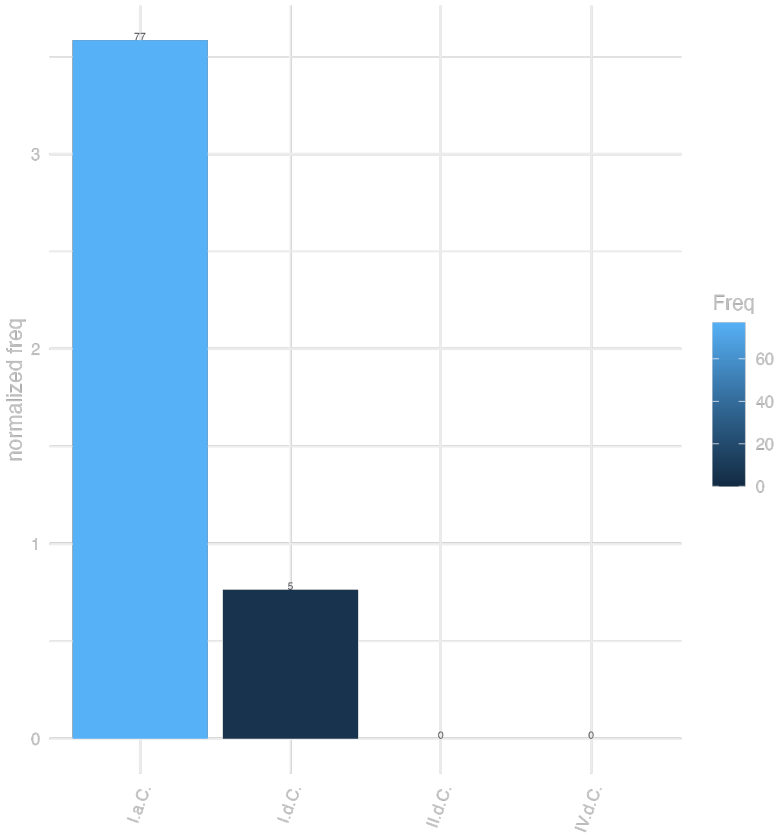
Freq = frequência absoluta; **Norm** = frequência normalizada (por 10 mil *tokens*).
Fonte: elaboração própria.

O *lexicographR* oferece recursos para a extração de dados sobre frequência, colocações e exemplos, considerados os principais tipos de evidências de interesse lexicográfico associadas ao uso de *corpus* (Heid, 2008; Tiberius et al., 2022).

O cálculo de frequência é realizado automaticamente assim que o *corpus* é inserido no sistema. Com a inserção de um arquivo adicional contendo os metadados de cada texto do *corpus* (i.e. indicações de autor, gênero e período), pode-se calcular a dispersão do lema, permitindo assim observar o quão homogênea é a distribuição das ocorrências do lema por *subcorpora* (McEnery; Wilson, 2001). Uma vez que o *corpus* não é balanceado, o aplicativo considera a frequência normalizada (i.e., ocorrências para cada 10 mil *tokens*), neutralizando assim a diferença de tamanho dos *subcorpora*. A Tabela 16 traz uma amostra da planilha resultante para lemas iniciados por *faci–*, considerando o critério cronológico de delimitação de subcorpora (= frequência normalizada por século). A mesma planilha é usada como fonte para a produção de uma visualização gráfica dos dados, que certamente favorece uma interpretação mais ágil. A Figura 26 ilustra esse recurso através de um gráfico de barras relativo ao uso do lema FACINUS, que permite comparar as frequências normalizadas (representadas pela altura da barra), ao mesmo tempo que informa a frequência absoluta através de um indicador numérico no alto de cada barra; no caso, percebe-se claramente que, nos limites do *corpus* utilizado, o uso do referido lema perde força no século I d.C., caindo completamente em desuso nos séculos seguintes.

O cálculo de possíveis colocações, ou melhor, dos candidatos a colocados em relação a cada lema, é feito com base em parâmetros regulados pelo usuário, que permitem indicar eventuais lemas a serem desconsiderados (*stopwords*), ajustar a janela de observação (*kwic window*) – isto é, o número de palavras antes e depois do lema em foco –, determinar uma frequência mínima para o colocado, definir a quantidade de candidatos a ser extraída bem como os limites mínimos para os resultados relativos a cada uma das fórmulas estatísticas utilizadas pelo aplicativo – a saber, Log-Likelihood (LogL), Log Ratio (LogR), Mutual Information (MI) e LogDice (LogD). A esse respeito, é importante ter em mente que tais fórmulas, embora compartilhem o objetivo de capturar associações entre as palavras, resultam cada uma em uma métrica distinta,

Figura 26 – Gráfico de barras indicando a frequência do lema *facinus* normalizada por século.



Fonte: elaboração própria.

sendo adequadas para contextos distintos (Tiberius et al., 2022; Kilgarriff; Kosem, 2012). O cálculo da MI, por exemplo, detecta pares de palavras que aparecem frequentemente juntas e raramente com outras palavras, de modo que costuma ser bem sucedido na detecção de frases feitas e termos técnicos; porém, enfatiza combinações de palavras raras e com erros de grafia, não sendo adequado para grandes *corpora* (Tiberius et al., 2022). Já cálculo do Log-Likelihood oferece melhores candidatos, porém nesse caso a lista resultante costuma ser dominada por palavras funcionais. O cálculo do LogDice, por sua vez, se mostra uma opção melhor quando se quer comparar *corpus* de tamanhos variados, por não ser afetado pelo tamanho do *corpus*; esse indicador mede a tipicalidade, isto é, a força da associação entre as palavras com base em sua frequência individual e em combinação.

Diante dessas diferenças, optou-se neste trabalho, a título de experimentação, por produzir quatro listas de colocados – cada uma privilegiando uma das fórmulas, com os demais parâmetros comuns (janela de uma palavra antes ou depois e frequência mínima de dez ocorrências para o colocado) – e, em seguida, unificá-las, de modo a incluir os candidatos a colocado indicados por cada fórmula em uma única tabela. Assim, o usuário, ao acessá-la, pode observar os colocados segundo o índice que sirva melhor aos seus propósitos. Além disso, para uma melhor visualização dos resultados, cinco gráficos em formato de nuvem de palavras são produzidos,

Tabela 17 – Candidatos a colocados para o lema *senatus*.

lema	colocado	classe	LogL	LogR	MI	LogD	Total
senatus	consultum	substantivo	579,00	10,70	6,10	12,30	608,10
senatus	in	preposição	153,50	2,40	2,20	10,50	168,60
senatus	frequens	adjetivo	121,30	7,60	5,30	11,10	145,30
senatus	auctoritas	substantivo	107,40	5,20	4,00	11,10	127,70
senatus	decerno	verbo	56,80	5,20	4,00	10,70	76,70
senatus	ab	preposição	60,80	2,50	2,20	10,40	75,90
senatus	ad	preposição	55,80	2,30	2,10	10,30	70,50
senatus	ex	preposição	31,80	2,30	2,10	10,20	46,40
senatus	refero	verbo	25,80	4,00	3,30	10,10	43,20
senatus	decretum	substantivo	22,50	5,60	4,20	9,80	42,10
senatus	populus	substantivo	21,80	2,70	2,30	10,10	36,90
senatus	appello	verbo	17,20	3,90	3,20	9,80	34,10

LogL = log-likelihood; **LogR** = logRatio; **MI** = Mutual Information; **LogD** = logDice.
Fonte: elaboração própria.

Figura 27 – Nuvem de palavras com os candidatos a colocado da classe dos substantivos para o lema *senatus*.



Fonte: elaboração própria.

um para cada classe lexical (substantivo, adjetivo, verbo, advérbio) e um contendo as palavras funcionais, utilizando como valor de referência para o tamanho das palavras a soma dos quatro índices. Uma amostra da lista de candidatos a colocados para o lema *senatus* é apresentada na Tabela 17, ao passo que a Figura 27 mostra a nuvem de palavras produzida a partir dessa lista com os candidatos a colocados da classe dos substantivos.

Para a extração de exemplos lexicográficos do *corpus*, o *lexicographR* permite utilizar os princípios e parâmetros do GDEx (Kilgarriff et al., 2008). Trata-se de uma técnica de extração automática de "bons exemplos lexicográficos" ("good dictionary examples", daí a sigla *GDEx*) de um *corpus* por meio da atribuição de uma pontuação para cada período representado na linha de concordância, calculada com base em um sistema de valores numéricos positivos (*bonuses*) e

negativos (*penalties*) conforme certas propriedades pré-definidas sejam consideradas desejadas ou indesejadas, respectivamente. De acordo com essa proposta, um bom exemplo lexicográfico deve ter: (1) tipicidade, isto é, deve mostrar padrões de uso frequentes e com boa dispersão; (2) informatividade, isto é, deve ajudar a esclarecer a definição e o significado; e (3) legibilidade, isto é, deve ser inteligível ao usuário, evitando incluir sem necessidade palavras raras e estruturas difíceis, nomes próprios obscuros ou distrativos, bem como referências anafóricas ou dêiticas que não podem ser compreendidas sem acesso a um contexto mais amplo (Kilgarriff et al., 2008, p.2). A tradução desses princípios para parâmetros práticos e mensuráveis inclui propriedades como tamanho da sentença, presença de palavras raras e formas de anáfora, e presença de candidatos a colocado, que podem ser ajustadas de modo personalizado no *lexicographR*.

Para a compilação do protótipo do **eLatim**, buscou-se extrair até cinquenta sentenças para cada lema, utilizando-se os seguintes parâmetros:

- *tamanho da sentença*: entre seis e dez palavras; para cada palavra a menos, o sistema atribui uma penalização de –5 pontos; para cada palavra extra, –1 ponto;
- *presença de colocados*: uma sentença com um ou mais colocados recebe um bônus de +5 pontos;
- *presença de palavras pouco frequentes*: penalização de –1 ponto para cada palavra na sentença fora das 5 mil palavras mais frequentes (conforme a frequência dos lemas no *corpus* da Lila Knowledge Base);
- *presença de anáfora*: penalização de –1 ponto para cada pronome presente na sentença (conforme a lista de lemas classificados como pronomes no LiLa Lemma Bank).

O resultado é uma planilha com quase 8 mil sentenças indexadas pelo lema e organizadas por tamanho (em número de palavras), às quais se agregam as respectivas tradução, nome abreviado do tradutor e a referência abreviada à fonte do exemplo. A Tabela 18 traz uma amostra com as doze primeiras sentenças dessa planilha. Vale notar que o texto latino foi comparado ao texto utilizado pelo tradutor; quando houve diferença significativa, a indicação "de outra edição" foi acrescentada às iniciais do tradutor.

4.3.4 Uma base de dados sintático-semânticos

Por fim, uma base de dados sintático-semânticos é extraída do **TranslatinumCor-ptree**, o *corpus* paralelo obtido dos textos latinos sintaticamente anotados, descrito anteriormente neste capítulo.

Nesse caso, o processamento não utiliza o *lexicographR*, mas é feito à parte através de um programa em linguagem R escrito especialmente para tal fim. Para cada palavra em posição de regente, o programa extrai as palavras indicadas como seus dependentes e a função sintática

Tabela 18 – Amostra da planilha com exemplos extraídos automaticamente segundo o modelo GDEx, acompanhados de tradução e referência.

lemma	loc.cit.	texto	traducao	tradutor
ab	Cic.Mil.59	ab Appio	"Da parte de Ápio."	[JDD]
bonus	Cic.Mil.59	di boni	Bons deuses!	[JDD]
ciuis	Cic.Att.4.15.7	o civem	O cidadão!	[JDD]
concedo	Cic.Att.4.2.4	non concedebant	Não concediam.	[JDD]
credo	Cic.Att.5.19.3	non credo	"Não acredito"	[JDD, ≠ edição]
deus	Cic.Mil.59	di boni	Bons deuses!	[JDD]
dico	Cic.Ver.1.33	dicam aperte	Vou dizer abertamente.	[JDD]
dico	Cic.Att.2.19.1	dices fortasse	Dirás talvez.	[JDD]
falsus	Sen.Ep.121.3	falsum est	É falso.	[JDD]
fors	Cic.Att.2.19.1	dices fortasse	Dirás talvez.	[JDD]
hic	Cic.Att.2.11.1	haec igitur	É isso então	[JDD]
igitur	Cic.Att.2.11.1	haec igitur	É isso então	[JDD]

Fonte: elaboração própria.

de cada uma delas, conforme anotação em padrão UD do *treebank* utilizado como fonte. Em seguida, classifica o tipo de sintagma que constituem (nominal, adjetival, adverbial, preposicional e oracional) e atribui ao dependente um traço semântico (humano, animado, concreto, abstrato), se disponível.[47]

A formulação dos rótulos utilizados para essa caracterização do dependente segue os seguintes parâmetros:

- sintagmas nominais são rotulados por um N maiúsculo acompanhados da indicação abreviada do caso morfológico entre parênteses, e.g. *N(abl), N(acc), N(dat), N(gen)*, etc.
- sintagmas preposicionais são rotulados pela abreviação 'Sprep' seguida da preposição entre parênteses, e.g. *Sprep(ab), Sprep(ad), Sprep(cum), Sprep(de), Sprep(ex), Sprep(pro)*, etc.
- sintagmas oracionais são rotulados pela abreviação 'Or' acompanhados do modo do verbo subordinado, precedido de conjunção conforme o contexto, e.g. *Or(cum+ind), Or(cum+sub), Or(cur+sub), Or(dum+ind), Or(ne+sub), Or(quia+ind)*, etc.
- sintagmas adjetivais e adverbiais são indicados simplesmente pelas abreviaturas *Adj* e *Adv*, respectivamente.

Ao final do trabalho, é produzida uma planilha completa contendo todos os pares regente-dependente encontrados no *corpus*, acompanhados de caracterização do regente (lema) e do dependente (sintagma, traço, função), bem como da sentença completa com referência e tradução

[47] A lista de palavras anotadas com traço semântico se resume a 540 lemas e foi adaptada do trabalho de Márquez Cruz e Fernández-Pampillón (2021).

alinhada. Essa planilha é utilizada como fonte para a produção do **eLatim-vallex**, uma ferramenta de consulta sintático-semântica que utiliza a caracterização acima como filtro de seleção de sentenças para observação direta de dados linguísticos, como se verá adiante.

4.4 CONCLUSÃO DO CAPÍTULO

Este capítulo se dedicou a selecionar e integrar em uma única base de dados relacionais, indexada por lemas, uma variedade de recursos lexicais e textuais, a fim de utilizá-los como fonte para a produção do dicionário digital latim-português **eLatim**.

Foram descritos os procedimentos de coleta de dados em algumas bases lexicais computacionais sobre o léxico latino, bem como os métodos utilizados para transformar os textos digitalizados de dicionários tradicionais em uma coleção de dados estruturados e legíveis por máquina (a **latinolusitanicum-db**), tornando-se assim passíveis de processamento pelo aplicativo de escrita de dicionários *lexicographR*. Também se descreveram os princípios e parâmetros para a compilação de dois *corpora* paralelos de textos latinos com tradução em português, um mais amplo porém apenas lematizado (o **translatinumCor-pt**), outro mais restrito porém sintaticamente anotado (o **translatinumCor-ptree**), bem como os diferentes métodos de extração de dados e produção de visualizações gráficas.

O próximo capítulo se volta para a produção de interfaces de consulta a esse conjunto de dados, desde a compilação dos verbetes que, uma vez reunidos, constituem o dicionário digital latim-português propriamente dito, até a criação de uma ferramenta de consulta direta ao *corpus* paralelo sintaticamente anotado.

5 CONSTRUÇÃO DE INTERFACES DE CONSULTA

Bases de dados como a descrita no capítulo anterior são recursos computacionais, isto é, são compiladas e estruturadas para leitura e processamento por computadores. No caso da leitura por humanos, em contrapartida, é necessária uma interface de consulta que forneça ao usuário uma ou mais formas de acesso aos dados reunidos. Quando se trata de consulta a dados sobre o léxico, pode-se dizer que a interface constitui o próprio dicionário digital.

Dois modelos de interface de consulta são explorados neste capítulo. O primeiro organiza os dados sobre os itens lexicais em verbetes indexados por lema, à semelhança dos dicionários tradicionais, e tem como resultado o protótipo de dicionário digital denominado **eLatim**. O segundo consiste em uma ferramenta de consulta direta ao *corpus* paralelo, denominada **eLatim-vallex**, que busca oferecer ao usuário a possibilidade de observar dados linguísticos autênticos e filtrá-los com base na caracterização sintático-semântica dos constituintes das sentenças. Uma vez que são ferramentas distintas, os próximos tópicos apresentam como se dá a construção de cada um, individualmente.

5.1 ELATIM: DICIONÁRIO DIGITAL LATIM-PORTUGUÊS

O **eLatim** é um protótipo de dicionário digital cujos verbetes são produzidos de forma automatizada por meio do aplicativo de escrita de dicionários *lexicographR*, a partir da base de dados compilada no capítulo anterior.

A possibilidade de automatizar o processo de produção de verbetes pressupõe a existência de um *template*, isto é, um modelo a ser utilizado para a geração de verbetes para todos os lemas definidos na nomenclatura. Segundo Abel (2012), a criação de um *template* pressupõe a definição de padrões para três aspectos do verbete: (1) os tipos de informação que contêm; (2) a distribuição dessas informações de forma estruturada; e (3) os estilos e o formato de exibição.

Os tipos de informação são aqueles disponíveis na base de dados lexicais, de cuja compilação se tratou no capítulo anterior. Esta, em resumo, é constituída por diversas planilhas indexadas por lema – parte delas produzida externamente (dados extraídos de dicionários e demais bases lexicais), parte produzida através do mesmo aplicativo (dados extraídos de *corpus*) –, reunindo dados sobre forma, significado e uso.

Para a distribuição estruturada das informações, podem ser utilizados como critério tanto o tipo de informação quanto o tipo de fonte de onde a informação é coletada. No **eLatim**, a estruturação se baseia no tipo de fonte por duas razões. A primeira é a conveniência de se preservar a coerência interna das informações obtidas; esta se mostra especialmente importante no caso dos verbetes de dicionários tradicionais, que apresentam cada um uma distribuição particular, de modo que uma exibição unificada de formas e acepções produziria redundâncias e excesso

de informações.[1] A segunda reside no fato de que cada fonte de informação possui um grau diferente de consistência e confiabilidade, de modo que é preciso garantir ao usuário meios de exercer sua preferência por determinado tipo de fonte.

Os estilos e formato de exibição, por sua vez, dependem das opções oferecidas pelo aplicativo de escrita de dicionários. Para a composição e compilação dos verbetes do **eLatim**, o sistema utilizado é o QuartoBook,[2] uma estrutura de publicação pensada para a publicação de livros digitais como páginas *Web* (i.e., escritas em linguagem HTML) e utilizada pelo *lexicographR* como modelo de organização do dicionário. Na prática, cada verbete funciona como se fosse um capítulo de livro, cujo título é o lema; assim, o sumário se torna a própria nomenclatura, na medida em que contém a lista de todos os lemas contemplados pelo dicionário. Entre as vantagens da adoção de um modelo estruturado como página *Web* está a possibilidade de publicação em qualquer serviço de hospedagem gratuita, ficando acessível a qualquer dispositivo conectado à Internet e dotado de um navegador *Web*. No caso do **eLatim**, o serviço escolhido foi o *GitHub Pages*, pelo fato de estar integrado à plataforma de desenvolvimento de softwares GitHub, onde o código-fonte e a base de dados estão igualmente disponibilizados para acesso aberto e reúso.[3]

A Figura 28 mostra a página inicial do protótipo do **eLatim**, em que se podem observar os elementos básicos resultantes da utilização dessa estrutura. A página se divide verticalmente em duas porções desiguais. A porção menor, à esquerda, traz duas estruturas complementares de acesso ao conteúdo: a primeira, logo abaixo do título, é um campo de busca textual direcionado para todo o conteúdo do dicionário (permitindo, por exemplo, localizar um verbete através de uma forma flexionada, desde que esta esteja presente na microestrutura); a segunda traz a nomenclatura, em forma de lista numerada, cujos itens servem de acesso direto aos verbetes correspondentes. A porção maior ocupa todo o restante da página e exibe o conteúdo do verbete selecionado. No caso da página capturada, o conteúdo exibido é o prefácio do dicionário, onde se fornecem as informações básicas sobre seu conteúdo e sobre as fontes utilizadas para sua constituição; estas, a propósito, ocorrem no texto de forma abreviada, com sua forma completa podendo ser acessada seja por meio de um simples passar do *mouse* sobre a referência (como a referência a Saraiva (1927) exibida no centro da Figura), seja por meio de um clique que leva à bibliografia ao final do texto.

