



# Lexical Variation and Standardization in *Prosodia* (1634–1750): Evidence from a Digitally Processed Historical Corpus

## Research Article

Correspondence: Helena Freire Cameron  
<[helenac@ipportalegre.pt](mailto:helenac@ipportalegre.pt)>

Associate Professor, Polytechnic University of  
Portalegre; University of Évora – CIDEHUS,  
Portugal

## Publication Details

**Received:** June 25, 2025

**Accepted:** November 25, 2025

**Published:** November 30, 2025

## Abstract

This paper examines linguistic innovation and lexical variation in the bilingual Portuguese Jesuit dictionary *Prosodia*, published between 1634 and 1750. Through an analysis of the two dictionaries, the Latin-Portuguese and Portuguese-Latin, components of the lexicographic set, the study reveals how *Prosodia* not only reflected but also contributed to the development and stabilization of the Portuguese language at the end of the 17th century. The research draws on the *Corpus Lexicográfico do Português* and applies digital humanities methodologies, described in the article, to uncover patterns of lexical change, the coexistence of archaic and innovative forms, and the editorial strategies underlying the dictionary's evolution. Three case studies—variants that coexist in the *Prosodia corpus* (e.g., the words for 'knee' and 'fruit'), predominance of a variant over the other (e.g., words ending in -airo), and registration of the renovated form mentioning the disused one (e.g., 'aspeito')—illustrate the persistence and eventual selection of standard forms in Portuguese. The findings highlight *Prosodia*'s value as a resource for documenting linguistic renovation and its lasting influence on subsequent Portuguese lexicography.

**Keywords:** Portuguese language, Lexicography, Jesuit dictionaries, Linguistic innovation, Corpus Lexicográfico do Português



## 1. Introduction

Bilingual Latin-European dictionaries serve not only as lexical collections but also as invaluable testimonies to linguistic change. By systematically recording word forms, meanings, and usage across different periods, dictionaries provide direct evidence of how languages evolve, document the coexistence of archaic and innovative forms, and reveal the processes of lexical selection and standardization. Their entries capture not only vocabulary but also the cultural and historical contexts in which words were used, turning each dictionary into a testimony of linguistic and social change. This makes historical dictionaries essential sources for tracing and understanding the dynamics of language development over time.

It is important to note that lexicographic work on Portuguese appeared later than in other European countries. While major European lexicographic traditions had their origins earlier, systematic Portuguese dictionary production began only in the 16th century (Verdelho, 2002), placing Portuguese lexicography somewhat behind its European counterparts in terms of chronology and development.

The *Prosodia*, compiled by Jesuit Bento Pereira in the 17<sup>th</sup> century, includes the Latin-Portuguese *Prosodia*, the Portuguese-Latin *Tesouro*, and additional para-lexicographic works, establishing it as one of the largest bilingual lexical *corpora* in Portuguese lexicography. Published continuously from 1634 to 1750, it functioned as a principal educational guide for Latin and Portuguese in Jesuit schools. The seventh edition, published in 1697, represents the peak of its editorial history and features significant lexical renewal by Jesuit Matias de S. Germano in the Latin-Portuguese dictionary. The volume comprises approximately 75,000 Latin-Portuguese entries and 24,000 Portuguese-Latin entries (Cameron, 2012). Due to its longevity, importance in Portuguese lexicography, and the breadth of its *corpus*, this dictionary serves as a primary source for identifying key markers of linguistic renovation in Portuguese.

The *Prosodia*'s extensive documentation of lexical items, including both archaic and innovative forms, allows researchers to observe the dynamics of vocabulary change and the processes of standardization that shaped modern Portuguese. Because the entries and words inside definitions span more than a century and reflect multiple editorial interventions, the dictionary also captures shifts in usage, thus offering a unique lens into the language's evolution registration. Moreover, the set provides valuable evidence of the state of Portuguese at the end of the 17<sup>th</sup> century, showing not only which forms persisted but also which were in obsolescence. As such, *Prosodia* stands as a testament to the role of lexicographic works as historical witnesses to linguistic change, revealing broader social and cultural transformations as reflected in the language.

This article aims to explore the role of the Jesuit bilingual dictionary *Prosodia* (1634–1750) in documenting linguistic innovation in Portuguese during the late 17<sup>th</sup> century. By analysing the coexistence and selection of lexical variants within *Prosodia* and comparing their registration in prior and subsequent dictionaries, the study seeks to uncover patterns of lexicographical registration of language change in early modern Portuguese, presenting enlightening examples of lexicographic registration of forms whereas there is a coexistence of archaic and renovated forms, but also a predominance of renovated forms *versus* old variants, that evaluate the dictionary's significance as a historical testimony to linguistic evolution.

Furthermore, the research demonstrates the value of digital humanities methodologies in lexicographic *corpora*, enabling the construction of pre-contemporary linguistic resources.