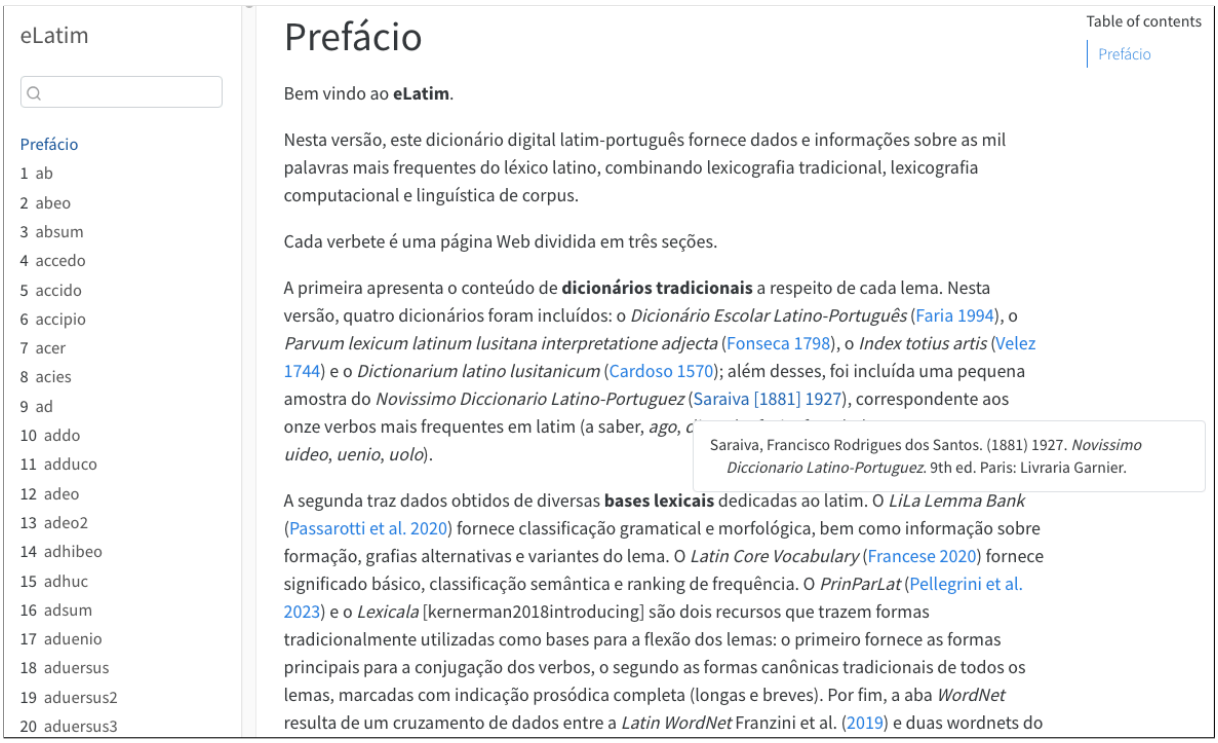
A Figura 29 ilustra a primeira seção do verbete, dedicada aos conteúdos extraídos de dicionários tradicionais. Cinco abas abrigam, cada uma, um dos dicionários processados até o momento, identificados pelo nome do autor. O caso ilustrado se refere ao lema *debeo* e tem selecionada a aba referente ao verbete extraído de Saraiva (1927). Nota-se que o conteúdo é

[1] Sobre os problemas da sobrecarga de informações para o usuário de dicionários digitais, cf. Gouws e Tarp (2016).

[2] Quarto é um sistema de publicação técnica e científica de código aberto que permite a publicação de obras digitais em diversos formatos eletrônicos, de página *Web* a PDF e ePub; seu diferencial reside na possibilidade de utilização integrada às linguagens de programação, a exemplo da linguagem R utilizada pelo *lexicographR*. Para mais informações, cf. <https://quarto.org/>.

[3] Cf. <https://github.com/lucascdz/eLatim>. A licença pública atribuída ao material é a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (<https://creativecommons.org/licenses/by/4.0/>).

Figura 28 – Reprodução de tela da página inicial do **eLatim**, com amostra do prefácio.



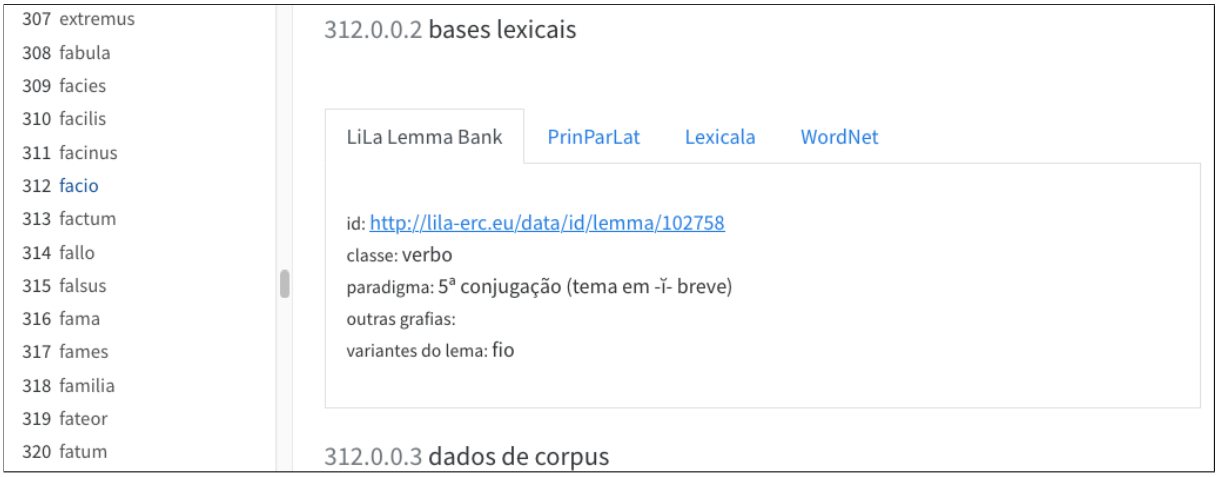
Fonte: <https://lucascdz.github.io/eLatim/index.html>.

Figura 29 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído de dicionários tradicionais.



Fonte: <https://lucascdz.github.io/eLatim/debeo.html>.

Figura 30 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído de bases lexicais.



Fonte: <https://lucascdz.github.io/eLatim/facio.html>.

apresentado de forma menos adensada em relação à versão impressa, aproveitando a possibilidade ilimitada de utilização do espaço da página digital. Essa opção se materializa mediante a utilização de linhas individuais para cada elemento do verbete, preservando todavia a numeração e a ordem em que aparecem no verbete original. Assim, na imagem observam-se: (1) o lema composto de formas canônicas, iniciando o verbete, formatado em negrito; (2) a observação gramatical e etimológica, logo abaixo, em letra menor; (3) as acepções individualizadas, incluindo eventuais divisões internas em subgrupos de equivalentes (evidentes pela repetição da numeração); (4) os exemplos de uso fornecidos, agrupados por acepção (indicada pela numeração que precede cada grupo).

A Figura 30 ilustra a seção que abriga o conteúdo das quatro bases lexicais incluídas no protótipo, igualmente identificadas pelo nome do projeto original. Na amostra em tela, a aba selecionada exibe os dados fornecidos pelo *LiLa Lemma Bank*, referentes às classificações gramaticais e às variações lexicais e ortográficas.

As Figuras 31–34 ilustram a seção dedicada aos dados extraídos do *corpus*. Esta seção se divide em dois conjuntos de abas, um contendo informações sobre possíveis colocados, outro contendo exemplos de uso e informação sobre frequência. Os casos ilustrados são todos referentes ao lema verbal *facio*.

Assim, a Figura 31 ilustra o conteúdo da primeira aba (‘Possíveis colocados’), exibido em forma de tabela, que permite ao usuário ordenar os resultados por qualquer uma das colunas apresentadas – ou seja, quer por lema ou classe dos colocados, quer por um dos índices estatísticos. As demais abas trazem, cada uma, uma visualização gráfica do tipo nuvem de palavras dos possíveis colocados, divididos por classe de palavras; a Figura 32 ilustra esse recurso mostrando a nuvem de substantivos que provavelmente constituem colocações típicas com o verbo *facio*.

A Figura 33, por sua vez, oferece uma amostra de exemplos de uso extraídos do *corpus*, organizados em blocos contendo o texto latino, a indicação bibliográfica da fonte (segundo o

Figura 31 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (tabela de candidatos a colocados).



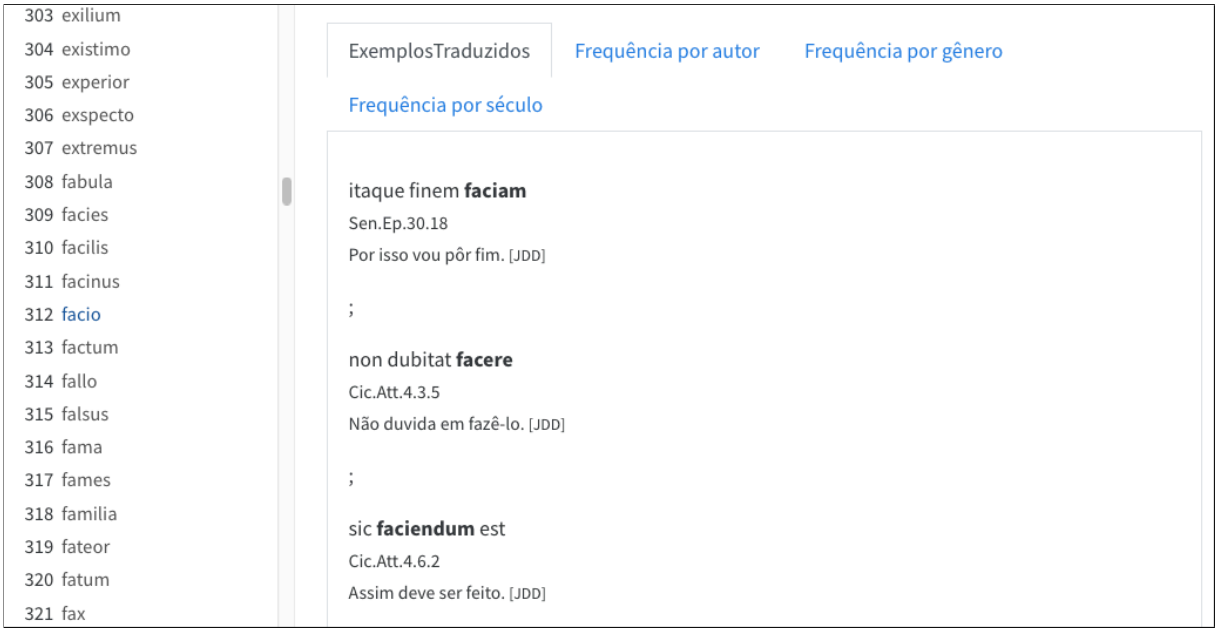
Fonte: <https://lucascdz.github.io/eLatim/facio.html>.

Figura 32 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (nuvem de substantivos candidatos a colocados).



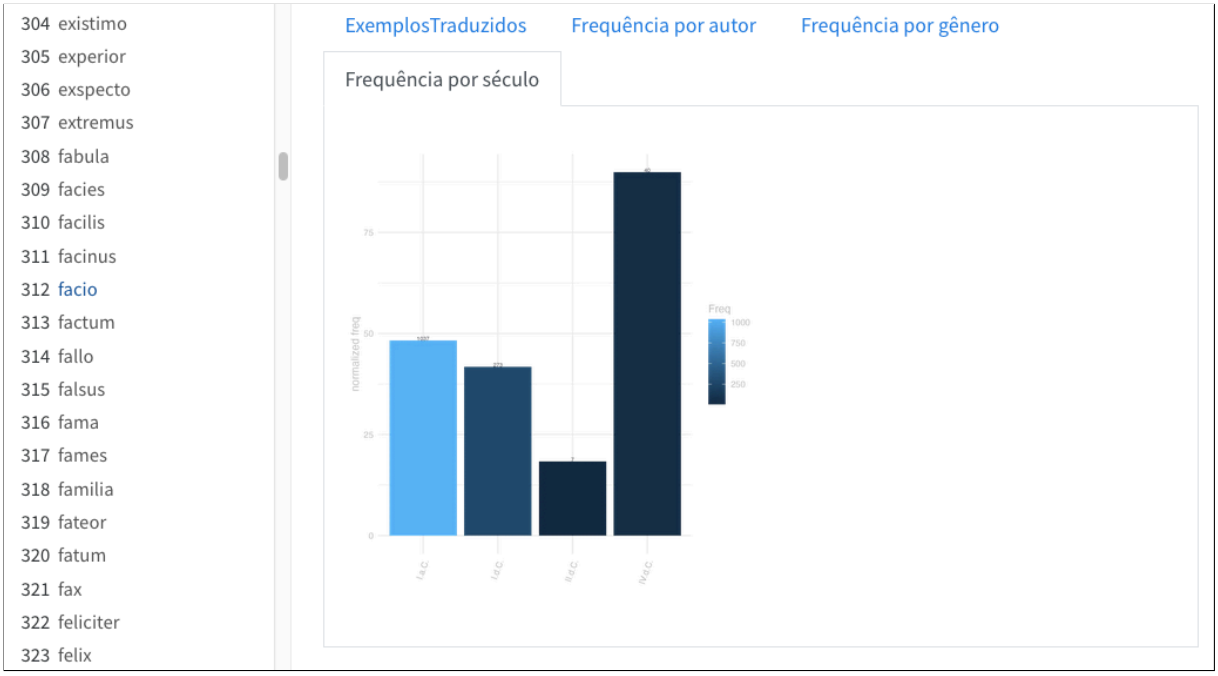
Fonte: <https://lucascdz.github.io/eLatim/facio.html>.

Figura 33 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (exemplos de uso).



Fonte: <https://lucascdz.github.io/eLatim/facio.html>.

Figura 34 – Reprodução de tela da página do **eLatim** com amostra de conteúdo extraído do *corpus* paralelo (gráfico de frequência).



Fonte: <https://lucascdz.github.io/eLatim/facio.html>.

formato convencional utilizado em filologia) e a tradução disponível. Por fim, as demais abas trazem cada uma uma visualização gráfica dos dados de frequência extraídos do *corpus*, com indicação tanto da frequência normalizada (relativa a 10 mil lemas e representada pela altura da barra) quanto da frequência absoluta (representada pela cor e indicada em legenda). A Figura 34 ilustra esse recurso com um gráfico da frequência do verbo *facio* por século.

O protótipo do **eLatim**, com todos os dados reunidos para os mil lemas mais frequentes do léxico latino, está publicado na plataforma *GitHub Pages* e pode ser acessado de forma gratuita no seguinte endereço: <https://lucascdz.github.io/eLatim>. O código-fonte utilizado para sua produção, bem como as planilhas de dados utilizadas, estão disponíveis no repositório GitHub do projeto, acessível em: <https://github.com/lucascdz/eLatim-dev>.

5.2 ELATIM-VALLEX: FERRAMENTA DE PESQUISA DE *CORPUS*

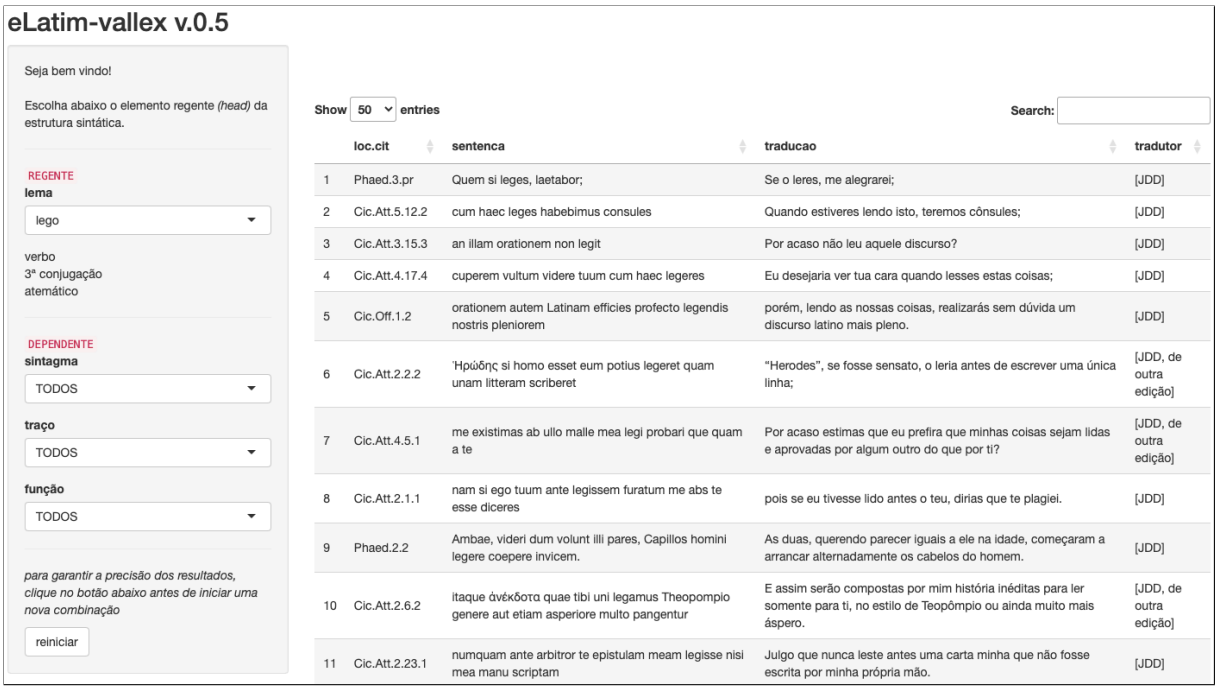
O **eLatim-vallex** é aplicativo de consulta de *corpus* construído em formato ShinyApp[4] de forma autônoma (i.e., sem a utilização de um DWS). A ideia é prover ao usuário um instrumento que lhe permita observar produções linguísticas autênticas e explorar as possibilidades de combinação sintático-semântica entre um lema e seus elementos dependentes de forma dinâmica, utilizando como base de dados o conteúdo do **translatinumCor-ptree**, o *corpus* paralelo com anotação sintático-semântica dos textos latinos compilado no capítulo anterior. Mesmo não sendo exatamente um dicionário bilíngue no sentido tradicional, o aplicativo cumpre uma função semelhante (senão de forma mais sofisticada), na medida em que provê sugestões de tradução para palavras e combinações de palavras inseridas em contexto.

O funcionamento do **eLatim-vallex** busca ser simples e intuitivo. A Figura 35 mostra a página inicial do aplicativo, configurada para exibir todas as sentenças em que ocorre o lema *lego* na posição de núcleo. A interface apresenta uma tela dividida verticalmente em duas partes. À esquerda, uma porção estreita, com fundo cinza, é composta de quatro seletores: o primeiro diz respeito à seleção do lema que ocupa a posição de núcleo (ou nó); os outros três definem propriedades sintático-semânticas para o elemento dependente (quais sejam, tipo de sintagma, traço semântico e função sintática). À direita, os resultados da combinação dos valores atribuídos a cada seletor são apresentados na forma de uma tabela com cinco colunas (ou sete, ver adiante) que apresentam, para cada sentença, respectivamente: uma identificação da sentença (baseada simplesmente em numeração); a referência bibliográfica no padrão filológico (coluna ‘loc.cit.’); o texto latino (coluna ‘sentence’); a tradução para o português (coluna ‘traducao’); e a autoria da tradução (coluna ‘tradutor’); esta, a propósito, é indicada por meio das iniciais do nome do tradutor, eventualmente acompanhadas de uma observação sobre discrepâncias no texto latino utilizado, servindo como alerta para possíveis incompatibilidades texto-tradução. Vale dizer que, por padrão, os resultados são organizados por número de palavras da sentença, em ordem crescente; não obstante, são passíveis de reorganização alfabética por qualquer uma das colunas exibidas, bastando clicar em seu título. Por fim, um campo de busca (‘search’), localizado no alto, à direita, e direcionado para todo o conteúdo da tabela, oferece um meio suplementar para filtrar os resultados obtidos.

O potencial de uso da ferramenta pode ser compreendido através de um exemplo de pesquisa. Mencionou-se, no Capítulo 1, que a estrutura argumental dos verbos é um lugar privile-

[4] <https://www.shinyapps.io/>.

Figura 35 – Reprodução de tela do **eLatim-vallex**, com exemplos de uso do verbo *lego*.

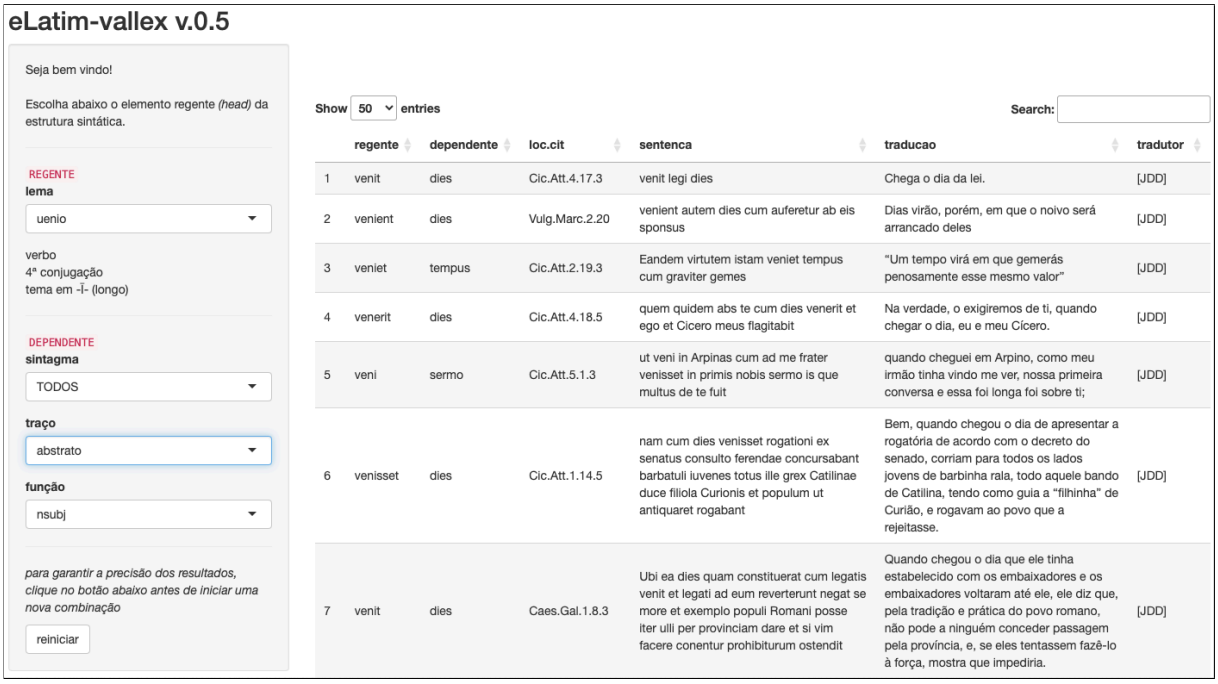


Fonte: <https://github.com/lucascdz/eLatim-vallex>.

giado de observação da conexão sintático-semântica. Um exemplo de aplicação desse modelo se encontra em Cabrillana (1997), que analisa os "marcos predicativos" do verbo latino *uenio* ('vir', 'chegar', etc.) a partir das ocorrências encontradas em um *corpus* historiográfico limitado às obras de Salústio e César. Com base na pressuposição de que um verbo de movimento deve incluir um argumento direcional, a autora distingue de início duas possibilidades de estrutura argumentativa: usos 'absolutos' (i.e., em que o argumento direcional não é expresso); e usos com o elemento direcional. Tal distinção se reflete no tipo de item lexical que ocupa a função de sujeito em uma e outra estrutura. Assim, no primeiro caso, a entidade a que se atribui o movimento é *inanimada* e, portanto, não pode exercer controle sobre a ação verbal; o significado obtido corresponde a uma concepção existencial do movimento, algo como '(chegar a) produzir-se', 'ocorrer' (e.g. *dies uenit*, 'chega o dia', *bellum uenit*, 'vem a guerra'). No segundo caso, o sujeito do movimento tende a ser uma entidade animada, capaz de assumir uma atividade de movimento. Em termos semânticos, pode-se dizer que "cuando el término que llena la primera posición tiene los rasgos /+animado/ y /+control/ hay una mayor tendencia a que aparezca el elemento direccional que en caso de la presencia en tal término de los rasgos contrarios (/-animado/, /-control/)" (Cabrillana, 1997, p.123).

Talvez a principal utilidade do **eLatim-vallex** seja prover meios práticos de observar, no *corpus* paralelo, exemplos traduzidos desse tipo de distinção. A Figura 36 mostra sete dos onze resultados encontrados pela ferramenta na atual versão do *corpus* quando se configura como lema "uenio", como traço semântico "abstrato" e como função sintática "nsubj" (que é o código para sujeito nominal no esquema da Universal Dependencies). Antes de tudo, é preciso notar que, em comparação com a tela inicial, aqui a tabela de resultados apresenta duas colunas adicionais,

Figura 36 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com sujeito abstrato.



Fonte: <https://github.com/lucascdz/eLatim-vallex>.

denominadas 'regente' e 'dependente'. Tais colunas são ativadas automaticamente pela seleção de configurações relacionadas às propriedades do dependente e exibem o resultado específico da combinação selecionada, extraído das formas que ocorrem nas respectivas sentenças; assim, no caso em tela, podem-se localizar rapidamente, entre os exemplos, as várias ocorrências em que a posição de sujeito é preenchida pelo lema *dies* (e.g. *uenit dies*, 'chega o dia'), bem como uma de *tempus* (*ueniet tempus*, 'um tempo virá…').

Quanto ao uso de *uenio* com sujeito animado, Cabrillana (1997) distingue cinco configurações morfossintáticas que a expressão do elemento direcional pode assumir. Três delas envolvem o caso acusativo, que pode ocorrer sozinho ou vir acompanhado das preposições *in* ou *ad*; outras duas se referem ao emprego do dativo e ao emprego de advérbios de lugar.

Segundo a autora, o uso do acusativo sem preposição é bem delimitado, ocorrendo exclusivamente com nomes próprios designando lugar menor (cidade ou ilha). Exemplos desse uso podem ser localizados no *corpus* pela seleção de "N(acc)" como tipo de sintagma e "iobj" como função sintática[5]; sete exemplos dentre os 23 encontrados pela ferramenta podem ser vistos na Figura 37.

Já o uso do acusativo preposicionado parece claramente vinculado ao valor que é produzido pela combinação da preposição com as características semânticas do lema empregado. De fato, a preposição *in* costuma ocorrer normalmente com lemas designadores de lugar (e.g. *ueni*

[5] No esquema das Universal Dependencies, a função iobj, i.e. o objeto indireto, é definida como "qualquer sintagma nominal que é argumento principal do verbo mas não é seu sujeito ou objeto direto" (no original, "the indirect object of a verb is any nominal phrase that is a core argument of the verb but is not its subject or direct object", cf. <https://universaldependencies.org/docs/u/dep/iobj.html>).

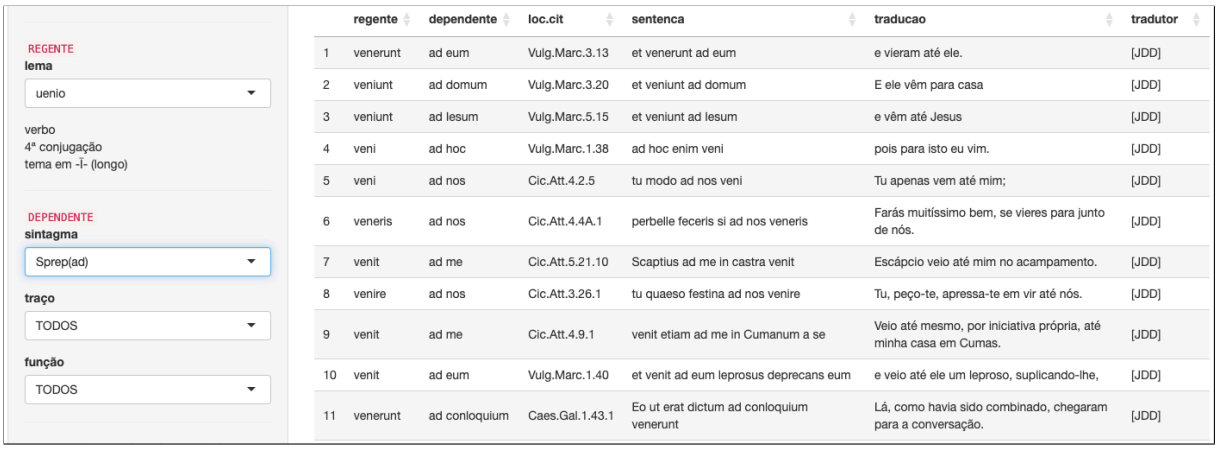
Figura 37 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento no acusativo.



	regente	dependente	loc.cit	sentenca	traducao	tradutor
1	veni	Brundisium	Cic.Att.4.1.4	Brundisium veni Nonis Sextilibus	Cheguei a Brundísio nas nonas de sextil [5 de agosto]	[JDD, de outra edição]
2	veni	Tarentum	Cic.Att.5.6.1	Tarentum veni a.d.xv Kal.Iunias	Cheguei em Tarento no décimo quinto dia antes das calendas de junho [18 de maio]	[JDD, de outra edição]
3	venire	Atuatucam	Caes.Gal.6.35.9	Tribus horis Atuatucam venire potestis	Em três horas podeis chegar em Aduátuca:	[JDD]
4	venimus	Delum	Cic.Att.5.12.1	sexto die Delum Athenis venimus	Em seis dias chegamos de Atenas a Delos.	[JDD]
5	veni	Laodiceam	Cic.Att.5.15.1	Laodiceam veni pridie Kal Sextilis	Cheguei em Laodicéia na véspera das calendas de sextil [31 de julho];	[JDD, de outra edição]
6	veni	Tarsum	Cic.Att.5.20.2	Tarsum veni a.d.III Nonas Octobris	Cheguei a Tarso no terceiro dia antes da nonas de outubro [5 de outubro].	[JDD, de outra edição]
7	veni	Brundisium	Cic.Att.3.7.1	Brundisium veni a. d. xiiii Kal. Maias.	Cheguei a Brundísio em 18 de abril.	[JDD]

Fonte: <https://github.com/lucascdz/eLatim-vallex>.

Figura 38 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por *ad*+acusativo.



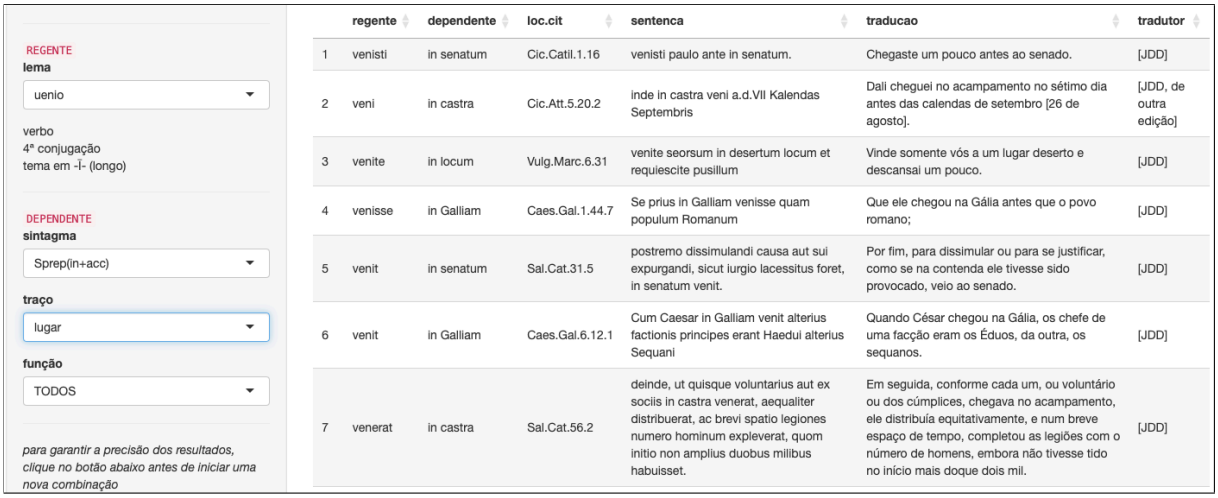
	regente	dependente	loc.cit	sentenca	traducao	tradutor
1	venerunt	ad eum	Vulg.Marc.3.13	et venerunt ad eum	e vieram até ele.	[JDD]
2	veniunt	ad domum	Vulg.Marc.3.20	et veniunt ad domum	E ele vêm para casa	[JDD]
3	veniunt	ad Iesum	Vulg.Marc.5.15	et veniunt ad Iesum	e vêm até Jesus	[JDD]
4	veni	ad hoc	Vulg.Marc.1.38	ad hoc enim veni	pois para isto eu vim.	[JDD]
5	veni	ad nos	Cic.Att.4.2.5	tu modo ad nos veni	Tu apenas vem até mim;	[JDD]
6	veneris	ad nos	Cic.Att.4.4A.1	perbelle feceris si ad nos veneris	Farás muitíssimo bem, se vieres para junto de nós.	[JDD]
7	venit	ad me	Cic.Att.5.21.10	Scaptius ad me in castra venit	Escápcio veio até mim no acampamento.	[JDD]
8	venire	ad nos	Cic.Att.3.26.1	tu quaeso festina ad nos venire	Tu, peço-te, apressa-te em vir até nós.	[JDD]
9	venit	ad me	Cic.Att.4.9.1	venit etiam ad me in Cumanum a se	Veio até mesmo, por iniciativa própria, até minha casa em Cumas.	[JDD]
10	venit	ad eum	Vulg.Marc.1.40	et venit ad eum leprosus deprecans eum	e veio até ele um leproso, suplicando-lhe,	[JDD]
11	venerunt	ad conloquium	Caes.Gal.1.43.1	Eo ut erat dictum ad conloquium venerunt	Lá, como havia sido combinado, chegaram para a conversação.	[JDD]

Fonte: <https://github.com/lucascdz/eLatim-vallex>.

in Trebulanum ‘cheguei na chácara de Trébula’), ao passo que a preposição *ad* é preferencialmente utilizada com lemas de traço humano (e.g. *ueni ad Pontium* ‘cheguei na casa de Pôncio’); quando é utilizada com lemas não humanos, parece possuir uma referência especialmente limitada, transitando da função de Direção para a Finalidade do movimento expresso pelo verbo (e.g. *ad comitia Romam ueniret* ‘[que] viesse a Roma para os comícios’).

A pesquisa no **eLatim-vallex** pode confirmar essas observações com dados quantitativos (dentro, é claro, das limitações do *corpus* descritas no Capítulo 3). A seleção de “Sprep(ad)” como tipo de sintagma dependente de *uenio* tem como resultado 64 exemplos, do quais 46 trazem sintagmas preposicionados constituídos por um substantivo ou pronome designativo de pessoa. A Figura 38 mostra os onze primeiros resultados, nos quais se vê a predominância de pronomes pessoais (e.g. *ad me*, *ad nos*, *ad eum*) e exemplos esporádicos de argumento de finalidade (*ad conloquium* ou *ad hoc*). A seleção do tipo de sintagma como “Sprep(in+acc)”, por sua vez, também resulta, coincidentemente, em 64 sentenças, das quais 43 contêm o referido sintagma constituído por um substantivo designador de lugar (seja um nome próprio como *in Galliam*, seja um nome comum como *in portum* ou *in senatum*); esse tipo de construção pode ser observada

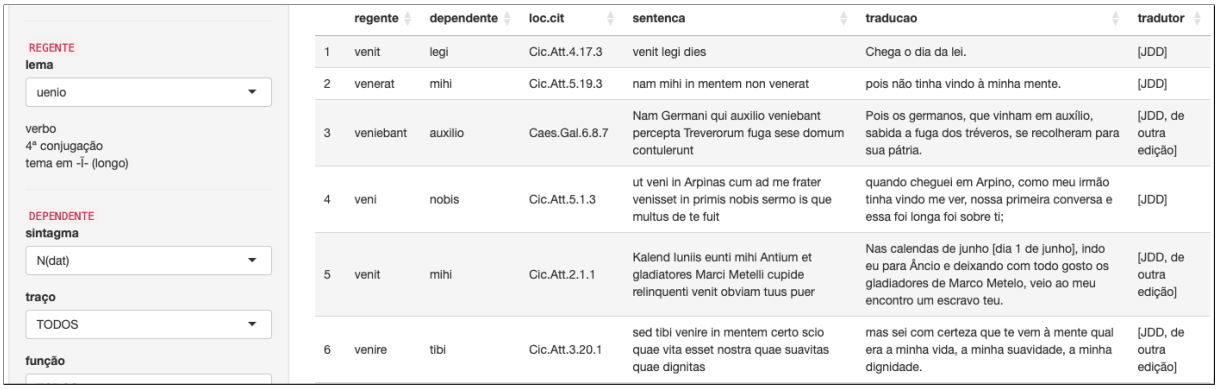
Figura 39 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por *in*+nome abstrato no acusativo.



	regente	dependente	loc.cit	sentenca	traducao	tradutor
1	venisti	in senatum	Cic.Catil.1.16	venisti paulo ante in senatum.	Chegaste um pouco antes ao senado.	[JDD]
2	veni	in castra	Cic.Att.5.20.2	inde in castra veni a.d.VII Kalendas Septembris	Dali cheguei no acampamento no sétimo dia antes das calendas de setembro [26 de agosto].	[JDD, de outra edição]
3	venite	in locum	Vulg.Marc.6.31	venite seorsum in desertum locum et requiescite pusillum	Vinde somente vós a um lugar deserto e descansai um pouco.	[JDD]
4	venisse	in Galliam	Caes.Gal.1.44.7	Se prius in Galliam venisse quam populum Romanum	Que ele chegou na Gália antes que o povo romano;	[JDD]
5	venit	in senatum	Sal.Cat.31.5	postremo dissimulandi causa aut sui expurgandi, sicut iurgio lacessitus foret, in senatum venit.	Por fim, para dissimular ou para se justificar, como se na contenda ele tivesse sido provocado, veio ao senado.	[JDD]
6	venit	in Galliam	Caes.Gal.6.12.1	Cum Caesar in Galliam venit alterius factionis principes erant Haedui alterius Sequani	Quando César chegou na Gália, os chefe de uma facção eram os Éduos, da outra, os sequanos.	[JDD]
7	venerat	in castra	Sal.Cat.56.2	deinde, ut quisque voluntarius aut ex sociis in castra venerat, aequaliter distribuerat, ac brevi spatio legiones numero hominum expleverat, quom initio non amplius duobus millibus habuisset.	Em seguida, conforme cada um, ou voluntário ou dos cúmplices, chegava no acampamento, ele distribuía equitativamente, e num breve espaço de tempo, completou as legiões com o número de homens, embora não tivesse tido no início mais doque dois mil.	[JDD]

Fonte: <https://github.com/lucascdz/eLatim-vallex>.

Figura 40 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por nome no dativo.



	regente	dependente	loc.cit	sentenca	traducao	tradutor
1	venit	legi	Cic.Att.4.17.3	venit legi dies	Chega o dia da lei.	[JDD]
2	venerat	mihi	Cic.Att.5.19.3	nam mihi in mentem non venerat	pois não tinha vindo à minha mente.	[JDD]
3	veniebant	auxilio	Caes.Gal.6.8.7	Nam Germani qui auxilio veniebant percepta Treverorum fuga sese domum contulerunt	Pois os germanos, que vinham em auxílio, sabida a fuga dos tréveros, se recolheram para sua pátria.	[JDD, de outra edição]
4	veni	nobis	Cic.Att.5.1.3	ut veni in Arpinas cum ad me frater venisset in primis nobis sermo is que multus de te fuit	quando cheguei em Arpino, como meu irmão tinha vindo me ver, nossa primeira conversa e essa foi longa foi sobre ti;	[JDD]
5	venit	mihi	Cic.Att.2.1.1	Kalend Iuniis eunti mihi Antium et gladiatores Marci Metelli cupide relinquenti venit obviam tuus puer	Nas calendas de junho [dia 1 de junho], indo eu para Âncio e deixando com todo gosto os gladiadores de Marco Metelo, veio ao meu encontro um escravo teu.	[JDD, de outra edição]
6	venire	tibi	Cic.Att.3.20.1	sed tibi venire in mentem certo scio quae vita esset nostra quae suavitas quae dignitas	mas sei com certeza que te vem à mente qual era a minha vida, a minha suavidade, a minha dignidade.	[JDD, de outra edição]

Fonte: <https://github.com/lucascdz/eLatim-vallex>.

mediante a seleção de "lugar" como traço semântico, como mostra a Figura 39.