Portuguese is a language spoken on every continent, a result of both History and the diaspora. Currently, the Portuguese language has two distinct norms: European Portuguese and Brazilian Portuguese. African norms are also under development. The convergence of the two main norms is very high, differing primarily in some syntactic and orthographic issues. However, worldwide, the Portuguese language is still under-represented on the internet, especially in open-access, fully searchable linguistic resources, and even more scarce in the historical domain. The construction of linguistic *corpora*, especially in pre-contemporary stages of the Portuguese language, is essential. For that, studies based on historical linguistic *corpora*, especially those built from dictionaries, are a way to promote and enlarge the online presence of the Portuguese language. The discussion of linguistic issues regarding lexicographic evidence, especially using the rich *Prosodia corpus*, is an opportunity to try to contribute to this dissemination.

To better give notice of examples that are not in contemporary stage of the language, in this article, the examples from dictionaries maintain the original spelling and lexicographic registration, in Latin and in Portuguese. No standardization was applied to forms registration. However, regarding archaic and renovated forms, to better enlighten the reader, we updated the subtitles' graphical accentuation in the forms still in use. Also, as the article is written in English, we propose translation of the Portuguese examples. Regarding the translation of Camões' verses, we resort to the Translation of Richard Francis Burton (1880), not updating English spelling proposed by the translator.

## 2. Digital Humanities and old bilingual Portuguese lexicographic *corpora*

Texts from earlier stages of languages present significant challenges for both historians and language researchers. Access to these texts is often limited, as many historical documents remain in library collections with restricted availability, sometimes subject to conservation policies that prevent handling or reproduction.

The automatic processing of ancient texts remains a challenge for Digital Humanities research (Finatto, Quaresma, & Gonçalves, 2018). Although considerable efforts have been made to digitize patrimonial books, old dictionaries, and other historical documents, many are only available digitally in image format. This complicates both human and machine reading, as texts are sometimes digitized from earlier supports such as microfilm, thereby diminishing the image quality, which is essential for perceiving document details. Also, many digital documents are no longer available in libraries for in-person consultation, making it difficult to read when they contain bleed-through, erasures, or other markings on paper.

Furthermore, the diversity of scripts, orthographies, and linguistic conventions used in earlier periods adds additional layers of complexity to their interpretation and analysis. The spelling pattern differs from the contemporary one, and by the end of the 17<sup>th</sup> century, the Portuguese language had many spelling variants, making it more difficult to access and process information (see Cameron et al. (2023) regarding an 18<sup>th</sup>-century Portuguese textual collection). These factors collectively hinder the complete access to texts and highlight the necessity of digitization,

preservation, and the development of specialized tools for processing and studying historical language data.

Applying Natural Language Processing (NLP) techniques to non-contemporary texts introduces additional difficulties. Earlier language stages often exhibit spelling, lexical, and syntactic variation, complicating machine processing. Nevertheless, the application of NLP methods offers new perspectives for linguistic studies (Banza & Gonçalves, 2018), and the benefits of such approaches for historical texts are substantial (Gonçalves & Banza, 2013).

Another challenge in Digital Humanities regarding historical documents concerns the FAIR principles. The documents must be Findable, Accessible, Interoperable, and Reusable (see Wilkinson, 2016), so they can be fully searchable in the digital online world. Most digitized historical documents don't fulfil the Fair principles, making it more difficult to achieve a global approach to textual data and hindering human and machine usability. Also, in an open linked science, if data are not formatted according to FAIR principles, they cannot really have a truly digital presence. This reinforces the effort made in the transcription of *Prosodia*, making it findable and accessible, at least, as described below.

To build a digital version of *Prosodia*, some activities were developed, not merely making a support transfer but transforming the paper 17<sup>th</sup>-century edition in a fully digital searchable *corpus*, as detailed. Regarding the Latin-Portuguese dictionary in *Prosodia*, the author transcribed and organized a manually transcribed digital version, as detailed in (Cameron, 2012). Although some prior attempts were made to capture the texts using OCR, the results at that time were not accurate enough to allow the constitution of a fully digitally process-able version of the dictionary. For that reason, the dictionary was fully manually transcribed.

Regarding the transcription process, it was issued as a preferential diplomatic edition of the complete lexicographic text of *Prosodia*, preserving all graphic variants. In the edition made, it was maintained the variation in the vocal or consonantal use of <j> and <v> and in nasal diphthongs (-am/-ão/-aõ). The use of double consonants without linguistic value or pseudo-etymological spelling was also preserved. Also, '&' with the value of 'et' was maintained.

To be processed automatically by a machine, we put occasional spaces to separate words that, for typographic reasons, are typed together in the definitions. Also, the two languages, Latin and Portuguese, were conventionally separated in a working version built solely for computational processing, to allow searching for Latin or Portuguese words separately.

The corpus was processed using the Concordancer tool *AntConc*<sup>1</sup>. It is a widely used open-source tool that can recognize Latin characters like 'ç' and graphic accents like 'ã', 'õ', and others that are characteristic of the Portuguese language. With AntConc, we generate three types of lists: alphabetical, frequency ranking, and ordered by the end of the word. At the same time, we generate a concordance *corpus* containing all Portuguese words in context. The *corpus* was not lemmatized. It contains 703,725 occurrences and 46,067 single Portuguese words. (See Cameron, 2012). From the Portuguese *corpus*, we build a Dataset to better process lexical data.