Quanto ao uso do dativo, por fim, a autora verifica que boa parte dos usos corresponde à expressão de finalidade, outra parte se dividindo entre as designações de destinatário, beneficiário e receptor. Independente disso, o sentido do verbo varia de concreto a abstrato conforme o lema que ocupa a posição de sujeito tenha os traços /+humano//+controle/ (e.g. *aliquos ab tergo hostibus uenturos* 'alguns chegariam aos inimigos pelas costas') ou /-humano//-controle/ (e.g. *illi et gloriam et regnum uenturum* 'a glória e o reino chegariam para ele'), respectivamente. Ocorrências dessas construções no *corpus* podem ser observadas no **eLatim-vallex** mediante a seleção de "N(dat)" como tipo de sintagma; na atual versão do **translatinumCor-ptree**, essa busca resulta em 13 ocorrências, com exemplos tanto de dativo de Finalidade (*ueniebant auxilio*, 'vinham em socorro') quanto de dativo indicando os demais papéis semânticos, como o de Receptor em *nuntii nobis tristes uenerant* ('notícias tristes nos chegaram'); a Figura 40 mostra outros casos semelhantes.

Uma outra possibilidade, não explorada por Cabrillana (1997) e mencionada por Pinkster

Figura 41 – Reprodução de tela do **eLatim-vallex**, com exemplo de uso do verbo *uenio* com argumento constituído por *in*+nome abstrato no acusativo.



Fonte: <https://github.com/lucascdz/eLatim-vallex>.

(2015, p.75), é o uso de *uenio* como verbo suporte em expressões idiomáticas. Nessa estrutura há uma combinação de duas formas de argumento vistas acima, a saber, acusativo precedido de *in* e dativo. Também aqui as características do sujeito são determinantes. De um lado, lemas com traços /+humano/ tanto na posição de sujeito (nominativo) e quanto na posição de Receptor (dativo) participam de construções em que *in* rege a forma acusativa de um lema abstrato, designando uma categoria em que o sujeito 'vem a ser incluído' (e.g. *uenisse Germanis in amicitiam* 'tinha sido incluído na aliança com os germanos' ou '…tornara-se aliado dos germanos') ou um sentimento ou situação que o sujeito passa a experimentar (e.g. *uenire in spem* 'tornar-se esperançoso', *uenire in iudicium* 'ser levado a julgamento').[6] Ocorrências desse uso metafórico do verbo *uenio* podem ser observadas mediante a seleção de "Sprep(in+acc)" como tipo de sintagma e "abstrato" como traço semântico, como mostra a Figura 41.

Em suma, o **eLatim-vallex** que favorece a pesquisa de estruturas argumentais e outros tipos de relações de dependência sintática entre elementos lexicais. Vale observar que, se o exemplo aqui apresentado corresponde ao um tipo de pesquisa fundamentada em *corpus* (ou *corpus-based*, no sentido de que os critérios de seleção de exemplos dependem de hipóteses previamente formuladas), é perfeitamente possível usar a ferramenta para realizar pesquisas guiadas pelos próprios resultados obtidos (i.e. uma pesquisa do tipo *corpus-driven*, em que a observação do *corpus* dá ensejo à formulação de hipóteses).

Para utilizar o programa, o usuário precisa realizar uma instalação local em sua máquina, uma vez que, até o momento, não foi possível viabilizar uma instalação para acesso *online* dentro do critério de custo zero que tem orientado o presente trabalho. Não obstante, a ferramenta possui compatibilidade com qualquer dispositivo e sistema operacional. O protótipo do **eLatim-vallex** apresentado, bem como as instruções para instalação, estão disponíveis para *download* gratuito no repositório GitHub do autor.[7]

[6] Pinkster (2015, p.75) nota que algumas dessas expressões podem ocorrer com outros verbos tipicamente de suporte, como *sum* e *habeo*.

[7] Cf. <https://github.com/lucascdz/eLatim-vallex>.

6 CONCLUSÃO

Os protótipos de dicionário bilíngue (**eLatim**) e de ferramenta de consulta de *corpus* (**eLatim-vallex**) apresentados no capítulo anterior representam o término de um percurso de pesquisa que buscou constituir uma base teórico-metodológica que desse condições de suprir a carência de um dicionário digital latim-português acessível e atualizado com as novas tecnologias da informação e comunicação.

O trajeto passou por quatro etapas: planejamento do dicionário, levantamento de fontes, extração e processamento de dados e organização de material para consulta humana.

Na primeira, tratou-se de conhecer as principais decisões que a construção de dicionários requer dos lexicógrafos, relativas ao tipo de dicionário almejado, bem como às fontes, qualidades e modos de organização das informações apresentadas. Sabe-se que boa parte dessas decisões depende do conhecimento das situações e necessidades de consulta lexicográfica apresentadas pelos usuários, que podem ser relativas a informações sobre a forma, o significado e o uso dos itens lexicais.

Um esforço especial foi dedicado ao levantamento das possíveis fontes de informação a serem utilizadas, perscrutando dois milênios de produção de repertórios lexicais a respeito do latim, desde aqueles registrados em suportes manuscritos até os mais recentes recursos lexicais produzidos para leitura por computadores. No meio do caminho, toda uma tradição de dicionários latim-português publicados de forma impressa entre os séculos XVI e XX, muitos dos quais eram (e ainda são) desconhecidos de boa parte de estudantes e acadêmicos dedicados aos Estudos Clássicos.

A grande quantidade de material encontrado exigiu um esforço igualmente intenso de transformação e adaptação, a fim de estabelecer um padrão de estruturação que viabilizasse a criação uma base de dados lexicais unificada, ainda que constituída de dados heterogêneos, extraídos tanto de dicionários tradicionais e bases computacionais quanto de um *corpus* paralelo latim-português de textos clássicos. Esse trabalho utilizou modelos baseados em ontologias (e.g. *OntoLex*, *LiLa Knowledge Base*) e foi auxiliado pelo uso de um aplicativo de escrita de dicionários (DWS) de código aberto e uso gratuito desenvolvido de forma concomitante com essa pesquisa (*lexicographR*). Aliás, importa destacar que não apenas os modelos e o aplicativo, mas também as fontes de conteúdo utilizadas – quer dizer, os *corpora*, os dicionários tradicionais digitalizados e as demais bases lexicais – foram escolhidos tendo como um dos principais critérios a gratuidade e a publicação sob licença aberta, seguindo uma filosofia de divulgação e construção compartilhada do conhecimento científico acumulado. Por conseguinte, as bases de dados[1] e os protótipos de interface de consulta[2] apresentados como resultado de pesquisa foram

[1] Cf. <https://github.com/lucascdz/latinolusitanicum-db> e <https://github.com/lucascdz/translatinumCor-pt>.
[2] Cf. <https://github.com/lucascdz/eLatim> e <https://github.com/lucascdz/eLatim-vallex>.

desenvolvidos e publicados sob o mesmo tipo de licença, como forma de retribuição à sociedade acadêmica internacional pelos recursos deixados à disposição e imprescindíveis para a realização deste trabalho.

Ao mesmo tempo que representam o término, os dois protótipos apresentados são apenas o início de um caminho mais longo rumo à produção de dicionários e ferramentas lexicais cada vez abrangentes e afinados com as necessidades específicas dos usuários reais. No horizonte dessa nova trajetória estão as diversas possibilidades não só de aprimoramento contínuo da interface e do conteúdo atualmente disponível, mas também de inclusão de novos materiais e fontes de dados.

Para começar, a nomenclatura pode ser ampliada para além das mil palavras latinas mais frequentes. Isso depende de adotar uma lista frequencial baseada em um *corpus*, bem como de encontrar soluções técnicas apropriadas para o armazenamento e processamento de dezenas de milhares de verbetes.

O conteúdo dos verbetes, por sua vez, pode ser ampliado de diversas maneiras. No caso dos dicionários tradicionais, há um bom número de itens bibliográficos esperando serem digitalizados e convertidos em dados estruturados. Dentre eles, o dicionário de Saraiva (1927) tem prioridade devido a sua nomenclatura abrangente e sua extraordinária quantidade de exemplos lexicográficos oferecidos. A propósito, a dimensão desse *corpus* justificaria a criação de uma base de dados especializada em fraseologia – provisoriamente denominado **exemplaElocutionum-db**[3] – mediante um trabalho de anotação sintático-semântica das frases latinas nos mesmos moldes utilizados no *corpus* paralelo latim-português (**translatinumCor-ptree**). Desse modo, essa nova fonte de dados poderia ser explorada através do **eLatim-vallex** para fornecer um conjunto valioso de probabilidades de ancoragem fraseológica de significado, como pressupõe Hanks (2022, p.426–7).[4] Quanto aos possíveis aprimoramentos, um dos mais úteis seria a inclusão de um identificador semântico das acepções, que permitisse não só associar as diferentes formas de descrever o significado apresentada pelos diversos dicionários entre si, mas também associar os exemplos extraídos de *corpus* com cada acepção específica (nesse caso, obviamente, desde que o *corpus* seja semanticamente anotado segundo o mesmo padrão).

Também os recursos computacionais podem ser ampliados pela inclusão de outras bases lexicais existentes e ainda por vir, considerando a vitalidade de produções nesse campo de pesquisa; por ora, caberia uma revisão dos dados apresentados, na medida em que muitos deles foram extraídos por processos automatizados e estão sujeitos a erros de interpretação pela máquina.

O *corpus* paralelo pode ser ampliado em várias dimensões: quer pela inclusão de tradução paralela aos textos dotados de anotação sintático-semântica; quer pela inclusão de anotação para os textos atualmente apenas lematizados; quer ainda pela inclusão de novos textos acompanha-

[3] Cf. <https://github.com/lucascdz/exemplaElocutionum-db>.

[4] O uso de exemplos de dicionário como *corpus* de pesquisa é discutido brevemente por Kennedy (1998, pp.35–36) e mais detidamente por Hoffmann (2004) e Brinton, Dollinger e Fee (2012), entre outros.

dos tanto de anotação quanto de tradução. Como foi dito acima, o acréscimo de um identificador de significado ao esquema de anotação tornaria possível associar o conteúdo extraído do *corpus* às acepções de dicionário, permitindo ao dicionário digital fornecer não apenas exemplos de uso semanticamente definidos mas também os padrões sintático-semânticos típicos de cada significado – considerando como uma das funções prioritárias das citações textuais incluídas nos verbetes a de "ilustrar o uso (sintagmático) das unidades lexicais" (Heid, 2008, p.142). Outro aprimoramento possível, com importante repercussão, seria aumentar a granularidade do alinhamento das traduções, do nível das sentenças para o nível das palavras, viabilizando assim a extração de equivalentes de tradução "reais", por assim dizer, no sentido de que são resultado do trabalho de um tradutor sobre o texto em contexto.

Por último, mas certamente não menos importante, é a possibilidade de submeter os dois protótipos a protocolos de pesquisa de uso de dicionário, que permitiria definir com mais exatidão os tipos de informação e as formas de apresentação mais adequadas a cada perfil de usuário em potencial. Na ausência desse parâmetro específico, tentou-se criar interfaces de consulta e visualização dos dados com base em padrões mínimos de adequação para a leitura humana. Isso se revela na tentativa de distribuir o conteúdo dos dicionários de forma mais arejada, menos densa em relação às práticas da lexicografia impressa, no uso de critérios claros de segmentação do verbete (tipo de fonte) e no uso de elementos gráficos para facilitar a apreensão da informação. Sem dúvida, há outras possibilidades a se explorar para aprimorar o *design* gráfico e a experiência do usuário, que dependem de um conhecimento mais profundo das situações e das necessidades de informação de quem utiliza dicionários latim-português.

Tantas perspectivas de aprimoramento só são possíveis na medida em que uma estrutura lexicográfica consolidada de processamento e organização dos dados lexicais tenha sido concebida e produzida. Embora a quantidade de dados processados até o momento seja modesta, os resultados obtidos parecem suficientes para comprovar a eficácia da metodologia aqui desenvolvida, uma vez que define padrões e procedimentos suficientes para a inclusão futura de informação lexical obtida de quaisquer tipos de fontes. Este é, sem dúvida, o principal resultado deste trabalho de pesquisa.

REFERÊNCIAS

ABEL, A. Dictionary writing systems and beyond. In: GRANGER, S.; PAQUOT, M. (Ed.). *Electronic Lexicography*. [S.l.]: Oxford University Press, 2012. Citado 2 vezes nas páginas 165 e 204.

ACERO DURÁNTEZ, I. Lexicografía plurilingüe de los siglos xvi y xvii. *Philologia Hispalensis*, n. 22, p. 7–36, 2008. Citado na página 101.

ADAMS, J. N. *The Latin sexual vocabulary*. Baltimore: Johns Hopkins University Press, 1990. Citado na página 72.

AIJMER, K. Parallel and comparable corpora. In: LÜDELING, A.; KYTÖ, M. (Ed.). *Corpus Linguistics*: An international handbook. Berlin/New York: Walter de Gruyter, 2008. v. 1, p. 275–92. Citado 2 vezes nas páginas 186 e 188.

ALMAS, B. Perseids: Experimenting with infrastructure for creating and sharing research data in the digital humanities. *Data Sci. J.*, v. 16, p. 19, 2017. Disponível em: <https://doi.org/10.5334/dsj-2017-019>. Citado na página 18.

ALMEIDA, J. M. de. O primeiro lexicógrafo português da Língua Latina: Jerónimo Cardoso. *Euphrosyne*, v. 2 (1st series), p. 139–152, 1959. Citado 3 vezes nas páginas 105, 106 e 110.

ALMEIDA, J. M. de. Lexicógrafos da Língua Latina em Portugal [pt.2]. *Revista de Guimarães*, v. 79, n. 3–4, p. 193–226, 1969. Citado 5 vezes nas páginas 127, 130, 133, 134 e 136.

ALUÍSIO, S. M.; ALMEIDA, G. M. de B. O que é e como se constrói um corpus?: Lições aprendidas na compilação devários corpora para pesquisa lingüística. *Calidoscópio*, Universidade do Vale do Rio dos Sinos, v. 4, n. 3, p. 156–178, 2006. Citado na página 186.

ALVES, D. As humanidades digitais como uma comunidade de práticas dentro do formalismo académico: dos exemplos internacionais ao caso português. *Ler história*, n. 69, p. 91–103, 2016. Citado na página 18.

AMARANTE, J. *Latinitas: uma introdução à língua latina através dos textos*. [S.l.]: SciELO-EDUFBA, 2018. Citado na página 50.

AMBLER, S. W. *Agile Modeling*: Effective practices for extreme programming and the unified process. New York: John Wiley & Sons, 2002. Citado na página 165.

AMOS, T. L. *The Fundo Alcobaça of the Biblioteca Nacional, Lisbon, vols.1–3*. Collegeville, Minnesota: Hill Monastic Manuscript Library, 1988–1990. Citado na página 95.

ANDRADE, W. O.; NEVES, D. A. de B. Análise documental e representação da informação: aportes teóricos à utilização simultânea visando a recuperação da informação em arquivos. In: FUJITA, M. S. L.; NEVES, D. A. de B.; DAL'EVEDOVE, P. R. (Ed.). *Leitura documentária*: estudos avançados para a indexação. Marília/São Paulo: Oficina Universitária/Cultura Acadêmica, 2017. p. 93–112. Citado na página 174.

ARAÚJO, P. H. L. *As letras ramistas em dois roteiros de viagem do século VIII*. Dissertação (Mestrado) — Universidade de São Paulo, 2007. Citado na página 50.

AREVALO, F. (Ed.). *S. Isidori Hispalensis Episcopi Opera Omnia*: Tomus v. libri differentiarum et opera biblica. Romae: Apud Antonium Fulgonium, 1802. Citado na página 88.

ARRUDA, F. E. C. *Microestrutura de um dicionário de latim para alunos iniciantes*. Tese (Doutorado) — Universidade Federal do Ceará, 2019. Citado na página 20.

ASSUNÇÃO, C.; FERNANDES, G. Amaro de Roboredo, gramático e pedagogo português seiscentista, pioneiro na didáctica das línguas e nos estudos linguísticos. In: ROBOREDO, Amaro. *Methodo grammatical para todas as linguas*. Vila Real: Centro de Estudos em Letras da Universidade de Trás-os-Montes e Alto Douro, 2007. Edição facsimilada. Citado na página 117.

ATKINS, B. S.; RUNDELL, M. *The Oxford guide to practical lexicography*. [S.l.]: Oxford University Press, 2008. Citado 11 vezes nas páginas 34, 37, 45, 49, 51, 53, 54, 62, 68, 74 e 75.

ATKINS, S. Theoretical lexicography and its relation to dictionary-making. In: FONTENELLE, T. (Ed.). *Practical lexicography*: a reader. Oxford: Oxford University Press, 2008. p. 31–50. Citado 7 vezes nas páginas 25, 26, 27, 34, 37, 43 e 44.

ATKINS, S.; CLEAR, J.; OSTLER, N. Corpus design criteria. *Journal of Literary and Linguistic Computing*, Oxford University Press, v. 7, n. 1, p. 1–33, 1992. Citado 3 vezes nas páginas 185, 187 e 188.

AZEVEDO, F. d. *Pequeno dicionário latino-português*. São Paulo: Companhia Editora Nacional, 1944. Citado na página 146.

BAKER, P.; HARDIE, A.; MCENERY, T. *Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2006. Citado 2 vezes nas páginas 166 e 191.

BALBUS, J. *Catholicon*. Mainz: Guttenberg, 1460. Reimp. Westmead: Gregg International Publishers Limited, 1971. Citado na página 93.

BAMMAN, D.; CRANE, G. The Latin dependency treebank in a cultural heritage digital library. In: SPORLEDER, C.; BOSCH, A. van den; GROVER, C. (Ed.). *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007)*. Prague, Czech Republic: Association for Computational Linguistics, 2007. p. 33–40. Disponível em: <https://aclanthology.org/W07-0905>. Citado na página 189.

BAMMAN, D.; CRANE, G. Computational linguistics and classical lexicography. *Digital Humanities Quarterly*, v. 3, n. 1, 2009. Citado na página 162.

BAMMAN, D.; CRANE, G. Linguística de corpus, treebanks e a reinvenção da filologia. In: FERREIRA, A. D. (Ed.). *Introdução aos textos clássicos na era digital do terceiro milênio*. Araraquara, SP: Letraria, 2015. p. 19–32. Citado 2 vezes nas páginas 18 e 19.

BAÑOS BAÑOS, J. M. Presentación: estructuras de complementación en griego y latín. In: BAÑOS, J. M. B. et al. (Ed.). *Praedicativa*: complementación en griego y latín. Santiago de Compostela: Servicio de Publicaciones de la Universidade de Santiago de Compostela, 2003. Citado 4 vezes nas páginas 64, 71, 72 e 75.

BAÑOS BAÑOS, J. M. *Sintaxis del latín clásico*. Madrid: Liceus, 2009. Citado 2 vezes nas páginas 64 e 72.

BARATIN, M.; DESBORDES, F. *L'analyse linguistique dans l'Antiquité classique*: Tome i: Les théories. Paris: Klincksieck, 1984. Citado na página 80.

BARNEY, S. A. et al. *The Etymologies of Isidore of Seville*. Cambridge: Cambridge University Press, 2006. Citado 4 vezes nas páginas 86, 87, 88 e 89.

BARWICK, K. (Ed.). *Flavii Sosipatri Charisii Artis Grammaticae Libri V*. Lipsiae: Teubner, 1964. Citado 2 vezes nas páginas 84 e 85.

BATHE, W. (Ed.). *Ianua linguarum*: sive modus maxime accommodatus, quo patefit ad omnes linguas intellegendas. Salmanticae: Apud Franciscum de Cea Tesa, 1611. Disponível em: <http://archive.org/details/bub_gb_sXeXLiNMxEwC>. Citado 2 vezes nas páginas 116 e 117.

BECK, C. A copious and critical latin-english lexicon... *The North American Review*, v. 73, n. 152, p. 71–90, 1851. Disponível em: <https://www.jstor.org/stable/40794541>. Acesso em: 26.mai.2021. Citado 2 vezes nas páginas 137 e 139.

BENEDUZI, R. *Colocações substantivo+ adjetivo: propostas para sua identificação e tratamento lexicográfico em dicionários ativos português-espanhol*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008. Citado na página 65.

BERGENHOLTZ, H.; TARP, S.; DUVÅ, G. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: J. Benjamins, 1995. (Benjamins translation library). ISBN 9789027216120. Citado 2 vezes nas páginas 32 e 57.

BERNERS-LEE, J. H. T.; LASSILA, O. The semantic web. *Scientific American*, maio 2001. Citado 2 vezes nas páginas 32 e 155.

BERNERS-LEE, T. *Linked Data*. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 11.jan.2025. Citado 2 vezes nas páginas 32 e 156.

BERNERS-LEE, T. *Levels of Abstraction: Net, Web, Graph*. 2007. Disponível em: <https://www.w3.org/DesignIssues/Abstractions.html>. Acesso em: 11.jan.2025. Citado na página 156.

BIBER, D. Representativeness in corpus design. *Literary and linguistic computing*, Oxford University Press, v. 8, n. 4, p. 243–57, 1993. Citado na página 186.

BIDERMAN, M. T. C. Dimensões da palavra. *Filologia e linguística portuguesa*, Universidade de São Paulo, v. 2, n. 1, p. 81–118, 1998. Citado na página 56.

BIDERMAN, M. T. C. *Teoria Linguística: teoria lexical e linguística computacional*. São Paulo: Martins Fontes, 2001. Citado 5 vezes nas páginas 21, 34, 37, 39 e 75.

BLACKWELL, C.; CRANE, G. Cyberinfrastructure, the scaife digital library and classics in a digital age. *Digital Humanities Quarterly*, The Alliance of Digital Humanities Organizations and The Association for Computers and the Humanities, v. 3, n. 1, 2009. Disponível em: <https://www.digitalhumanities.org/dhq/vol/3/1/000035/000035.html>. Citado na página 18.

BOGAARDS, P. Uses and users of dictionaries. In: STERKENBURG, P. (Ed.). *A Practical Guide to Lexicography*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 1996. p. 26–33. Citado 7 vezes nas páginas 29, 30, 31, 37, 41, 45 e 71.

BOISSON, C.; KIRTCHUK, P.; BÉJOINT, H. Aux origines de la lexicographie: les premiers dictionnaires monolingues et bilingues. *International Journal of Lexicography*, Oxford University Press, v. 4, n. 4, 1991. Citado na página 79.

BORBA, F. d. S. Propriedades sintáticas, semânticas e pragmáticas do léxico. *Revista (Con)textos linguísticos*, v. 1, n. 1, 2007. Citado na página 74.

BORBA, F. da S. *Dicionário gramatical de verbos do Português contemporâneo do Brasil*. São Paulo: Editora Unesp, 1992. Citado 4 vezes nas páginas 62, 64, 70 e 75.

BORBA, F. da S. *Uma gramática de valências para o português*. São Paulo: Ática, 1996. Citado 2 vezes nas páginas 62 e 64.

BORBA, F. da S. *Organização de dicionários: uma introdução à lexicografia*. São Paulo: Editora Unesp, 2003. Citado 8 vezes nas páginas 38, 39, 49, 62, 63, 65, 66 e 73.

BORGES, A. M. La reforma pombalina de la enseñanza: de la prosodia de bento pereira al parvum lexicon de pedro da fonseca. In: *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*. [S.l.: s.n.], 2008. p. 1081–1095. Citado 3 vezes nas páginas 127, 128 e 132.

BORGES, A. M. de A. *Pedro José da Fonseca e a sua obra lexicográfica*. Tese (Doutorado) — Universidade de Aveiro, 2011. Citado 6 vezes nas páginas 128, 130, 132, 133, 134 e 136.

BOSCHETTI, F. Semantic analysis and thematic annotation. In: BERTI, M. (Ed.). *Digital Classical Philology*. Berlin, Boston: De Gruyter Saur, 2019. p. 321–340. Disponível em: <https://doi.org/10.1515/9783110599572-018>. Citado 2 vezes nas páginas 158 e 159.

BOSQUE-GIL, J.; GRACIA, J. (Ed.). *The OntoLex Lemon Lexicography Module: Final Community Group Report*. Cambridge, MA, 2019. Disponível em: <https://www.w3.org/2019/09/lexicog/>. Citado na página 177.

BOTELHO, J. M. *Pequeno dicionário de latim-português*. Rio de Janeiro: Botelho, 2014. Citado na página 152.

BRINTON, L. J.; DOLLINGER, S.; FEE, M. Balanced corpora and quotation databases: Taking shortcuts or expanding methodological scope. *Studies in Variation, Contacts and Change in English*, v. 10, p. s.p., 2012. Disponível em: <https://varieng.helsinki.fi/series/volumes/10/brinton_dollinger_fee/>. Acesso em: 08.abr.2025. Citado na página 217.

BRITANNICA, T. E. o. E. Latin alphabet. In: ____. *Encyclopedia Britannica*. [s.n.], 2024. Disponível em: <https://www.britannica.com/topic/Latin-alphabet>. Acesso em: 29 out. 2024. Citado na página 79.

BRUNO, H. Latim e formação linguística. In: LIMA, A. D. (Ed.). *Latim, da fala à língua*. Araraquara: Faculdade de Ciências e Letras da UNESP, 1992. p. 39–43. Citado na página 29.

BUGUEÑO MIRANDA, F. V. Para uma taxonomia de paráfrases explanatórias. *ALFA: Revista de Linguística*, v. 53, n. 1, p. 243–260, 2009. Citado 3 vezes nas páginas 54, 55 e 56.

BURKE, P. Problemas causados por gutenberg: a explosão da informação nos primórdios da europa moderna. *Estudos avançados*, SciELO Brasil, v. 16, p. 173–185, 2002. Citado na página 98.

BUSARELLO, R. *Dicionário Básico Latino-Português*. Florianópolis: Editora da UFSC, 1988. Citado na página 152.

BUSDRAGHI, P. et al. (Ed.). *Osberno: Derivazioni*. Firenze: Centro italiano di studi sull'alto medioevo / Spoleto, 1996. Citado 2 vezes nas páginas 91 e 92.

BUSSARELLO, R. *Contribuição linguística do modelo dependencial de Lucien Tesnière para o estudo do latim em cursos de Letras*. Dissertação (Mestrado) — Pontificia Universidade Catolica do Rio Grande do Sul, 1981. Citado na página 62.

BUYSE, K. "Por el amor del estudiante": hacia una lexicografía integradora. *Cadernos de Tradução*, Florianópolis, v. 2, n. 18, p. 195–213, 2006. Citado na página 20.

CABRAL, M. de P. *Magnum Lexicon Latinum et Lusitanicum*. Olyssipone: Typis Simonis Thaddaei Ferreira, 1802. Citado 3 vezes nas páginas 133, 134 e 135.

CABRAL, M. de P.; RAMALHO, J. A.; FERREIRA, M. J. *Magnum Lexicon Latinum et Lusitanicum*. Parisiis: Apud Vam. J. P. Aillaud, Guillard et Sem., 1867. Citado na página 134.

CABRILLANA, C. Complementos direccionales en el marco predicativo de venio. *Minerva. Revista de Filología Clásica*, n. 11, p. 117–137, 1997. Citado 5 vezes nas páginas 64, 72, 211, 212 e 214.

CALEPINUS, A. *Dictionarium*. Rhegii lingobardiae: Dionysius Berthochius, 1502. Disponível em: <http://archive.org/details/bub_gb_v-JKAAAAcAAJ>. Acesso em: 25 nov. 2024. Citado na página 101.

CANFORA, L. *Storia della letteratura greca*. Roma/Bari: Laterza, 1986. Citado na página 98.

CARDOSO, J. *Dictionarium latino lusitanicum & vice versa lusitanico latinu[m]*. Coimbra: João de Barreira, 1570. Disponível em: <http://purl.pt/14265>. Acesso em: 18.7.2023. Citado 8 vezes nas páginas 106, 107, 108, 162, 172, 181, 182 e 183.

CARDOSO, J. Dictionarium latino lusitanicum & vice versa lusitanico latinu[m]. In: VERDELHO, T. dos S.; SILVESTRE, J. P. M. (Ed.). *Corpus Lexicográfico do Português*. Aveiro, Portugal: Universidade de Aveiro and Centro de Linguística da Universidade de Lisboa, 2002. Disponível em: <http://clp.dlc.ua.pt/DICIweb/default.asp?url=Obras&Livro=5&Edicao=5>. Acesso em: 18.7.2023. Citado na página 107.

CARDOSO, Z. d. A. *A literatura latina*. Porto Alegre: Mercado Aberto, 1989. Citado na página 191.

CARTER, H. H. A fourteenth-century latin–old portuguese verb dictionary. *Romance Philology*, Brepols / University of California Press, v. 6, n. 2/3, p. 71–103, 1952. Citado 2 vezes nas páginas 95 e 96.

CAVALLO, G.; CHARTIER, R. Introducción. In: CAVALLO, G.; CHARTIER, R. (Ed.). *Historia de la lectura en el mundo occidental*. Madrid: Taurus, 2004. Citado na página 80.

CECCHINI, F. et al. Udante: First steps towards the universal dependencies treebank of dante's latin works. In: MONTI, J.; DELL'ORLETTA, F.; TAMBURINI, F. (Ed.). *Proceedings of the Seventh Italian Conference on Computational Linguistics*. [S.l.]: CEUR Workshop Proceedings, 2020. p. 99–105. Citado na página 190.

CECCHINI, F. M. et al. Challenges in converting the *Index Thomisticus* treebank into universal dependencies. In: *Proceedings of the Universal Dependencies Workshop 2018 (UDW 2018)*. Brussels, Belgium: [s.n.], 2018. Citado na página 190.

CELANO, G. G. Lemmatization and morphological analysis for the latin dependency treebank. *Studi e Saggi Linguistici*, v. 58, n. 1, p. 21–38, 2020. Citado na página 47.

CHAFE, W. L. *Significado e estrutura linguística*. Rio de Janeiro: Livros Técnicos e Científicos, 1979. Citado na página 62.

CIMIANO, P. et al. Lexinfo: A declarative model for the lexicon-ontology interface. *Journal of Web Semantics*, Elsevier, v. 9, n. 1, p. 29–51, 2011. Citado na página 56.

CIMIANO, P.; McCRAE, J. P.; BUITELAAR, P. *Lexicon Model for Ontologies: Final Community Group Report*. Cambridge, MA, 2016. Disponível em: <https://www.w3.org/2016/05/ontolex/>. Citado 4 vezes nas páginas 32, 162, 176 e 178.

CODOÑER, C. Evolución en los diccionarios de antonio de nebrija, 1492–1512. *Historiographia Linguistica*, v. 23, n. 3, p. 267–85, 1996. Citado na página 100.

CODOÑER, C. Isidore de séville : différences et vocabulaires. In: HAMESSE, J. (Ed.). *Les manuscrits des lexiques et glossaires de l'Antiquité tardive à la fin du moyen âge*. Louvain-la-Neuve: Fédération Internationale des Instituts d'Études Médiévales, 1996. p. 58–77. Citado na página 87.

CODOÑER, C. Las Etymologiae y el Liber glossarum. *Dossiers d'HEL*, n. 10, p. 179–198, 2016. Citado na página 89.

CODOÑER MERINO, C. De glosarios, vocabularios, definiciones y etimologías. In: MUÑOZ, M. J.; MARTÍN, P. C. y C. (Ed.). *La compilación del saber en la Edad Media*. Porto: Fédération Internationale des Instituts d'Études Médiévales, 2002. p. 61–84. Citado na página 90.

COLOMBAT, B. Le traitement de la construction verbale dans la grammaire latine humaniste. In: AUROUX, S. (Ed.). *History of Linguistics 1999*: selected papers from the eight international conference on the history of the language sciences, 14–19 september 1999, fontenay-st. cloud. Amsterdam/Philadelphia: John Benjamins, 2003, (Amsterdam Studies in the Theory and History of Linguistic Science. Series III, Studies in the History of the Language Sciences, v. 99). p. 63–81. ISBN 90-272-4588-6. Citado na página 115.

CONBOY, K.; FITZGERALD, B. Toward a conceptual framework of agile methods: a study of agility in different disciplines. In: ZANNIER, C.; ERDOGMUS, H.; LINDSTROM, L. (Ed.). *Extreme Programming and Agile Methods*: Xp/agile universe 2004. proceedings of the 4th conference on extreme programming and agile methods calgary, canada, august 15-18, 2004. Berlin/Heidelberg: Springer, 2004. p. 105–116. Citado na página 165.

CONSIDINE, J. *Dictionaries of Early Modern Europe*: Lexicography and the making of heritage. Cambridge: Cambridge University Press, 2008. Citado 5 vezes nas páginas 98, 99, 100, 101 e 102.

CONSIDINE, J. John jamieson, franz passow, and the double invention of lexicography on historical principles. *Journal of the History of Ideas*, JSTOR, v. 75, n. 2, p. 261–281, 2014. Citado na página 137.

CONSIDINE, J. Historical dictionaries. history and development; current issues. In: DURKIN, P. (Ed.). *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press, 2016. Citado na página 137.

CONSIDINE, J. *Sixteenth-Century English Dictionaries*. [S.l.]: Oxford University Press, 2022. v. 1. Citado na página 93.

CONSIDINE, J.; OGILVIE, S.; SAFRAN, G. The unfinished business of eighteenth-century european lexicography. In: OGILVIE, S.; SAFRAN, G. (Ed.). *The whole world in a book*: Dictionaries in the nineteenth century. Oxford: Oxford University Press, 2020. Citado na página 137.

CORMIER, M. C. New perspectives on seventeenth- and eighteenth-century european lexicography1. *International Journal of Lexicography*, v. 23, n. 2, p. 133–5, 03 2010. ISSN 0950-3846. Disponível em: <https://doi.org/10.1093/ijl/ecq011>. Citado na página 97.

COSERIU, E. *Teoría del linguaje y lingüistica general*. Madrid: Editorial Gredos, 1967. Citado na página 28.

COSTA, R.; SALGADO, A.; ALMEIDA, B. SKOS as a key element for linking lexicography to digital humanities. In: GOLUB, K.; LIU, Y.-H. (Ed.). *Information and Knowledge Organisation in Digital Humanities*: global perspectives. London/New York: Routledge, 2021. p. 178–204. Citado 2 vezes nas páginas 175 e 179.

CRANE, G. The perseus project. In: BAINBRIDGE, W. S. (Ed.). *Leadership in Science and Technology*: A reference handbook. Thousand Oaks, CA: Sage, 2012. p. 644–53. Citado 2 vezes nas páginas 154 e 155.

CRETELLA JÚNIOR, J.; CINTRA, G. d. U. *Dicionário latino-português*. São Paulo: Cia. Ed. Nacional, 1944. Citado na página 146.

CRUSE, D. A. *Lexical Semantics*. Cambridge: Cambridge University Press, 1986. Citado na página 46.

CUZZOLIN, P.; HAVERLING, G. Syntax, sociolinguistics, and literary genres. In: BALDI, P.; CUZZOLIN, P. (Ed.). *New perspectives on historical Latin syntax*: Volume 1: Syntax of the sentence. Berlin/New York: Mouton de Gruyter, 2009. p. 19–64. Citado na página 190.

DALY, L. W.; DALY, B. A. Some techniques in mediaeval latin lexicography. *Speculum*, v. 39, n. 2, p. 229–239, 1964. Citado 2 vezes nas páginas 90 e 93.

DE ANGELIS, V. (Ed.). *Papiae Elementarium*: Littera a. Milano: Cisalpino-Goliardica, 1978. Citado 2 vezes nas páginas 90 e 91.

DE ANGELIS, V. L'elementarium di Papia: método e prassi di un lessicografo. *Voces: revista de estudios de lexicología latina y antigüedad tardía*, Ediciones Universidad de Salamanca, v. 8/9, p. 121–139, 1998. Citado na página 90.

DEE, S.; FORADI, M.; ŠARIĆ, F. Learning by doing: learning to implement the tei guidelines through digital classics publication. In: BODARD, G.; ROMANELLO, M. (Ed.). *Digital Classics Outside the Echo-Chamber: Teaching, Knowledge Exchange & Public Engagement*. London: Ubiquity Press, 2016. p. 15–32. Citado na página 18.

DELATTE, L. et al. *Dictionnaire fréquentiel et index inverse de la langue latine*. Liège: Laboratoire d'Analyse statistique des Langues anciennes de l'Université de Liège (L.A.S.L.A.), 1981. Citado na página 152.

DEZOTTI, J. D. O papel do texto-documento. In: LIMA, A. D. (Ed.). *Latim, da fala à língua*. Araraquara: Faculdade de Ciências e Letras da UNESP, 1992. p. 29–31. Citado na página 29.

DEZOTTI, L. C.; PASSAROTTI, M.; MAMBRINI, F. Modelling and linking an old Latin-Portuguese dictionary to the LiLa knowledge base. In: CALZOLARI, N. et al. (Ed.). *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Torino, Italia: ELRA and ICCL, 2024. p. 11537–11547. Disponível em: <https://aclanthology.org/2024.lrec-main.1008>. Citado na página 162.

DI STEFANO, A. (Ed.). *Arusiani Messii Exempla elocutionum*. Mörlenbach: Weidmann, 2011. Disponível em: <http://hdl.handle.net/10077/33780>. Citado na página 84.

DICCIONARIO latim-portuguez, etymologico, prosodico e orthographico. 12. ed. Rio de Janeiro/Lisboa: Livraria Francisco Alves/Livraria Bertrand, 1930? (Os diccionarios do povo). Citado 3 vezes nas páginas 142, 143 e 144.

DIK, S. C. *Functional Grammar*. Dorderecht/Cinnaminson: Foris Publications, 1978. Citado na página 61.

DURAN, M. S. A lexicografia pedagógica e sua contribuição para a mudança do paradigma lexicográfico. In: XATARA, C.; BEVILACQUA, C.; HUMBLÉ, P. (Ed.). *Lexicografia Pedagógica*: pesquisas e perspectivas. Florianópolis: UFSC, 2008. p. 82–93. Citado na página 20.

DURAN, M. S.; XATARA, C. M. Dicionários semibilíngües: uma inovação? *Revista de Estudos da Linguagem*, v. 13, n. 1, p. 45–57, 2005. Citado na página 56.

DURAN, M. S.; XATARA, C. M. A metalexicografia pedagógica. *Cadernos de Tradução*, Florianópolis, v. 2, n. 18, p. 41–66, 2006. Citado na página 20.

DURAN, M. S.; XATARA, C. M. Critérios para categorização de dicionários bilíngues. In: OLIVEIRA, A. M. P. P.; ISQUERDO, A. N.; ALVES, I. M. (Ed.). *As ciências do léxico*: lexicologia, lexicografia, terminologia. São Paulo: Editora Humanitas, 2007. v. 3, p. 311–320. Citado na página 57.

DURKIN, P. *The Oxford Guide to Etymology*. Oxford: OUP, 2009. Citado na página 51.

DURKIN, P. N. Root and branch: Revising the etymological component of the oxford english dictionary. *Transactions of the Philological Society*, Wiley Online Library, v. 97, n. 1, p. 1–49, 1999. Citado na página 52.

ERNOUT, A.; MEILLET, A. *Dictionnaire étymologique de la langue latine*: Histoire des mots. Paris: Klincksieck, 2001. Citado na página 53.

FANTOLI, M. et al. Linking the LASLA corpus in the LiLa knowledge base of interoperable linguistic resources for Latin. In: DECLERCK, T. et al. (Ed.). *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, 2022. p. 26–34. Disponível em: <https://aclanthology.org/2022.ldl-1.4>. Citado na página 189.

FARIA, E. *Vocabulário Latino-Português*: significação e história das palavras, agrupadas por famílias, segundo os programas atuais. Rio de Janeiro: F. Briguiet, 1943. Citado na página 149.