<sup>1</sup>AntConc - A freeware corpus analysis toolkit for concordancing and text analysis, available at <https://www.laurenceanthony.net/software/antconc/>

The Latin-Portuguese version of the *Prosodia* was later integrated into the digital collection of the *Corpus Lexicográfico do Português*<sup>2</sup> (Portuguese Lexicographical Corpus), a project of the University of Aveiro that aggregates major bilingual Latin-Portuguese-Latin dictionaries. The *Prosodia* headwords can be consulted both in the Latin entries and in the Portuguese words within the definitions.

To access the lexical mass of other Portuguese bilingual dictionaries, this study used the *Portuguese Lexicographic Corpus*, where headwords are fully searchable via the DiCiWeb Platform, and the results obtained show the complete entry of each dictionary. This resource provides comprehensive access to dictionary entries and all words within their definitions, choosing Latin, Portuguese, or other languages present in the *Lexicographic Corpus*.

The digital transformation of the *Prosodia corpus* was essential not only for the development of studies that take advantage of computational tools and resources but also for making *Prosodia* available to linguists, historians, and the broader public.

### 3. The lexicographic set of *Prosodia* (1634-1750)

Bento Pereira, S.J., is one of the great lexicographers of Portuguese lexicography (Verdelho, 2002). He was born in Borba in 1605, in the Alentejo region of southern Portugal. At 15, he joined the Society of Jesus. After his novitiate in Lisbon, he became a Professor in the University of Évora, where he teaches Humanities and Rhetoric, and where he also started a very fruitful production of didactic texts, manuals, and dictionaries, for teaching Latin and Portuguese Language, History, Philosophy, theology, etc. (Verdelho, 1992).

In 1634, he published, at his own expense, the Latin-Portuguese-Spanish *Prosodia*, a trilingual volume containing about 50,000 entries (Cameron, 2012). The Latin words are all indexed alphabetically, and each of them has the corresponding Portuguese term, the citation of the term in Latin authors, and the prosodic information of the Latin term. The Spanish terms are rare in all the dictionary.

In 1647, Pereira published the Portuguese-Latin dictionary *Tesouro*. It was published 13 years after the 1<sup>st</sup> dictionary, *Prosodia*, but we believe it was already conceived earlier, during the elaboration of the Latin-Portuguese dictionary, as editorial licences may enlighten. This time gap was a special and unstable period in the History of Portugal, marked by the Restoration of Independence in 1640, when Portugal regained its independence from the Spanish, who had dominated Portugal since 1580.

Bento Pereira also covered the cost of the *princeps* edition of *Tesouro* (Cameron, 2018). In this dictionary, Portuguese entries are listed in alphabetical order, frequently interrupted by the insertion of lexical families in certain words. All entries have the corresponding Latin term. The nomenclature of *Tesouro* remained unchanged until the last edition in 1750. The headwords were collected among Portuguese reference authors, accordingly to the “Authors List” of the *princeps* edition.

<sup>2</sup>*Corpus Lexicográfico do Português*, at <http://clp.dlc.ua.pt/Corpus.aspx>

From the second edition, in 1653, both dictionaries were published together in the same volume, and, from the 3rd edition, also with the para-lexicographic work *Florilegio* (that had a former and single edition in 1655), these works formed an important lexicographic set known as *Prosodia*, published till 1750, for 12 editions, published in Lisbon, by the Craesbeeck House, and in Évora, by the University of Évora. The reunion of these dictionaries in this large volume certainly serves didactic purposes, allowing students to come into contact with the Latin language, supported by the Portuguese entries of *Tesouro*, all in the same volume (Verdelho, 1992).

Bento Pereira died in 1681, but the editorial life of *Prosodia* continued, receiving a significant revision in its seventh edition in 1697 by the Jesuit Matias de S. Germano, in the Lexicographic school, in University of Évora. The few Spanish terms were removed from the glosses, and the dictionary was presented as bilingual. The Latin nomenclature was notably expanded with “barbarisms” and Latin forms not authorized in classical tradition and collected in European Latin dictionaries. The Portuguese definitions were broadened, receiving not only more meanings but also encyclopedic vocabulary. It is worth to note that many spelling variants were introduced inside Portuguese definitions.

The 1697 edition represents the editorial zenith of this work, establishing *Prosodia* not merely as a pedagogical manual for the teaching of Latin and Portuguese but as a lexicographical reference for both languages, unrivalled for many years in the tradition of Portuguese lexicographical publishing. This substantial edition comprises 932 folio pages.

The Jesuits were already preparing another major reformulation of the work when the volume was forbidden and then ordered burned, in accordance with the new policies of the Marquis de Pombal, the Prime Minister of King D. José, in mid-18<sup>th</sup> century (Cameron, 2018). The Jesuits were expelled from Portugal by the King in 1759, and the Society of Jesus was suppressed by Pope Clement XIV in 1773. With this order, two Jesuit teaching manuals, *Prosodia* by Bento Pereira and *Grammatica* by Manuel Álvares, were prohibited and ordered destroyed (Mendes de Almeida, 1967).

*Prosodia* was, at that time, an important resource in the teaching mission of the Portuguese and Latin languages in the Society of Jesus, not only in Portugal but also in Brazil and in the Orient, within the Society's mission spaces (Cameron, 2012). However, *Prosodia* left an important heritage in the subsequent monolingual Portuguese Lexicography. After the order of destruction of *Prosodia*, Portugal had no Manuals for teaching Latin, and Professor Pedro José da Fonseca was asked to make a new dictionary, publishing *Parvum Lexicon* in 1762, where the Portuguese entries of *Tesouro* are easily perceived in the nomenclature, perpetuating the “old” *Prosodia*, *mutatis mutandis* (Borges, 2011).

The *Prosodia* survived the destruction order, and many volumes can still be found in national and municipal libraries, both public and private. When the destruction order was issued, some volumes had their title pages torn out, presumably to deceive the inspectors who came to collect the copies. According to Cameron (2018), consulting online catalogues, there are still more than 100 volumes of all the twelve *Prosodia* editions in libraries' collections in Europe, North America, and South America.

#### 4. Notes on the lexicographic registration of innovation

By the end of the 17<sup>th</sup> century, the Portuguese language was marked by linguistic stabilization, significant lexical innovation, and extensive lexical variation. This era witnessed a robust influx of neologisms (Verdelho, 1987). During this period, several linguistic features stabilized (Cardeira, 2006). Lexical forms exhibited considerable diversity at the linguistic level (Cardeira & Mateus, 2008), with both archaic and innovative forms coexisting (Teyssier, 1997), and at the orthographic level, with numerous orthographic and typographic variants (Gonçalves, 2003).

The lexicographical set *Prosodia* documents both the cultural and linguistic memory of a period of stabilization and the renovation of Portuguese. Its significance lies in revealing forms not recorded in subsequent dictionaries, as well as forms or variants first registered in *Prosodia*. This volume provides valuable insight into linguistic change, as it contains two dictionaries produced at different times but compiled in a single volume: the Portuguese definitions in the Latin-Portuguese *Prosodia* were extensively updated in the seventh edition of 1697, incorporating new or revised forms absent from earlier editions, while the Portuguese nomenclature in the Portuguese-Latin *Tesouro* remains essentially unchanged from the 1647 autonomous edition. The temporal gap between these two lexicographic works clearly illustrates the linguistic renovation of the period.

From our perspective, *Prosodia* serves as a significant marker in lexicographic records, illuminating the language's development. First, it illustrates how obsolete and renewed forms coexist within the same dictionary. Second, *Prosodia* signifies a pivotal change in lexicographic practice, as it abandons the recording of disused forms. Finally, while some obsolete forms are still referenced, they are no longer included as main entries in *Prosodia*'s nomenclature.

Regarding the coexistence of obsolete/renovated forms, we use the examples of 'joelho' and 'fruto', each with its variants.

Concerning the preference for lexicographic registration of the renovated form to the detriment of the old form, from the *Prosodia* dictionary and subsequent dictionaries, we used examples of words ending in -airo, such as 'boticário', 'calendário' and 'vigário', and their respective variants.

The 3<sup>rd</sup> matter concerns the inclusion of renovated forms while maintaining the mention of disused words. In this matter, we use one example from a previous well-known literary work: the word 'aspeito', from *Lusíadas* (1572).

#### 4.1 Coexistence of obsolete and renovated forms

##### 4.1.1 Geolho/giolho/joelho/juelho ['knee']

The actual Portuguese word for 'knee' is 'joelho'. During the History of the Portuguese language, others occurred, like 'geolho' and 'giolho', but only 'joelho' is in use. We can find all three forms in the *Prosodia* lexical corpus, but with different registrations.

- (1) Genu, u, n. g. O joelho, alii geolho.[...] (Pereira-Prosodia, 1697)
- (2) Giolho. Genu, Indekl. (Pereira-Tesouro, 1647)

The form ‘giolho’ appears only once in the Portuguese-Latin *Tesouro*. The forms ‘geolho’ and ‘joelho’ are recorded exclusively in the Latin-Portuguese dictionary, with five and twenty-three occurrences, respectively. Notably, the form ‘joelho’, which remains in use today, already had the highest frequency.

To establish the historical trajectory of these forms, a previous dictionary, Cardoso (1569-1570), and three subsequent dictionaries—Bluteau (1712-1728), Folqman (1755), and Fonseca (1798)—were also consulted. Another variant, ‘juelho’, appears in two 18<sup>th</sup>-century dictionaries, Bluteau and Folqman, but is no longer in use. The following table summarizes the registration of these variants in the six dictionaries considered:

Table 1: Giolho/geolho/joelho/juelho

	<b>Jerónimo Cardoso</b> <i>Dictionarium</i> (1560-70)	<b>Bento Pereira</b> <i>Tesouro</i> -1647	<b>Bento Pereira</b> <i>Prosodia</i> -1697	<b>Bluteau</b> <i>Vocabulario</i> (1712-28)	<b>Folqman</b> <i>Diccionario</i> -1755	<b>Fonseca Parvum lexicon</b> -1798
<b>giolho</b>	X	x		x		
<b>geolho</b>			x			
<b>joelho</b>		x	x	x	x	x
<b>juelho</b>				x	x	

The variant ‘giolho’ has only 3 occurrences in the Bluteau dictionary, but in one of them, the term ‘juelho’ is registered as equivalent to the variant:

- (3) Giolho. Vid. Juelho, [...] (Bluteau, 1712-1728)

The variant ‘joelho’ has its first lexicographic occurrence in 1647 and is included in the nomenclature or definitions of all subsequent dictionaries. On the contrary, ‘giolho’ is registered till the first half of the 17<sup>th</sup> century, with only three occurrences in Bluteau. The form ‘geolho’ is registered only by Bento Pereira in 1697, and the variant ‘juelho’ is noted only by Bluteau and Folqman.