FARIA, E. *Gramatica superior da Língua Latina*. Rio de Janeiro: Livraria Acadêmica, 1958. Citado na página 50.

FARIA, E. *Dicionário escolar latino-português*. Rio de Janeiro: Ministério da Educação e Cultura, 1962. Citado 2 vezes nas páginas 20 e 150.

FARIA, E. *Dicionário Escolar Latino-Português*. 6. ed. Rio de Janeiro: Ministério da Educação e Cultura/Fundação de Assistência ao Estudante, 1994. Citado 6 vezes nas páginas 151, 172, 179, 181, 182 e 183.

FARIA, E. *Dicionário escolar latino-português*. Rio de Janeiro: Ministério da Educação e Cultura, 2015. Disponível em: <https://www.dicionariolatino.com>. Acesso em: 11.7.2023. Citado na página 20.

FELLBAUM, C. *WordNet: An electronic lexical database*. Cambridge, Massachusetts: MIT press, 1998. Citado na página 158.

FERNANDES, G. A ianua linguarum dos jesuítas irlandeses (salamanca, 1611) e a porta de linguas de amaro de roboredo (lisboa, 1623). *Boletim de Estudos Clássicos*, Coimbra, n. 42, p. 165–81, 2004. Citado 3 vezes nas páginas 116, 117 e 118.

FERNANDES, G. As gramáticas do portugués de fernão de oliveira (1536) e de bento pereira (1672). *Confluência*, Rio de Janeiro, n. 33/34, p. 127–141, 2008. Citado na página 121.

FERNANDES, G. Textos gramaticais latino-portugueses na idade média. In: MORO VICENTE CALVO FERNÁNDEZ, P. P. J. E. B. (Ed.). *Historiografía lingüística*: líneas actuales de investigación. Münster: Nodus Publikationen, 2012. p. 326—39. Citado na página 94.

FERNANDES, T. *A Literatura Latina no Brasil: uma história de traduções*. Tese (Doutorado) — Universidade Federal de Santa Catarina, Florianópolis, 2017. Citado na página 190.

FERREIRA, A.; BLACKWELL, C.; PALLADINO, C. Edições digitais nas letras clássicas: elementos de gênero na produção e leitura da língua grega e do latim. *Linha D'Água (Online)*, São Paulo, v. 33, n. 2, p. 113–135, 2020. Citado na página 18.

FERREIRA, A. G. *Dicionário de Latim-Português*. Porto: Porto Editora, 1966. Citado 2 vezes nas páginas 69 e 152.

FILGUEIRAS, J. M. A produção de materiais didáticos pelo mec: da campanha nacional de material de ensino à fundação nacional de material escolar. *Revista Brasileira de História*, SciELO Brasil, v. 33, p. 313–35, 2013. Citado 3 vezes nas páginas 148, 149 e 150.

FILLMORE, C. J. The case for case. In: ____. *Form and Meaning in Language*: Vol. I: Papers on semantic roles. Chicago: The University of Chicago Press, 2003. p. 21–119. Citado na página 62.

FIORIN, J. L. A leitura e interpretação de textos. In: LIMA, A. D. (Ed.). *Latim, da fala à língua*. Araraquara: Faculdade de Ciências e Letras da UNESP, 1992. p. 19–23. Citado na página 71.

FIRMINO, N. *Dicionário latino-português*. Lisboa: Editora Académica de D. Felipa, 1945. Citado na página 148.

FLOBERT, P. La lexicographie latine en france de robert estienne à gaffiot. *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, Persée-Portail des revues scientifiques en SHS, v. 145, n. 1, p. 385–403, 2001. Citado 2 vezes nas páginas 139 e 150.

FLOBERT, P. La lexicographie latine en france. avant et après félix gaffiot. *Atala Cultures et sciences humaines*, Lycée Chateaubriand, n. 6, p. 195–209, 2003. Citado na página 139.

FONSECA, P. J. da. *Parvum lexicum latinum lusitana interpretatione adjecta*. Lisboa: Typographia Regia, 1798. Disponível em: <http://books.google.it/books?id=xZtEAAAAcAAJ>. Acesso em: 18.7.2023. Citado 9 vezes nas páginas 10, 129, 131, 134, 162, 172, 181, 182 e 183.

FONSECA, P. J. da. Parvum lexicum latinum lusitana interpretatione adjecta. In: VERDELHO, T. dos S.; SILVESTRE, J. P. M. (Ed.). *Corpus Lexicográfico do Português*. Aveiro, Portugal: Universidade de Aveiro and Centro de Linguística da Universidade de Lisboa, 2002. Disponível em: <http://clp.dlc.ua.pt/DICIweb/default.asp?url=Obras&Livro=16&Edicao=17>. Acesso em: 18.7.2023. Citado na página 130.

FRANCESE, C. *Core Latin and Ancient Greek Vocabularies*. Carlisle, PA: Dickinson College Commentaries, 2020. Disponível em: <https://dcc.dickinson.edu/vocab/core-vocabulary>. Citado 3 vezes nas páginas 113, 152 e 166.

FRANZINI, G. et al. Nunc est aestimandum: Towards an evaluation of the latin wordnet. In: *CLiC-it*. [S.l.: s.n.], 2019. Citado 2 vezes nas páginas 158 e 159.

FREIRE CAMERON, H. M. S. R. S. *A Prosodia de Bento Pereira: Contributos para o estudo lexicográfico e filológico*. Tese (Doutorado) — Universidade de Aveiro (Portugal), 2012. Citado 6 vezes nas páginas 122, 123, 124, 126, 127 e 128.

FREITAS, H. R. O prof. Ernesto Faria e sua importância para os estudos de latim. *Confluência*, n. 31, p. 49–57, 2006. Citado na página 149.

FURNO, M. Problèmes de lexicographie humaniste: un savant et un dictionnaire, paolo manuzio et le calepin. In: FERRI, R. S. (Ed.). *The Latin of Roman Lexicography*. Pisa/Roma: Farbizio Serra Editore, 2011. p. 187–222. Citado na página 104.

GAFFIOT, F. *Dictionnaire illustrée latin-français*. Paris: Hachette, 1934. Citado 2 vezes nas páginas 150 e 151.

GARCEA, A.; PLANCQ, C.; CINATO, F. Corpus grammaticorum latinorum: un projet de traitement informatique autour des grammairiens latins. *Rivista di Cultura Classica e Medioevale*, Edizioni dell'Ateneo, v. 52, n. 2, p. 377–400, 2010. Citado na página 154.

GARCIA, J. M.; CASTRO, J. A. R. O. *Dicionário gramatical de latim*. Brasília: Editora da UnB, 2003. Citado na página 152.

GARDNER, D. D. *A Frequency Dictionary of Latin Words*. Tese (Doutorado) — Stanford University, 1971. Citado na página 152.

GLARE, P. G. *Oxford Latin Dictionary*. Oxford: Clarendon Press, 1968. Citado 2 vezes nas páginas 53 e 191.

GOUWS, R. H.; TARP, S. Information overload and data overload in lexicography. *International Journal of Lexicography*, v. 30, n. 4, p. 389—415, 2016. Citado na página 205.

GRANGER, S. Electronic lexicography: From challenge to opportunity. In: GRANGER, S.; PAQUOT, M. (Ed.). *Electronic Lexicography*. Oxford: Oxford University Press, 2012. p. 1–12. Citado na página 185.

GRÉCO, G. (Ed.). *Le Gaffiot numérisé*. 2016. Formato LaTeX. Disponível em: <http://gerardgreco.free.fr/IMG/zip/gaffiot-a-z-komarov-1-1-20160502.zip>. Acesso em: 04.abr.2023. Citado na página 151.

GREGIS, H. *Desenho de um dicionário bilíngue latino-português para alunos em fase inicial de aprendizagem*. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2018. Citado na página 20.

GROSS, M. Lexicon-grammar and the syntactic analysis of french. In: *International Conference on Computational Linguistics (COLING)*. [S.l.: s.n.], 1984. p. 275–282. Citado na página 67.

GROSS, M. La construction de dictionnaires électroniques. *Annals of Telecommunications*, v. 44, n. 1-2, p. 4–19, 1989. Citado na página 67.

HALLIDAY, M. A. K.; MATTHIESSEN, C. *Introduction to functional grammar*. New York: Hodder Arnold, 2004. Citado na página 61.

HAMESSE, J. Introduction. In: HAMESSE, J. (Ed.). *Les manuscrits des lexiques et glossaires de l'Antiquité tardive à la fin du moyen âge*. Louvain-la-Neuve: Fédération Internationale des Instituts d'Études Médiévales, 1996. p. ix–xiii. Citado na página 89.

HANKS, P. Corpus evidence and electronic lexicography. In: GRANGER, S.; PAQUOT, M. (Ed.). *Electronic Lexicography*. Oxford: Oxford University Press, 2012. p. 57–82. Citado na página 185.

HANKS, P. Lexicography and technology in the renaissance and now. In: *EFNIL Conference, Budapest 2012*. Frankfurt am Main: Peter Lang, 2012. p. 23–50. Citado na página 98.

HANKS, P. *Lexical analysis: norms and exploitations*. Cambridge, MA/London: MIT Press, 2013. Citado na página 72.

HANKS, P. Lexicography from earliest times to the present. In: ALLAN, K. (Ed.). *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press, 2013. p. 503–536. Citado na página 78.

HANKS, P. Lexicography. In: MITKOV, R. (Ed.). *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, 2022. Citado 6 vezes nas páginas 35, 56, 62, 152, 170 e 217.

HANNAY, M. Types of bilingual dictionaries. In: STERKENBURG, P. (Ed.). *A Practical Guide to Lexicography*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 1996. p. 145–153. Citado na página 29.

HARTMANN, R. R. K. Structural and typological perspectives. In: ____. *Teaching and Researching Lexicography*. 2nd.. ed. London/New York: Routledge, 2001. p. 57–79. Citado 8 vezes nas páginas 29, 34, 37, 45, 46, 78, 172 e 177.

HARTMANN, R. R. K.; JAMES, G. *Dictionary of Lexicography*. London and New York: Routledge, 2002. Citado na página 37.

HAUG, D. T.; JØHNDAL, M. Creating a parallel treebank of the old indo-european bible translations. In: SPORLEDER, C.; RIBAROV, K. (Ed.). *Proceedings of the second workshop on language technology for cultural heritage data (LaTeCH 2008)*. [S.l.: s.n.], 2008. p. 27–34. Citado 2 vezes nas páginas 162 e 189.

HEID, U. Corpus linguistics and lexicography. In: LÜDELING, A.; KYTÖ, M. (Ed.). *Corpus Linguistics*: An international handbook. Berlin/New York: Walter de Gruyter, 2008. v. 1, p. 131–53. Citado 6 vezes nas páginas 185, 186, 188, 197, 198 e 218.

HJØRLAND, B. Knowledge organization (ko). *Knowledge Organization*, Nomos Verlagsgesellschaft, v. 43, n. 6, p. 475–484, 2016. Disponível em: <https://www.isko.org/cyclo/knowledge_organization>. Citado na página 179.

HOCKEY, S. The history of humanities computing. In: SCHRIEBMAN, S.; SIEMENS, R.; UNSWORTH, J. (Ed.). *A Companion to Digital Humanities*. Oxford: Blackwell, 2004. Disponível em: <https://companions.digitalhumanities.org/DH/>. Citado na página 152.

HOFFMANN, S. Using the OED quotations database as a corpus: a linguistic appraisal. *ICAME journal*, v. 28, n. 4, p. 17–30, 2004. Citado na página 217.

HOLTZ, L. *Donat et la tradition de l'enseignement grammatical*: Étude sur l'ars donati et sa diffusion (ive-ixe siècle) et édition critique. Paris: Centre National de la Recherce Scientifique, 1981. Citado na página 80.

HOLTZ, L. Glossaires et grammaire dans l'antiquité. In: HAMESSE, J. (Ed.). *Les manuscrits des lexiques et glossaires de l'Antiquité tardive à la fin du moyen âge*. Louvain-la-Neuve: Fédération Internationale des Instituts d'Études Médiévales, 1996. p. 1–21. Citado 4 vezes nas páginas 80, 86, 87 e 88.

HORATIUS FLACCUS, Q. Epistulae. In: KLINGNER, F. (Ed.). *Q. Horati Flacci Opera*. Leipzig: Teubner, 1959. Citado na página 80.

HOUAISS, A. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva Instituto Antônio Houaiss de Lexicografia, 1986. Citado na página 52.

HUMBLÉ, P. *Dictionaries and language learners*. Frankfurt-am-Main: Haag und Herchen, 2001. Citado na página 36.

HUNSTON, S. Collection strategies and design decisions. In: LÜDELING, A.; KYTÖ, M. (Ed.). *Corpus Linguistics*: An international handbook. Berlin/New York: Walter de Gruyter, 2008. v. 1, p. 154–68. Citado 4 vezes nas páginas 186, 187, 188 e 191.

HUNSTON, S.; FRANCIS, G. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins, 2015. Citado na página 65.

HUNT, T. *Teaching and learning Latin in thirteenth-century England*. Cambridge: D.S.Brewer, 1991. v. 1. Citado 4 vezes nas páginas 91, 92, 93 e 94.

IDE, N.; PUSTEJOVSKY, J. What does interoperability mean, anyway: Toward an operational definition of interoperability for language technology. In: *Proceedings of the Second International Conference on Global Interoperability for Language Resources (ICGL 2010)*. Hong Kong: [s.n.], 2010. Disponível em: <http://www.cs.vassar.edu/~ide/papers/ICGL10.pdf>. Citado na página 175.

IDE, N.; VÉRONIS, J. Encoding dictionaries. *Computers and the Humanities*, Springer, v. 29, n. 2, p. 167–179, 1995. Citado na página 155.

IKEN, S. *Index totius artis* (1599-1755): algumas reflexões sobre o índice lexicográfico latino-português da gramática de Manuel Álvares, elaborado por António Velez. In: KEMMLER, R.; SCHÄFER-PRIEß, B.; SCHÖNBERGER, A. (Ed.). *Estudos de história da gramaticografia e lexicografia portuguesas*. Frankfurt am Main: Domus Editoria Europaea, 2002, (Beihefte zu Lusorama. Erste Reihe: Studien zur portugiesischen Sprachwissenschaft, v. 9). p. 53–83. ISBN 3-927884-57-X. Citado 2 vezes nas páginas 111 e 113.

ILSON, R. Etymological information: can it help our students? *ELT journal*, Oxford University Press, v. 37, n. 1, p. 76–82, 1983. Citado na página 54.

JACINTO GARCÍA, E. J. *Forma y función del diccionario. Hacia una teoría general del ejemplo lexicográfico*. Jaén: Servicio de Publicaciones de la Universidad de Jaén, 2015. Citado 3 vezes nas páginas 68, 69 e 70.

JACKSON, H. *Lexicography: an introduction*. New York: Routledge, 2002. Citado na página 40.

JEŽEK, E.; SPRUGNOLI, R. *Linguistica Computazionale. Introduzione all'analisi automatica dei testi*. Bologna: il Mulino, 2023. Citado 9 vezes nas páginas 38, 39, 40, 41, 44, 58, 60, 63 e 65.

JOHNS, T. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, São Paulo, v. 4, p. 27–45, 1991. Citado na página 71.

JONES, S. E. *The emergence of the digital humanities*. London: Routledge, 2014. Citado na página 17.

KALLAS, J. et al. Lexicographic practices in europe: Results of the elexis survey on user needs. In: KOSEM, I. et al. (Ed.). *Electronic Lexicography in the 21st Century: Smart Lexicography*: Proceedings of the elex 2019 conference, sintra, portugal. Brno, Czech Republic: Lexical Computing CZ, 2019. p. 519–36. Citado na página 175.

KALTNER, L. F.; SANTOS, M. C. S. dos. Ernesto Faria (1906-1952) e o latim para a formação de filólogos e linguistas no Brasil do século XX: análise historiográfica. *Cadernos de Linguística*, v. 5, n. 2, 2024. Citado 3 vezes nas páginas 139, 148 e 152.

KEIL, H. (Ed.). *Grammatici latini*. Lipsiae: Teubner, 1857. v. 1. Citado 2 vezes nas páginas 85 e 119.

KEMMLER, R. De institutione grammatica libri tres (lisboa, 1573): a edição princeps da ars minor de manuel álvares. *Revista Portuguesa de Humanidades*, Aletheia Associação Científica e Cultural/Publicações da Faculdade de Filosofia de Braga, Braga, v. 17, n. 1, 2013. Citado na página 111.

KEMMLER, R. Handwritten annotations in the early editions of Manuel Alvares' *De institvtione grammatica libri tres*. *Philologica Jassyensia*, v. 28, p. 57–69, 2018. Citado na página 111.

KENNEDY, G. *An introduction to corpus linguistics*. London and New York: Longman, 1998. Citado 3 vezes nas páginas 186, 187 e 217.

KENSKI, V. M. *Educação e tecnologias: o novo ritmo da informação*. Campinas, SP: Papirus, 2007. Citado na página 18.

KERNERMAN, I. Introducing lexicala: crosslingual multilayer lexical data. *Lexicala Review*, n. 26, p. 1, 2018. Citado na página 168.

KHAN, A. *collapsibleTree: Interactive Collapsible Tree Diagrams using 'D3.js'*. [S.l.], 2023. R package version 0.1.8. Disponível em: <https://CRAN.R-project.org/package=collapsibleTree>. Citado na página 170.

KILGARRIFF, A. et al. The sketch engine. *Lexicography*, v. 1, n. 1, p. 7–36, 2014. Citado na página 165.

KILGARRIFF, A. et al. Gdex: Automatically finding good dictionary examples in a corpus. In: BERNAL, J. D. E. (Ed.). *Proceedings of the 13th EURALEX International Congress*. Barcelona, Spain: Institut Universitari de Linguistica Aplicada, Universitat Pompeu Fabra, 2008. p. 425–432. Citado 2 vezes nas páginas 200 e 201.

KILGARRIFF, A.; KOSEM, I. Corpus tools for lexicographers. In: GRANGER, S.; PAQUOT, M. (Ed.). *Electronic Lexicography*. Oxford: Oxford University Press, 2012. p. 31–56. Citado na página 199.

KOEHLER, H. *Pequeno dicionário escolar latino-português*. Porto Alegre: Editôra Globo, 1935. Citado na página 146.

KOPP, U. F. (Ed.). *Martiani Minei Felicis Capellae Afri Carthaginiensis De nuptiis Philologiae et Mercurii et de septem artibus liberalibus libri novem*. Francofurti ad Moenum: F. Varrentrapp, 1836. v. 1. Citado na página 84.

KORKIAKANGAS, T. Late latin charter treebank: contents and annotation. *Corpora*, Edinburgh University Press, v. 16, n. 2, p. 191–203, 2021. Citado na página 190.

KOSSÁRIK, M. A doutrina linguística de amaro de roboredo. In: *Actas do XII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: Associção Portuguesa de Linguística, 1997. v. 2, p. 429–443. Citado na página 117.

KOSSARIK, M. A. A obra de amaro de roboredo: questões de historiografia linguística portuguesa. In: ROBOREDO, A.; KOSSARIK, M. A. (Ed.). *Methodo Grammatical para todas as Linguas*. Lisboa: Imprensa Nacional/Casa da Moeda, 2002. Citado na página 117.

KRAMER, J. I glossari tardo-antichi di tradizione papiracea. In: HAMESSE, J. (Ed.). *Les manuscrits des lexiques et glossaires de l'Antiquité tardive à la fin du moyen âge*. Louvain-la-Neuve: Fédération Internationale des Instituts d'Études Médiévales, 1996. p. 23–55. Citado na página 85.

KRUYT, T. Multifunctional linguistic databases: Their multiple use. In: STERKENBURG, P. (Ed.). *A Practical Guide to Lexicography*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 1996. p. 194–203. Citado na página 36.

LANCASHIRE, I. The lexicons of early modern english. *Digital Studies/Le champ numérique*, Open Library of Humanities, n. 8, 2003. Citado na página 154.

LAUFER, B. Corpus-based versus lexicographer examples in comprehension and production of new words. In: TOMMOLA, H. et al. (Ed.). *Proceedings of the 5th EURALEX International Congress*. Tampere, Finland: Tampereen Yliopisto, 1992. p. 71–76. ISBN 951-44-3111-1. Citado na página 71.

LAZARD, S. La structuration du lexique dans le Vocabularium de Papias. *Revue de linguistique romane*, n. 261-62, p. 221–243, 2002. Citado na página 90.