#### 4.1.2 Fruito/ fructo/fruto [fruit]

In Portuguese, the variants ‘fruito’ and ‘fructo’ have existed, but the current standard form is ‘fruto’. Their documentation in dictionaries helps to clarify the historical usage of each form.

Cardoso (1569-79) and Bento Pereira, in *Prosodia*, both include ‘fruito’ in nomenclature or inside definitions:

- (4) Fruito. fructus(us). (Cardoso 1569-70)

- (5) Acarpia, ae, f. g. A falta de fruito.[...] (Pereira- *Prosodia*, 1697) [‘Thelackoffruit’]

Regarding the variant ‘fructo’, it is not present in the Portuguese-Latin Bento Pereira’s *Tesouro* (1647) or in Jerónimo Cardoso, but it is written inside the definitions of *Prosodia* (1697):

- (6) Morum, i, n. g. A amora, fructo da amoreira. [...] (Pereira-Prosodia, 1697) [‘Blackberry, fruit of mulberry’]

The form ‘fruto’, the only one that is still in use, has a lexicographic register in *Tesouro* and *Prosodia* and in previous and subsequent dictionaries:

- (7) Fructuosus(a. um). Cousa que traz fruto. (Cardoso, 1569-70) [‘fruitful, thing that has fruit’]
- (8) Flor da romãa, que nam dá fruto. Balaustium, ii. (Pereira-Tesouro, 1647) [‘flower of pomegranate, that doesn’t have fruit’]
- (9) Cariota, ae, f. g. A tamara, fructo da palma. [...]. (Pereira-Prosodia, 1697) [‘date, fruit of the palm tree’]
- (10) Abrunho. Fruto do abrunheiro. Prunum silvestre. [...] (Bluteau, 1712-1728) [‘Sloe. Fruit of the blackthorn tree’]
- (11) Alaternus, i, f. Plin. Certa arvore ou arbusto, que não produz semente, nem fruto. (Fonseca, 1798) [‘Certain tree or bush, that doesn’t produce seed nor fruit’]

This example demonstrates the coexistence of archaic and renewed forms, with only one variant persisting in contemporary usage.

## 4.2 Predominance of renovated form vs archaism in forms in –airo/-ario

The Portuguese forms that end in ‘–airo’ are no longer used in contemporary Portuguese. The renovated variants in ‘–ário’ are the ones that are used nowadays, but they coexisted for a long time. For that, it is important to note their lexicographic register to perceive the approximate date of the tendency of predominance of the contemporary form. We choose three words, very productive: boticário ‘apothecary’, calendário ‘calendar’, and vigário ‘vicar’, and their related disused forms, respectively boticairo, calendairo, and vigairo.

### 4.2.1 Boticairo/boticário [apothecary]

The words boticairo/boticario coexisted for some time. The dictionaries clarify their use and, importantly, when lexicographers dropped one of the forms from their main entries. For Cardoso, in his *Dictionarium*, and Bento Pereira, in *Tesouro*, both recorded the archaic forms:

- (12) Pharmacus(i). Ho feiticeiro. ou boticairo. (Cardoso, 1569-70) [‘the sorcerer, or apothecary’]
- (13) Boticairo, ou boticario. Fharmacopola, &Aromatarius. (Pereira-Tesouro, 1647) [‘apothecary’]

However, Bento Pereira, in *Tesouro*, writes both forms in the nomenclature, as equivalents. It’s worth noting that the Portuguese lexicographer Agostinho Barbosa, in his *Dictionarium*, in 1611, already includes the renovated form in the nomenclature:

- (14) Boticario. Pharmacopola, ae, [...]. (Barbosa, 1611)

*Prosodia*, inside the definitions, registers only the renovated form -ario:

- (15) *Medicamentarius*, ii, m. g. O boticario, ou Medico. [...] (Pereira-Prosodia, 1697) [‘the apothecary, or the Doctor’]

*Prosodia* is the first Latin-Portuguese dictionary to omit the obsolete form, registering only the updated variant. Subsequent dictionaries also record only the renovated form:

- (16) *xeringa*, *ouxeringa*. Instrumento de Boticario. Vid. *Seringa*. [...] (Bluteau, 1712-28) [‘syringe. Instrument of an apothecary’ ]  
 (17) *Medicamentariu*, ii, m. *Plin*. Boticario. (Fonseca, 1798)

This matter had been discussed in the literature at that time. Madureira Feijó, in the monumental work *Orthographia* (1734), condemns the use of the archaism, defending the use of the renovated form: “Botíca, Boticário, e naõBoticairo.”

#### 4.2.2 *Calendairo/calendário* [calendar]

Regarding this pair of forms, the desused form ‘*calendairo*’ is part of Cardoso and Bento Pereira’s *Tesouro* nomenclature:

- (18) *Calendairo*. *Fasti(orum)*. (Cardoso, 1569-70) [‘calendar’]  
 (19) *Callendairo de santos*. *Fasti, orum*. *Calendarium*, ii. (Pereira-Tesouro, 1647) [‘saints calendar’]

In contrast, the same volume of *Prosodia* registers only the updated form ‘*calendario*’ in its Latin-Portuguese dictionary:

- (20)\* *Calendarium*, ii, n. g. *Calendario*, ou folhinha de anno. 1. b. 3. 1. Item, livro dos cambios. [...] (Pereira-Prosodia, 1697) [‘Calendar or papersheet per year. Also, exchange book’ ]

As with *boticairo/boticario*, only the updated form ‘*calendário*’ is recorded in the lexicographic register from 1697 onward.

- (21) *Lunário*. *Calendario*, que conta por Luas. [...] (Bluteau, 1712-28) [‘Lunar. Calendar that counts by the moons’]  
 (22) *Calendarium*, ii, n. *Sen*. *Calendario*, livro, em que se escrevião as usuras para se cobrarem nas *Calendas* do mez. (Fonseca, 1798) [‘Calendar, book where people write usury to be collected in “Calendas” time of the month [first day]’]

Once again, Agostinho Barbosa’s *Dictionarium* (1611) records the updated form over 30 years before *Tesouro*:

- (23) *Calendario dos meses*. *Calendarium*, ii, [...]. (Barbosa, 1611)

#### 4.2.3 *vigairo/vigário* [vicar]

The recording of the third pair of archaic and updated forms in 16th-century bilingual dictionaries reveals subtle differences from the prior two pairs.

The archaism ‘*vigairo*’ has a lexicographic register in Cardoso and in Pereira’s *Tesouro*:

- (24) Vigairo. vicarius(i). (Cardoso, 1569-70) ['thevicar']  
 (25) Vigairo. Vicarius, ii. (Pereira-Tesouro, 1647) ['thevicar']

The *Prosodia* Latin-Portuguese dictionary registers only the renovated form 'vigário':

- (26) Petrus, i, m. g. Pedro nome de homem(S. Pedro Principe dos Apostolos,[ 498. 2] Vigario de Christo, & c. Claud. Epigr. NutantemquatitundaPetrum, cuiChristus in alto. (Pereira-Prosodia, 1697) [Peter, name of a man (Saint Peter, Prince of the Apostles), Vicar of Christ]

In contrast to previous examples, Bluteau's *Vocabulario* registers both the archaic and updated forms. A closer look reveals 'vigairo' appears 10 times, while the updated 'vigário' occurs 29 times, indicating his use of both as equivalents:

- (27) vigairo, ou vigario, ou vicario. O que faz as vezes, &funçoens do Prelado na sua ausencia. Vicarius, ii. [...] (Bluteau, 1712-1728). ['The vicar. Who acts like a Prelate in his absence']

However, the greater number of occurrences of the renovated form suggests a predominance of this form over the obsolete form.

Fonseca includes the renovated form in its dictionary:

- (28) Minerva, ae, f. Cic. Minerva, [...] prov. Ensina o Padre nosso ao Vigario (Fonseca, 1798) ['teachingtheLord'sPrayer to thevicar']

These three examples highlight the importance of the *Prosodia* Latin-Portuguese dictionary as a turning point in lexicographic registration; from this edition onward, subsequent dictionaries omit archaic forms and focus on those that persist to the present.

### 4.3 Old forms mentioned as equivalent - aspeito/aspeto

The word 'aspeito', meaning 'aspect', is not used in present-day Portuguese but appears in Camões's epic work *Os Lusíadas*, specifically in the episode of "The old man of Belem":

- (29) Mas um velho d'aspeito venerando,  
 Que ficava nas praias, entre a gente  
 [...]

(Camões, 1572, Canto IV, stanza 94)

['But now an agèd Sire of reverend mien, /Upon the foreshore throngèd by the crowd' [...] ]  
 (Camões, 1880)

The disused form 'aspeito' is registered by Cardoso (1569-70)

- (30) Aspeito. Aspectus(us). (Cardoso, 1569-70) ['aspect']

However, Cardoso was the last Portuguese lexicographer to include only this form in the nomenclature. Bento Pereira, in *Tesouro* (1647), and Bluteau (1712-1728) both mention the form, but they propose equivalence to the renovated form, in the same article:

- (31) *Aspeito*, id est, *aspecto*. *Aspectus*, us. *Forma*, ae. (Pereira-Tesouro, 1647).  
 (32) *Aspeito*. V. *Aspecto*. [...] (Bluteau, 1712-28)

The Prosodia Latin-Portuguese dictionary no longer includes the old form inside the definitions, as we can observe in one of the occurrences:

- (33) *Agriomorphus*, i, m. g. O que se he de *aspectorustico* (Pereira-Prosodia, 1697) [‘the one who has a rustic appearance’]

All the subsequent dictionaries register the renovated form. Cardoso was the last lexicographer to include only this archaism in the nomenclature.