LEITE, J. F. M.; JORDÃO, A. J. N. *Dicionário latino vernáculo*: etimologia-literatura-historia-mitologia-geografia. Rio de Janeiro: Laborvincit, 1944. Citado na página 148.

LEITE, L. R. Aprendendo a ler (e a falar) em latim. In: PRATA, P.; FORTES, F. (Ed.). *O latim hoje*: reflexões sobre cultura clássica e ensino. Campinas, SP: Mercado de Letras, 2015. Citado na página 29.

LENDINARA, P. Middle english glossaries. In: SEILER, A.; BENATI, C.; PONS-SANZ, S. (Ed.). *Medieval Glossaries from North-Western Europe*: Tradition and innovation. Turnhout: Brepols, 2023. p. 293–312. Citado na página 94.

LEVIN, B. *English verb classes and alternations*: A preliminary investigation. Chicago and London: The University of Chicago Press, 1993. Citado na página 62.

LEWIS, C.; SHORT, C. *A Latin Dictionary*. Oxford: Clarendon Press, 1879. Citado na página 19.

LEWIS, C. T.; SHORT, C. *A Latin Dictionary*. Medford, MA: Trustees of Tufts University/Perseus Project, 1997. Disponível em: <https://github.com/PerseusDL/lexica/blob/master/CTS_XML_TEI/perseus/pdllex/lat/ls/lat.ls.perseus-eng2.xml>. Citado na página 19.

LHOMMÉ, M.-K. De l'encyclopédie au glossaire: Festus et son adaptation par paul diacre. In: FERRI, R. S. (Ed.). *The Latin of Roman Lexicography*. Pisa/Roma: Farbizio Serra Editore, 2011. Citado na página 81.

LIMA, A. D. *Uma estranha língua?*: Questões de linguagem e de método. São Paulo: EdUnesp, 1995. Citado 5 vezes nas páginas 29, 36, 47, 57 e 71.

LINDSAY, W. M. (Ed.). *Nonii Marcelli De compendiosa doctrina libros XX*. Lipsiae: Teubner, 1903. Citado 2 vezes nas páginas 82 e 83.

LINDSAY, W. M. (Ed.). *Isidori Hispalensis Etymologiarum sive originum libri XX*. Oxonii: e typographeo Clarendoniano, 1911. Citado na página 87.

LINDSAY, W. M. (Ed.). *Sexti Pompei Festi De Verborum Significatu Quae Supersunt cum Pauli Epitome*. Lipsiae: Teubner, 1913. Citado na página 81.

LITTA, E.; PASSAROTTI, M.; MAMBRINI, F. The treatment of Word Formation in the LiLa Knowledge Base of linguistic resources for Latin. In: *Proceedings of the 2nd Int. Workshop on Resources and Tools for Derivational Morphology (DeriMo 2019)*. Prague, Czechia: [s.n.], 2019. p. 35–43. Disponível em: <https://doi.org/10.5281/zenodo.3403022>. Citado 2 vezes nas páginas 54 e 158.

LITTA MODIGNANI, E.; PEDONESE, G.; FRANZINI, G. *CIRCSE/latinWordnet-revision: v1.5*. Zenodo, 2023. Disponível em: <https://github.com/CIRCSE/latinWordnet-revision>. Citado na página 168.

LONGO, G. *Ensino de Latim: problemas lingüísticos e uso de dicionário*. Dissertação de Mestrado em Linguística e Língua Portuguesa — Universidade Estadual Paulista, Araraquara, SP, 2006. Citado 3 vezes nas páginas 36, 47 e 48.

LUGLI, L. Agile lexicography: Rapid dictionary prototyping. In: ARISTA, J. M.; LÓPEZ, A. E. O. (Ed.). *Structuring Lexical Data and Digitising Dictionaries*: Grammatical theory, language processing and databases in historical linguistics. Leiden: Brill, 2024. p. 121–149. Citado na página 165.

LUGLI, L. *lexicographR app*. figshare, 2024. Disponível em: <https://github.com/MangalamResearch-Lexicography/lexicographR>. Citado 3 vezes nas páginas 33, 165 e 167.

LUNARDINI, V. Su alcune glosse di un lessico monacense anonimo. *Archivum Latinitatis Medii Aevi*, v. 67, p. 273–9, 2009. Citado na página 91.

LUNARDINI, V. Lessici mediolatini on-line. l'esperienza di mirabile. *Archivum Latinitatis Medii Aevi*, v. 77, p. 401–9, 2019. Citado 2 vezes nas páginas 153 e 154.

LYONS, J. *Semântica I*. Lisboa: Editorial Presença/Martins Fontes, 1980. Citado 4 vezes nas páginas 38, 58, 59 e 60.

MACAMBIRA, D. D. *Impressões do tempo*: os almanaques no ceará (1870-1908). Dissertação (Mestrado) — Universidade Federal do Ceará, Fortaleza, 2010. Citado na página 142.

MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. d. F. G. M. Reflexões acerca das relações entre representação temática e descritiva. *Informação & Sociedade*, Universidade Federal da Paraíba-Programa de Pós-Graduação em Ciência da …, João Pessoa, v. 21, n. 1, p. 27–35, 2011. Citado na página 175.

MAMBRINI, F. et al. Linking the Lewis & Short dictionary to the LiLa Knowledge Base of interoperable linguistic resources for Latin. In: FERSINI, E.; PASSAROTTI, M.; PATTI, V. (Ed.). *Proceedings of the Eighth Italian Conference on Computational Linguistics*. Aachen, Germany: CEUR Workshop Proceedings, 2021. (AIxIA Series). ISSN 1084-6654. Citado 2 vezes nas páginas 19 e 162.

MAMBRINI, F. et al. Interlinking valency frames and WordNet synsets in the LiLa Knowledge Base of linguistic resources for Latin. In: ALAM, M. et al. (Ed.). *Further with Knowledge Graphs*. Berlin: AKA Verlag and IOS Press, 2021. Disponível em: <https://doi.org/10.3233/SSW210032>. Citado na página 160.

MARMORALE, E. V.; BARTOLOMEU JÚNIOR, J. *História da literatura latina*: Volume 1. Lisboa: Estúdios Cor, 1974. Citado na página 190.

MARQUES, F. C. [Recensão a] Ernesto Faria, a renovação actual dos estudos latinos; o latim e a cultura contemporânea. *Humanitas*, Faculdade de Letras da Universidade de Coimbra, Instituto de Estudos Clássicos, Coimbra, v. 1, p. 253–57, 1947. Citado na página 149.

MÁRQUEZ CRUZ, M. A lexicographical model based on the predicative framework theory (functional grammar) for sense disambiguation. An application to Latin author dictionaries. *Digital Scholarship in the Humanities*, v. 33, n. 2, p. 362–373, 08 2017. ISSN 2055-7671. Disponível em: <https://doi.org/10.1093/llc/fqx037>. Citado 2 vezes nas páginas 60 e 72.

MÁRQUEZ CRUZ, M.; FERNÁNDEZ-PAMPILLÓN, A. A morpho-semantic digital didactic dictionary for learners of latin at early stages. In: GAVRIILIDOU, Z.; MITSIAKI, M.; FLIATOURAS, A. (Ed.). *Lexicography for Inclusion*: Proceedings of the 19th euralex international congress, 7-9 september 2021, alexandroupolis, vol. 1. Komotini, Greece: Democritus University of Thrace, 2021. p. 193–202. Citado 4 vezes nas páginas 10, 72, 73 e 202.

MCENERY, T.; HARDIE, A. *Corpus linguistics: Method, Theory and Practice*. Cambridge, MA: Cambridge University Press, 2012. Citado na página 61.

MCENERY, T.; WILSON, A. *Corpus linguistics*: An introduction. Edinburgh: Edinburgh University Press, 2001. Citado 2 vezes nas páginas 187 e 198.

MCGEE, I. Collocation dictionaries as inductive learning resources in data-driven learning: an analysis and evaluation. *International Journal of Lexicography*, v. 25, n. 3, p. 319–361, 2012. Citado na página 71.

MCGILLIVRAY, B. *Methods in Latin Computational Linguistics*. Leiden/Boston: Brill, 2014. Citado na página 160.

MCGILLIVRAY, B.; PASSAROTTI, M.; RUFFOLO, P. The index thomisticus treebank project: Annotation, parsing and valency lexicon. *Traitement Automatique des Langues*, v. 50, n. 2, p. 103–127, 2009. Citado na página 19.

MCGILLIVRAY, B.; VATRI, A. Computational valency lexica for latin and greek in use: a case study of syntactic ambiguity. *Journal of Latin Linguistics*, De Gruyter, v. 14, n. 1, p. 101–126, 2015. Citado na página 19.

MCKITTERICK, R. *The Carolingians and the written word*. Cambridge: Cambridge University Press, 1989. Citado na página 89.

MĚCHURA, M. B. Introducing lexonomy: an open-source dictionary writing and publishing system. In: KOSEM, I. et al. (Ed.). *Electronic Lexicography in the 21st Century*: Lexicography from scratch. proceedings of the elex 2017 conference. Brno: Masaryk University, 2017. p. s.p. Citado na página 165.

MEDINA GUERRA, A. M. Cinco siglos en la historia de los diccionarios bilingües latín-español y español-latín (xv-xix). *Philologia hispalensis*, v. 22, p. 259–88, 2008. Citado 2 vezes nas páginas 99 e 100.

MELISS, M.; SÁNCHEZ HERNÁNDEZ, P. Nuevos retos para la lexicografía pedagógica bilingüe alemán-español. *Revista Internacional de Lenguas Extranjeras*, v. 4, p. 123–150, 2015. Citado na página 57.

MEL'ČUK, I.; POLGUÈRE, A. Theory and practice of lexicographic definition. *Journal of Cognitive Science*, v. 19, n. 4, p. 417–470, 2018. Citado 2 vezes nas páginas 54 e 55.

MERRILEES, B. The latin-french dictionarius of firmin le ver (1420-1440). In: *Zürilex '86 Proceedings*. Tübingen: A. Francke Verlag, 1988. p. 181–188. Citado na página 173.

MERRILEES, B. Métalexicographie médiévale: la fonction de la métalangue dans un dictionnaire bilingue du moyen âge. *Archivum Latinitatis Medii Aevi*, Persée-Portail des revues scientifiques en SHS, v. 50, n. 1, p. 33–70, 1990. Citado na página 116.

MERRILEES, B. The shape of the medieval dictionary entry. *Digital Studies/le Champ Numérique*, Open Library of Humanities, v. 4, 1996. Citado 3 vezes nas páginas 12, 173 e 174.

MILLER, G. A. Wordnet: a lexical database for english. *Communications of the ACM*, ACM New York, NY, USA, v. 38, n. 11, p. 39–41, 1995. Citado na página 158.

MINOZZI, S. Latin wordnet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'information retrieval. In: MASTANDREA, P. (Ed.). *Strumenti digitali e collaborativi per le Scienze dell'Antichita*. Venezia: Edizioni Ca' Foscari, 2017. p. 123–134. Citado na página 160.

MOLINA GARCIA, D. *Fraseologia Bilingue: Un Enfoque Lexicografico-Pedagogico*. Granada: Comares, 2006. Citado 2 vezes nas páginas 56 e 57.

MOMIGLIANO, A. *Os limites da helenização*: a interacao cultural das civilizacoes grega, romana, celtica, judaica e persa. Rio de Janeiro: Jorge Zahar, 1991. Citado na página 80.

MORESCHINI, C.; NORELLI, E. *História da literatura cristã antiga grega e latina*: Vol. ii, tomo 1: Do concílio de niceia ao início da idade média. São Paulo: Edições Loyola, 2000. Citado 3 vezes nas páginas 79, 80 e 86.

MORRISSEY, R. Texts and contexts: The artfl database in french studies. *Profession*, JSTOR, p. 27–33, 1993. Citado na página 154.

NATION, I. S. P. *Learning Vocabulary in Another Language*. Cambridge University Press, 2013. Disponível em: <http://dx.doi.org/10.1017/cbo9781139858656>. Citado na página 54.

NEBRIJA, E. A. d. *Lexicon hoc est dictionarium ex sermone latinum in hispaniensem*. Salmanticae: Juan de Porras, 1492. Citado na página 100.

NEVES, M. H. d. M. *A gramática funcional*. São Paulo: Martins Fontes, 1997. Citado na página 61.

NIVRE, J. et al. Universal dependencies v2: An evergrowing multilingual treebank collection. *CoRR*, abs/2004.10643, 2020. Disponível em: <https://arxiv.org/abs/2004.10643>. Citado na página 189.

NYHAN, J. Text encoding and scholarly digital editions. In: WARWICK, C.; TERRAS, M.; NYHAN, J. (Ed.). *Digital humanities in practice*. London: Facet Publishing, 2012. p. 117–38. Citado 2 vezes nas páginas 154 e 155.

PAIVA, V. de; RADEMAKER, A.; MELO, G. de. Openwordnet-pt: An open Brazilian Wordnet for reasoning. In: *Proceedings of COLING 2012*: Demonstration papers. Mumbai, India: The COLING 2012 Organizing Committee, 2012. p. 353–360. Published also as Techreport, cf. http://hdl.handle.net/10438/10274. Disponível em: <http://www.aclweb.org/anthology/C12-3044>. Citado 2 vezes nas páginas 160 e 170.

PASINI, G. *Vocabula Latini Italique Sermonis*. Venetiis: Apud Jo. Mariam Lazzaronum, 1735. Citado na página 134.

PASSAROTTI, M. et al. The lemlat 3.0 package for morphological analysis of latin. In: *Proceedings of the NoDaLiDa 2017 workshop on processing historical language*. [S.l.: s.n.], 2017. p. 24–31. Citado na página 48.

PASSAROTTI, M. et al. Interlinking through lemmas. the lexical collection of the LiLa Knowledge Base of linguistic resources for Latin. *Studi e Saggi Linguistici (SSL)*, v. 58, n. 1, p. 177–212, 2020. Citado 4 vezes nas páginas 45, 49, 156 e 157.

PASSAROTTI, M.; SAAVEDRA, B. G.; ONAMBELE, C. Latin vallex. a treebank-based semantic valency lexicon for latin. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. [S.l.: s.n.], 2016. p. 2599–2606. Citado 2 vezes nas páginas 19 e 160.

PEDRO, C. M. L. da Silva de O. *Sixteenth-century Print Culture in the Kingdom of Portugal: A Study on Typographical Specimens Used in the Printing Houses of João de Barreira and João Álvares*. Tese (Doutorado) — Universidade do Porto, 2018. Citado na página 106.

PELLEGRINI, M. et al. Prinparlat: a resource of latin principal parts. In: NAMER, F.; LIGNON, S. (Ed.). *Fourth International Symposium of Morphology*: Ismo, 13-15 september 2023. Nancy, France: Université de Lorraine/ATILF/CNRS, 2023. p. 162–165. Disponível em: <https://ismo2023.ovh/?fichier=programme>. Citado 2 vezes nas páginas 158 e 168.

PENA, L. d. A. *Mini Dicionário de Latim-Português*. Rio de Janeiro: Edições Spiker, 1959. Citado na página 152.

PEREIRA, B. *Prosodia in vocabularium bilingue, Latinum, et Lusitanum digesta*. Evora: Typographia Academiae, 1697. Disponível em: <http://purl.pt/30226>. Acesso em: 18.7.2023. Citado 2 vezes nas páginas 124 e 125.

PEREIRA, B. Prosodia in vocabularium bilingue, latinum, et lusitanum digesta. In: VERDELHO, T. dos S.; SILVESTRE, J. P. M. (Ed.). *Corpus Lexicográfico do Português*. Aveiro, Portugal: Universidade de Aveiro and Centro de Linguística da Universidade de Lisboa, 2002. Disponível em: <http://clp.dlc.ua.pt/DICIweb/default.asp?url=Obras&Livro=2&Edicao=2>. Acesso em: 18.7.2023. Citado na página 124.

PEREIRA, M. A. *Quintiliano gramático*: o papel do mestre de gramática na institutio oratoria. São Paulo: Humanitas, 2005. Citado 2 vezes nas páginas 80 e 81.

PERINI, M. A. *Estudos de gramática descritiva*. São Paulo: Parábola Editorial, 2008. Citado na página 64.

PIANTA, E.; BENTIVOGLI, L.; GIRARDI, C. Multiwordnet: developing an aligned multilingual database. In: *1st International Global WordNet Conference, January 21-25, 2002: Proceedings*. Mysore, India: Central Institute of Indian Languages, 2002. p. 293–302. Citado na página 159.

PIASECKI, M. et al. Introduction to the special issue: On wordnets and relations. *Language Resources and Evaluation*, Springer, v. 47, p. 757–767, 2013. Citado na página 159.

PIERSON, H. D. Using etymology in the classroom. *ELT journal*, Oxford University Press, v. 43, n. 1, p. 57–63, 1989. Citado na página 54.

PINELLI, L. Mirabile: condivisione del lavoro/condivisione dei saperi per lo studio del medioevo. In: CANETTIERI, P. et al. (Ed.). *Filologia medievale*: comparatistica, critica del testo e attualità. atti del convegno (viterbo, 26-28 settembre 2018). Roma: L'Erma di Bretschneider, 2019. p. 195–201. Citado 2 vezes nas páginas 153 e 154.

PINFIELD, S. et al. *Open access in theory and practice*: the theory-practice relationship and openness. New York: Taylor & Francis, 2021. (Routledge Critical Studies on Open Access). Citado na página 27.

PINKSTER, H. *The Oxford Latin Syntax I: the simple clause*. Oxford: Oxford University Press, 2015. Citado 3 vezes nas páginas 64, 72 e 215.

POSIT TEAM. *RStudio*: Integrated development environment for r. Boston, MA: Posit Software, PBC, 2023. Disponível em: <http://www.posit.co/>. Citado na página 165.

PRADO, J. B. T. Latim, moinhos e cavaleiros andantes. In: LIMA, A. D. (Ed.). *Latim, da fala à língua*. Araraquara: Faculdade de Ciências e Letras da UNESP, 1992. p. 71–74. Citado na página 144.

PRADO, J. B. T. O texto-documento: sugestões de trabalho. In: LIMA, A. D. (Ed.). *Latim, da fala à língua*. Araraquara: Faculdade de Ciências e Letras da UNESP, 1992. p. 33–37. Citado 2 vezes nas páginas 28 e 29.

PRATA, P.; FORTES, F. A sobrevivência do latim. In: PRATA, P.; FORTES, F. (Ed.). *O latim hoje*: reflexões sobre cultura clássica e ensino. Campinas, SP: Mercado de Letras, 2015. p. 23–30. Citado na página 29.

QUEIROZ, O. A. P. *Dicionário latim-português*. São Paulo: Ed. Lep, 1943. Citado na página 146.

QUICHERAT, L.; DAVELUY, A. *Dictionnaire latin-français*. Paris: Hachette, 1865. Citado 2 vezes nas páginas 140 e 141.

QUICHERAT, L.; DAVELUY, A.; CHATELAIN, É. *Dictionnaire latin-français*. Paris: Hachette, 1892. Citado 2 vezes nas páginas 140 e 141.

R CORE TEAM. *R*: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2023. Disponível em: <https://www.R-project.org/>. Citado na página 165.

RABUSKE, A. Francisco rodrigues dos santos saraiva: algo de sua vida e obra, máxime no rio grande do sul. *Perspectiva Teológica*, v. 8, n. 16, p. 27–57, 1976. Citado 2 vezes nas páginas 136 e 142.

RAMALHO, A. d. C. *Para a história do humanismo em Portugal Vol. 5*. Coimbra: Imprensa da Universidade de Coimbra, 2013. Citado na página 105.

RAMISCH, C. *A generic and open framework for multiword expressions treatment: from acquisition to applications*. Tese (Doutorado) — Université de Grenoble (France); Universidade Federal do Rio Grande do Sul, 2012. Citado na página 66.

REZENDE, A. M. d.; BIANCHET, S. B. *Dicionário do latim essencial*. Belo Horizonte: Autêntica Editora, 2014. Citado na página 152.

ROBERTS, W.; EGG, M. A comparison of selectional preference models for automatic verb classification. In: MOSCHITTI, A.; PANG, B.; DAELEMANS, W. (Ed.). *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014. p. 511–522. Disponível em: <https://aclanthology.org/D14-1057>. Citado na página 63.