## 5. Conclusions

Dictionaries, particularly historical bilingual editions, provide valuable evidence for tracing the register of a language's development over time. Their systematic organization and long publication history make them indispensable tools for linguists, historians, and philologists. Dictionaries collect individual word forms and meanings, alphabetized. In traditional print versions, linguistic information is readily accessible in alphabetized entries but is more difficult to locate within definitions or to analyse systematically at scale.

Searching inside the definitions reveals that they also document spelling and usage changes, the appearance and disappearance of variants, and the gradual process of language standardization. Therefore, a Digital Humanities approach is essential for gaining access to and analysing linguistic data inside the lexical mass of dictionaries, especially the historical ones. This process involves several stages: transferring the material to a digital environment, applying Natural Language Processing tools to identify lexical information in context throughout the text, and using computational methods to process large amounts of lexical information that would be difficult to collect manually.

Using the data collected in the heritage lexicographical corpora in *Corpus Lexicográfico do Português*, it was possible to compare entries across multiple editions and dictionaries and to reconstruct the trajectory of lexical items and identify when innovations were adopted or archaisms abandoned.

The *Prosodia corpus* plays a crucial role as a historical and linguistic resource for understanding the development of Portuguese. As a bilingual dictionary compiled and revised, *Prosodia* documents lexical items, offering a unique perspective of early modern Portuguese. Its Portuguese entries in *Tesouro* and lexical items inside the definitions of Latin-Portuguese *Prosodia* capture not only the vocabulary in use at various points in time but also reflect editorial decisions regarding which forms to include, update, or discard—thus revealing the processes of standardization and innovation that shaped the language.

A close analysis of the *Prosodia* lexicographic set reveals a coexistence of archaic and renovated forms, offering a microcosm of registration of linguistic transition in early modern Portuguese. Unlike earlier dictionaries, such as Cardoso (1569–70), which often register only archaic forms, or later dictionaries like Fonseca (1798), which embrace mainly the renovated forms, *Prosodia* sits at a transitional point. It documents both variants within the same entries or definitions, providing direct evidence of a period where linguistic norms were not yet fully stabilized.

For example, in the case of the word for ‘knee’, *Prosodia* registers ‘geolho’ and ‘joelho’ in the Latin-Portuguese section, while the Portuguese-Latin *Tesouro* includes the even older ‘giolho’. The frequency data show ‘joelho’—the contemporary form—gaining predominance, but the persistence of ‘geolho’ and the isolated registration of ‘giolho’ reflect the ongoing variation. In comparison, Bluteau (1712–28) still records several variants, but Fonseca (1798) restricts itself to the modern form, indicating normative consolidation.

Similarly, with ‘fruito’/‘fructo’/‘fruto’ (‘fruit’), *Prosodia* records all forms in different roles, some as headwords, others within definitions. Earlier dictionaries like Cardoso also include ‘fruito’, while subsequent ones, such as Bluteau and Fonseca, progressively limit the register to ‘fruto’.

A particularly telling trend appears in forms ending in –airo/–ário, such as ‘boticaíro/boticário’ and ‘calendairo/calendário’. *Prosodia* is the first major Latin-Portuguese dictionary to omit the archaic ‘-airo’ variants and register only the updated forms ‘-ário’. Bluteau and Fonseca also registered the renovated forms. This editorial choice may signal a shift toward standardization and reflects the broader linguistic movement of the era.

The forms ‘aspeito’ and ‘aspeto’ were recorded in Portuguese dictionaries. This issue highlights the transition from documenting both forms to standardizing only the renovated form, illustrating the process of lexical change and the shift in lexicographic practice from Cardoso through *Prosodia* and later dictionaries. This analysis clarifies how *Prosodia* and subsequent dictionaries are privileged evidence in reflecting lexical changes.

In summary, *Prosodia* occupies a pivotal role: it bridges the documentation styles of older dictionaries, which still collect archaisms, and later ones, which reflect linguistic renovation and standardization. By recording coexistence and the gradual predominance of renovated forms, *Prosodia* offers unique insight into the dynamics of language change—a value further enhanced when compared systematically with the lexical choices of its predecessors and successors.

This work also allowed a comparative lexical analysis of the lexicographic registers in some heritage Portuguese dictionaries, where it indicates the *Prosodia* lexicographic set as a boundary. Its contribution to the study of Portuguese linguistics is substantial, not only as a record of language in transition.

Furthermore, the breadth and depth of the *Prosodia corpus* make it invaluable for comparative studies. By providing a rich dataset for digital analysis, the *corpus* supports the application of computational tools to identify patterns of lexical change registration. In this way, *Prosodia* serves not only as a testimony to linguistic evolution but also as a foundation for interdisciplinary research that bridges traditional philology and digital humanities, advancing our understanding of the Portuguese linguistic history.

**Funding:** This study was not funded in any shape or form by any party.

**Conflict of Interest:** The author declares that he has no conflict of interest.

**Bio-note:**

**Helena Freire Cameron** is Associate Professor at the School of Education and Social Sciences of the Polytechnic University of Portalegre, Portugal, and an integrated researcher at the ChronosLab of CIDEHUS — Center for Studies in History, Cultures and Societies of the University of Évora. With a PhD in Linguistics (University of Aveiro) and a post-doctorate in Digital Humanities and Natural Language Processing (University of Évora), she has extensive experience in digital editing and computational processing of Portuguese historical texts (18th century), with a special focus on orthographic variation and the annotation of named entities, contributing to the development of corpora and lexicographic resources fundamental to Digital Humanities. <https://orcid.org/0000-0001-7719-6994>

**References**

- Anthony, L. (2024). AntConc (Version 4.3.1) [Computer Software]. Tokyo, Japan: WasedaUniversity. <https://www.laurenceanthony.net/software/AntConc>
- Banza, A. P., & Gonçalves, M. (2018). *Roteiro de História da Língua Portuguesa*. (U. C.-H. Heritage, Ed.) Évora: Universidade de Évora.
- Barbosa, A. (1611). *DictionarivmLvsitanicoLatinvm*. Braga: Frutuoso L. de Basto.
- Bluteau, R. (1712-1728). *Vocabulário Portuguez e Latino*. Coimbra: Colégio das Artes da Companhia de Jesus.
- Borges, A. M. (2011). *Pedro José da Fonseca e a sua obra lexicográfica*. Aveiro: Universidade de Aveiro.
- Cameron, H. F. (2012). *A Prosodia de Bento Pereira: Contributos para o estudo lexicográfico e filológico*. Aveiro: Universidade de Aveiro.
- Cameron, H. F. (2018). *O conjunto lexicográfico Prosodia (1634-1750), de Bento Pereira, S.J.* (Vol. 7). Évora: Coleções do CIDEHUS, Coleção Fontes & Inventários - série geral, available at: <https://books.openedition.org/cidehus/3321>
- Cameron, H.F., Olival, F., Vieira, R. (2023) Planear a normalização automática: tipologia de variação gráfica do corpus das Memórias Paroquiais (1758), *LaborHistórico*, Rio de Janeiro, ISSN 2359-6910, vol.9, n.1. DOI: 10.24206/lh.v9i1e52234
- Camões, L. V. (1572). *Os Lusíadas*. Obtained from: <https://oslusíadas.org>
- Camoës, L. V. (1880). *The Lusíads*. (R. F. Burton, Trad.) London: Tinsley Brothers.
- Cardeira, E. (2006). *O essencial sobre a História do Português*. Alfragide: Editorial Caminho.
- Cardeira, E., & Mateus, M. (2008). *Norma e Variação*. Alfragide: Editorial Caminho.
- Cardoso, J. (1569-1570). *DictionarivmLatinohusitanicum& vice versa lusitanicolatinu[m]*. Coimbra: Joan. Barrerius.

- Corpus Lexicográfico do Português. (s.d.). Retrieved from: <http://clp.dlc.ua.pt/inicio.aspx>
- DICIWEB. (s.d.). Obtained from: <http://clp.dlc.ua.pt/DICIweb/default.asp?url=Home>
- Finatto, M. J., Quaresma, P., & Gonçalves, M. (2018). Portuguese corpora of the 18th century: old medicine texts for teaching and research. In D. Fiser, & A. Pancur (Eds.), *Proceedings of the Conference on Language Technologies & Digital Humanities* (pp. 114-120). Ljubljana: University of Ljubljana, Slovenia, retrieved from: <http://hdl.handle.net/10174/23606>.
- Gonçalves, M. F. (2003). *As ideias ortográficas em Portugal - de Madureira Feijó a Gonçalves Viana (1734-1911)*. Lisboa: Fundação Calouste Gulbenkian.
- Gonçalves, M. F., & Banza, A. P. (Eds.). (2013). *Património Textual e Humanidades Digitais: da antiga à nova Filologia*. Évora: Publicações do CIDEHUS, available at: <https://books.openedition.org/cidehus/1073>.
- Mendes de Almeida, J. (1967). Lexicógrafos portugueses da língua latina: A Prosodia de Bento Pereira. *Revista de Guimarães*, LXXVII (1-2), 5-12.
- Pereira, B. (1647). *Thesouro da Lingoa Portuguesa*. Lisboa: Paulo Craesbeeck.
- Pereira, B. (1697). *Prosodia in Vocabularium bilingue, Latinum et Lusitanum digesta*. Évora: Universidade de Évora.
- Teyssier, P. (1997). *História da Língua Portuguesa*. Lisboa: Sá da Costa.
- Verdelho, T. (1987). Latinização na história da Língua Portuguesa - o testemunho dos dicionários. *Volume de homenagem a Paul Teyssier*. Arquivos do Centro Cultural Português, XXIII, pp. 157-187.
- Verdelho, T. (1992). Aspectos da obra lexicográfica de Bento Pereira. *XXe Congrès International de Linguistique et Philologie Romanes*, IV (VI), pp. 777-785.
- Verdelho, T. (2002). Dicionários Portugueses, breve história. In J. H. Nunes & M. Petter (Eds.), *História do saber lexical e constituição de um léxico brasileiro* (pp. 15-64). São Paulo: Humanitas, FFLCH/USP, Pontes.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* (3), 160018. <https://doi.org/10.1038/sdata.2016.18>