ROBOREDO, A. de. *Methodo grammatical para todas as linguas*. Lisboa: Pedro Craesbeeck, 1619. Citado 2 vezes nas páginas 117 e 118.

ROBOREDO, A. de. *Raizes da lingua Latina / Radices sermonis Latini*. Lisboa: Pedro Craesbeeck, 1621. Citado 4 vezes nas páginas 118, 119, 120 e 121.

ROBOREDO, A. de. *Porta de línguas ou modo muito accommodado para as entender / Ianua linguarum sive modus maxime accomodatus ad eas intelligendas*. Lisboa: Pedro Crasbeeck, 1623. Citado na página 121.

RODRIGUES, D. A. V. *Dicionário Latino Português*. [S.l.]: Publicações Brasil, 1943. Citado na página 146.

RODRIGUES, M. A. *Do humanismo à contra-reforma em Portugal*. Coimbra: Centro de História da Sociedade e da Cultura da Universidade de Coimbra, 1981. Citado na página 105.

ROELLI, P. The corpus corporum, a new open latin text repository and tool. *Archivum Latinitatis Medii*, n. 72, p. 289–304, 2014. Citado 2 vezes nas páginas 35 e 154.

ROMANELLO, M.; BODARD, G. Introduction. In: BODARD, G.; ROMANELLO, M. (Ed.). *Digital Classics Outside the Echo-Chamber*: Teaching, knowledge exchange & public engagement. London: Ubiquity Press, 2016. p. 1–12. Citado na página 18.

ROMARY, L.; TASOVAC, T. Tei lex-0: A target format for tei-encoded dictionaries and lexical resources. In: SUZUKI, C. (Ed.). *Proceedings of the 8th Conference of Japanese Association for Digital Humanities (JADH 2018): Leveraging Open Data*: Hitotsubashi hall, tokyo, september 9-11, 2018. Tokyo: Center for Open Data in the Humanities, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 2018. p. 274–5. Citado na página 175.

ROMEO, R. P. La pedagogía del latín en portugal durante la primera mitad del siglo xvii: cuatro gramáticos lusitanos. *Cuadernos de filología clásica. Estudios latinos*, Servicio de Publicaciones U.C.M., Madrid, n. 10, p. 217–28, 1996. Citado na página 116.

SACRAMENTO BLAKE, A. V. A. Francisco rodrigues dos santos saraiva. In: ____. *Diccionario Bibliographico Brazileiro*. Rio de Janeiro: Typographia Nacional, 1883. v. 3, p. 106–7. Citado 2 vezes nas páginas 136 e 141.

SALMON-ALT, S. Data structures for etymology: towards an etymological lexical network. *Bulletin de linguistique appliquée et générale*, Presses Universitaires de Franche-Comté, v. 31, p. 1–12, 2006. Disponível em: <https://hal.science/hal-00110971>. Citado na página 52.

SÁNCHEZ SALOR, E.; GÓMEZ GÓMEZ, J. M. Introdução. In: *Manuel Álvares, Instituição da Gramática, ampliada e explicada por António Velez*. Coimbra: Imprensa da Universidade de Coimbra, 2020, (Portvgaliae Monvmenta Neolatina, v. 24). p. 7–32. ISBN 978-989-26-1982-8. Citado na página 111.

SANTOS, M. M. d. Dionísio da trácia: Arte. *Letras clássicas*, n. 11, p. 153–179, 2007. Citado na página 80.

SANTOS SOBRINHO, J. A. O latim no Brasil na primeira metade do século xx. *Phaos: Revista de Estudos Clássicos*, Campinas,SP, n. 13, 2013. Citado 2 vezes nas páginas 50 e 148.

SARAIVA, F. R. d. S. *Novissimo Diccionario Latino-Portuguez*. 9. ed. Paris: Livraria Garnier, 1927. Citado 7 vezes nas páginas 53, 138, 141, 172, 181, 205 e 217.

SCHÄFER-PRIESS, B. *A Gramaticografia Portuguesa até 1822*: Condições da sua génese e critérios de categorização, no âmbito da tradição latina, espanhola e francesa. Vila Real: Centro de Estudos em Letras da Universidade de Trás-Os-Montes e Alto Douro, 2019. Citado na página 122.

SCHIBEL, W.; RYDBERG-COX, J. A. Early modern culture in a comprehensive digital library. *D-Lib Magazine*, Corporation for National Research Initiatives, v. 12, n. 3, 2006. Citado na página 154.

SEDELOW, S. Y. Computational lexicography. *Computers and the Humanities*, Paradigm Press, Inc., v. 19, n. 2, p. 97–101, 1985. Citado na página 152.

SHARPE, R. Vocabulary, word formation, and lexicography. In: MANTELLO, F. A. C.; RIGG, A. G. (Ed.). *Medieval Latin*: an introduction and bibliographical guide. Washington, DC: Catholic University of America Press, 1996. Citado 2 vezes nas páginas 93 e 98.

SHAW, J. F. *Contributions to a Study of The Printed Dictionary In France Before 1539*. Toronto: EDICTA, 1997. Citado 4 vezes nas páginas 88, 90, 91 e 92.

SHORT, W. M. Metaphor and the teaching of idioms in latin. In: ONIGA, R.; IOVINO, R.; GIUSTI, G. (Ed.). *Formal linguistics and the teaching of Latin*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2011. p. 227–44. Citado na página 73.

SILVA, A. C. d.; MONTAGNER, A. C. *Dicionário Latino-Português*. Petrópolis, RJ: Editora Vozes, 2009. Citado na página 152.

SILVESTRE, J. P. Palavras tabu e eufemismos nos dicionários de bento pereira e rafael bluteau. In: FERREIRA, A. M. (Ed.). *Percursos de Eros*: Representações do erotismo. Aveiro: Associação Labor de Estudos Portugueses, 2003. p. 223–229. Citado na página 126.

SILVESTRE, J. P.; BORGES, A. A escola lexicográfica de évora: um contributo jesuíta para a reforma pombalina. In: PEREIRA, S. M.; VAZ, F. L. (Ed.). *Universidade de Évora (1559-2009)*: 450 anos de modernidade educativa. Lisboa: Chiado Editora, 2012. p. 323–334. Citado 4 vezes nas páginas 127, 128, 132 e 133.

SIMÕES, A.; GUINOVART, X. G.; ALMEIDA, J. J. Enriching a portuguese wordnet using synonyms from a monolingual dictionary. In: CALZOLARI, N. et al. (Ed.). *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association, 2016. p. 2682–2687. Citado na página 170.

SINCLAIR, J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991. Citado 2 vezes nas páginas 64 e 65.

SOMER, S. *A organização das informações do Aoristo – dicionário eletrônico latim-português*. Dissertação de Mestrado em Estudos da Tradução — Universidade Federal de Santa Catarina, Florianópolis, SC, 2016. Citado na página 20.

SOUZA, F. A. d. *Novo diccionário latino-portuguez*. 2. ed. [S.l.]: Francisco Alves, 1916. Citado na página 144.

SPRINGHETTI, E. Storia e fortuna della grammatica di Emmanuele Alvares. *Humanitas (Coimbra)*, v. 13–14, p. 283–304, 1961–1962. Citado na página 111.

STEFANOWITSCH, A. *Corpus linguistics*: A guide to the methodology. Berlin: Language Science Press, 2020. Citado 3 vezes nas páginas 186, 187 e 188.

STEIN, G. Linking lemma and gloss. In: ____. *Sir Thomas Elyot as Lexicographer*. Oxford: Oxford University Press, 2014. p. 120–142. Citado 3 vezes nas páginas 104, 116 e 173.

STEIN, G. Early polyglot word lists. In: ____. *Word studies in the Renaissance*. Oxford: Oxford University Press, 2017. Citado na página 101.

STEPHANUS, R. *Dictionarium Latinogallicum*. Parisiis: Ex officini Roberti Stephani Typographi Regii, 1538. Disponível em: <https://www.milestone-books.de/pages/books/003100/>. Acesso em: 25 nov. 2024. Citado na página 104.

STEPHANUS, R. *Dictionarium, seu Latinae linguae Thesaurus*. Parisiis: Ex officini Roberti Stephani Typographi Regii, 1543. Citado na página 102.

STEURS, F. et al. *The Future of Academic Lexicography—A White Paper*. Leiden: Instituut voor de Nederlandse Taal., 2020. Disponível em: <https://ivdnt.org/wp-content/uploads/2021/02/The-Future-of-Academic-Lexicography-A-White-Paper.pdf>. Citado 3 vezes nas páginas 31, 159 e 176.

STOK, F. *Appendix Probi IV*. Napoli: Arte tipografica, 1997. Citado na página 83.

STUBBS, M. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell, 2001. Citado na página 65.

SZENDE, T. Problèmes d'équivalence dans les dictionnaires bilingues. In: BÉJOINT, H.; THOIRON, P. (Ed.). *Les dictionnaires bilingues*. Louvain: Aupelf-Uref/Duculot, 1996. p. 111–126. Citado na página 56.

TARP, S. Desafíos teóricos y prácticos de la lexicografía de aprendizaje. In: XATARA, C.; BEVILACQUA, C.; HUMBLÉ, P. (Ed.). *Lexicografia Pedagógica*: pesquisas e perspectivas. Florianópolis: UFSC, 2008. p. 46–73. Citado na página 20.

TARP, S. La ventana al futuro: despidiéndose de los diccionarios para abrazar la lexicografía. *RILEX. Revista sobre investigaciones léxicas*, v. 2, n. 2, p. 5–36, Dec. 2019. Disponível em: <https://revistaselectronicas.ujaen.es/index.php/RILEX/article/view/4964>. Citado 3 vezes nas páginas 19, 31 e 155.

TASOVAC, T. *The Historical Dictionary as an Exploratory Tool*: A digital edition of vuk stefanovic karadzic's lexicon serbico-germanico-latinum. Tese (Doutorado) — School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Dublin, 2020. Citado na página 152.

TEIXEIRA, M. C. Os dicionários no espaço escolar brasileiro. *Revista Interfaces*, v. 15, n. 2, p. 157–78, 2024. Citado 2 vezes nas páginas 148 e 149.

TERRAS, M. Digitization and digital resources in the humanities. In: WARWICK, C.; TERRAS, M.; NYHAN, J. (Ed.). *Digital humanities in practice*. London: Facet Publishing, 2012. p. 47–70. Citado na página 153.

TESNIÈRE, L. *Éléments de syntaxe structurale*. Paris: Librairie C. Klincksieck, 1959. Citado na página 62.

TEUBERT, W. Setting up a lexicographical data-base for german. In: HARTMANN, R. (Ed.). *LEXeter Proceedings*: Papers from the international conference on lexicography at exeter, 9-12 september 1983. Tübingen: Max Niemeyer Verlag, 1984. p. 425–429. Citado na página 36.

TEUBERT, W. Corpus linguistics and lexicography. *International Journal of Corpus Linguistics*, v. 6 (special issue), p. 125–53, 2001. Citado 2 vezes nas páginas 186 e 197.

TIBERIUS, C. et al. Towards the ELEXIS data model: defining a common vocabulary for lexicographic resources. In: KOSEM, I.; CUKR, M. (Ed.). *Electronic lexicography in the 21st century: Post-editing lexicography*: Proceedings of the elex 2021 conference (virtual). Brno, Czech Republic: Lexical Computing CZ, 2021. Citado 2 vezes nas páginas 175 e 176.

TIBERIUS, C. et al. *Introduction to Corpus-Based Lexicographic Practice*. Zagreb: DARIAH-Campus, 2022. Disponível em: <https://elexis.humanistika.org/resource/posts/introduction-to-corpus-based-lexicographic-practice>. Citado 4 vezes nas páginas 187, 197, 198 e 199.

TOLKIEHN, J. (Ed.). *Dosithei ars grammatica*. Leipzig: Theodor Weicher, 1913. Citado na página 85.

TORRINHA, F. *Dicionário Latino-Português*. 2. ed. Porto: Edições Maranus, 1942. Citado 2 vezes nas páginas 145 e 147.

TUFFANI, E. *Repertório brasileiro de Língua e Literatura Latina (1830-1996)*. Cotia,SP: Íbis, 2006. Citado na página 190.

UHLIG, G. (Ed.). *Grammatici Graeci*: recogniti et apparatu critico instructi. Hildesheim: Georg Olms, 1965. Citado na página 83.

VALE, O. A. *Expressões cristalizadas do português do Brasil: uma proposta de tipologia*. Tese (Doutorado) — Universidade Estadual Paulista, Araraquara,SP, 2001. Citado na página 67.

VALLE, R. do. Professor ernesto de faria júnior. *Confluência*, n. 31, p. 11–48, 2006. Citado 2 vezes nas páginas 149 e 150.

VAN DER LOO, M. P. J. The stringdist package for approximate string matching. *The R Journal*, v. 6, p. 111–122, 2014. Disponível em: <https://CRAN.R-project.org/package=stringdist>. Citado na página 181.

VARANTOLA, K. The dictionary user as decision maker. In: *EURALEX 1994 Proceedings*. Amsterdam: Vrije Universiteit, 1994. p. 606–611. Citado 2 vezes nas páginas 30 e 36.

VASCONCELLOS, P. S. Ensino e pesquisa de língua e literatura latina no Brasil de hoje. In: PRATA, P.; FORTES, F. (Ed.). *O latim hoje*: reflexões sobre cultura clássica e ensino. Campinas, SP: Mercado de Letras, 2015. p. 53–68. Citado na página 61.

VELEZ, A. Index totius artis. In: *Emmanuelis Alvari S. J. De Institutione Grammatica Libri Tres*. Eborae: Typographia Academiae, 1744. p. 366–661. Disponível em: <http://purl.pt/14821>. Acesso em: 18.6.2023. Citado 8 vezes nas páginas 111, 112, 113, 162, 172, 181, 182 e 183.

VELEZ, A. Index totius artis. In: VERDELHO, T. dos S.; SILVESTRE, J. P. M. (Ed.). *Corpus Lexicográfico do Português*. Aveiro, Portugal: Universidade de Aveiro and Centro de Linguística da Universidade de Lisboa, 2002. Disponível em: <http://clp.dlc.ua.pt/DICIweb/default.asp?url=Obras&Livro=8&Edicao=9>. Acesso em: 18.6.2023. Citado na página 111.

VERDELHO, T. Historiografia linguística e reforma do ensino. a propósito de três centenários: Manuel álvares, bento pereira e marquês de pombal. *Brigantia. Revista de Cultura*, v. 2, p. 347–383, 1982. Citado 2 vezes nas páginas 121 e 122.

VERDELHO, T. *As origens da gramaticografia e da lexicografia latino-portuguesas*. Aveiro: Instituto Nacional de Investigação Científica, 1995. Citado 7 vezes nas páginas 95, 96, 97, 104, 107, 113 e 116.

VERDELHO, T.; SILVESTRE, J. P. Projecto corpus lexicográfico do português. In: ____. *Dicionarística portuguesa*: inventariação e estudo do património lexicográfico. Aveiro: Universidade de Aveiro, 2007. p. 205–17. Citado na página 155.

VERDELHO, T.; SILVESTRE, J. P. *Lexicografia bilíngue*: a tradição dicionarística português–línguas modernas. Lisboa/Aveiro: Centrto de Linguística da Universidade de Lisboa/Universidade de Aveiro, 2011. Citado na página 144.

VERDELHO, T. dos S.; SILVESTRE, J. P. M. *Corpus Lexicográfico do Português*. Aveiro, Portugal: Universidade de Aveiro and Centro de Linguística da Universidade de Lisboa, 2002. Disponível em: <http://clp.dlc.ua.pt/DICIweb>. Acesso em: 18.7.2023. Citado 2 vezes nas páginas 20 e 172.

VERLINDE, J. B. S. Electronic pedagogical dictionaries: a state of the art. In: XATARA, C.; BEVILACQUA, C.; HUMBLÉ, P. (Ed.). *Lexicografia Pedagógica*: pesquisas e perspectivas. Florianópolis: UFSC, 2008. p. 169–183. Citado na página 20.

VIEIRA, J. L.; HASEGAWA, A. *Dicionário latim-português*: termos e expressões. São Paulo: Edipro, 2016. Citado na página 152.

VIVES, I. L. *De disciplinis libri XX*. Lugduni: Ioannes Frellonius, 1551. Disponível em: <https://archive.org/details/bub_gb_DrZqyAdpLqAC>. Acesso em: 25 nov. 2024. Citado 2 vezes nas páginas 102 e 106.

VOSSEN, P. Wordnet, eurowordnet and global wordnet. *Revue française de linguistique appliquée*, Cairn/Softwin, v. 7, n. 1, p. 27–38, 2002. Citado na página 159.

WELKER, H. A. *Dicionários: uma pequena introdução à lexicografia*. Brasília: Thesaurus Editora, 2004. Citado 15 vezes nas páginas 37, 39, 46, 49, 50, 52, 53, 54, 55, 58, 68, 71, 74, 75 e 173.

WELKER, H. A. A valência verbal em três dicionários brasileiros. *Linguagem & Ensino*, v. 8, n. 1, p. 73–100, 2005. Citado 2 vezes nas páginas 57 e 62.

WHARTON, D. Linguistic semantics and the representation of word meaning in latin dictionaries. In: ONIGA, R.; IOVINO, R.; GIUSTI, G. (Ed.). *Formal linguistics and the teaching of Latin*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2011. p. 255–78. Citado 2 vezes nas páginas 57 e 75.

WICKHAM, H. et al. Welcome to the tidyverse. *Journal of Open Source Software*, v. 4, n. 43, p. 1686, 2019. Citado na página 181.

WIDDOWSON, H. G. *O ensino de línguas para comunicação*. Campinas, SP: Pontes, 1991. Citado na página 29.

WIEGAND, H. E. On the structure and contents of a general theory of lexicography. In: HARTMANN, R. R. K. (Ed.). *LEXeter'83 Proceedings*: Papers from the international conference on lexicography at exeter, 9-12 september 1983. Tübingen: Max Niemeyer Verlag, 1984. (Lexicographica Series Maior, v. 1), p. 13–30. Citado 2 vezes nas páginas 24 e 34.

WIEGAND, H. E. Textual condensation in printed dictionaries: a theoretical draft. *Lexikos*, n. 6, p. 133–158, 1996. Citado 2 vezes nas páginas 33 e 47.

WILKINSON, M. D. et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data*, v. 3, 2016. Citado na página 156.

WOOLDRIDGE, R. Lexicography. In: SCHRIEBMAN, S.; SIEMENS, R.; UNSWORTH, J. (Ed.). *A Companion to Digital Humanities*. Oxford: Blackwell, 2004. Disponível em: <https://companions.digitalhumanities.org/DH/>. Citado 2 vezes nas páginas 153 e 155.

WOOLDRIDGE, T. R. *Le dictionnaire latin-français d'Estienne*. Toronto: [s.n.], 2002. Disponível em: <http://barthes.enssib.fr/translatio/rw/tiden/thesdict/>. Citado 3 vezes nas páginas 102, 103 e 173.

YOUSEF, T.; BERTI, M. The digital fragmenta historicorum graecorum and the ancient greek-latin dynamic lexicon. *Corpus-Based Research in the Humanities (CRH)*, Institute of Computer Science Warsaw, Poland, p. 117, 2015. Citado na página 19.

ZAPIRAIN, B. et al. Selectional preferences for semantic role classification. *Computational Linguistics*, MIT Press, v. 39, n. 3, p. 631–663, 2013. Citado na página 64.

ZAVAGLIA, A. Lexicografia bilíngüe e corpora paralelos: procedimentos e critérios experimentais. *Cadernos de Tradução*, Florianópolis, v. 2, n. 18, p. 19–39, 2006. Citado na página 57.

ZAVAGLIA, C.; SILVA, R. M. Dicionário multilíngüe de regência verbal: uma proposta. *Cadernos de Tradução*, Florianópolis, v. 2, n. 18, p. 67–92, 2006. Citado na página 71.

ZETZEL, J. E. *Critics, compilers, and commentators*: an introduction to roman philology, 200 bce–800 ce. New York: Oxford University Press, 2018. Citado 3 vezes nas páginas 83, 85 e 86.



Esta publicação foi produzida em 29 de maio de 2025
utilizando o sistema de composição tipográfica LaTeX,
com base no modelo de documento abnTeX2,
tipografia Times New Roman e GFS Artemisia,
compilador XeLaTeX, através da plataforma Overleaf.com.