

Notes on variation and lexical diachrony in the *Parish Memories-Alentejo* collection (1758)

Helena Freire Cameron¹, Fernanda Olival², Renata Vieira²

¹CIDEHUS- Portalegre Polytechnic University, Portugal, Portugal

²CIDEHUS-University of Évora, Portugal

helenac@ipportalegre.pt, mfo@uevora.pt, renatav@uevora.pt

Abstract. *Memórias Paroquiais-Alentejo (1758)* collects the responses of the parish priests from the largest region of Portugal (Alentejo) to a survey carried out by the Crown, asking about the state of the territory and its populations, and also about the effects of the earthquake 1755. This article discusses the transformative process from the manuscripts up to the processable digital stage. We described some individualising characteristics of the lexical “corpus” obtained with PLN. Marks of lexical diachrony also stand out, revealing society and its uses in the mid-18th century.

Resumo. *As Memórias Paroquiais-Alentejo (1758)* reúnem as respostas dos párocos da maior região de Portugal (Alentejo) a um inquérito feito pela Coroa, sobre o estado do território e das suas populações, e também sobre os efeitos do sismo de 1755. Neste artigo dá-se conta do processo transformativo dos manuscritos até à fase digital processável. Descrevem-se algumas características individualizadoras do corpus lexical realizadas através de PLN. Destacam-se, igualmente, marcas de diacronia lexical, reveladoras da sociedade de então e dos seus usos.

1. *Parish Memories: constitution and support transfers*

In order to obtain a better knowledge of the territory, the Portuguese crown, in 1758, through the Secretariat of State for Business of the Kingdom, carried out a survey of the entire continental territory of Portugal with 60 questions. He asked about the “land”, the “mountain”, the “river”, and the effects of the Lisbon earthquake of 1755. He tried to collect information about the territory, the buildings, the populations and aspects that today we would classify as material culture, such as religious festivities and local memory, based on identity factors, notable aspects and illustrious characters. This initiative had precedents. There was previous information resulting from other surveys, namely the one in 1732 and other subsequent ones (Capela (1998)). However, the earthquake not only destroyed much of this information, such as that collected by the Oratorian Father Luís Cardoso using a geographical dictionary, but also changed the country’s reality (Cosme & Varandas (2009)). Therefore, the crown launched a new survey, almost certainly with the help of the invoked Father. It should reach every parish in the kingdom. Each formed the smallest administrative unit in the territory, the parish, which then had only a religious reference.

The parish priests' responses were later gathered into a documentary collection, aiming to continue the creation of a future geographical dictionary initiated by Father Luís Cardoso (Bigotte Chorão (1987)). The collection was bound in the 19th century and is now in the custody of the National Archives of Torre do Tombo (ANTT) in Lisbon. It would later become known as *Memórias Paroquiais*, a name both “consecrated and incorrect”, according to Bigotte Chorão (2023), since it deals with the parish priests' responses to the inquiry and not the memories of the clergy themselves, as the title may suggest.

Given the high demand for consulting of this valuable collection, between 1993 and 2003, ANTT microfilmed the originals and digitized them in 2005, making them available online, in image format. Therefore, the face-to-face consultations of these species have ended on paper.

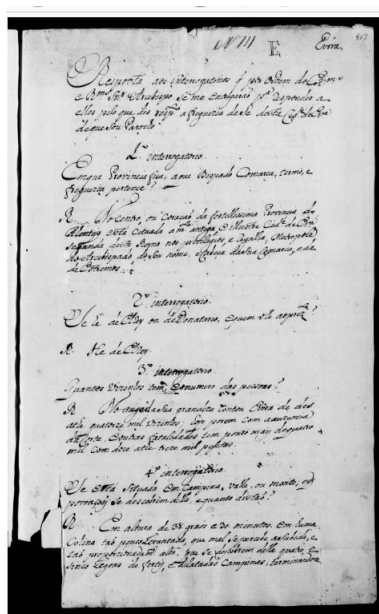


Figure 1: Extract from the *Parish Memory* of Évora, p. 1, ANTT, PT-TT-MPRQ-14-111, at <https://digitalrj.arquivos.pt/viewer?id=4240009>

In 2007, the CIDEHUS research centre (Interdisciplinary Center for History, Cultures and Societies of the University of Évora) began transcribing the *Parish Memoirs* of Alentejo (southern area of Portugal), onto digital media in a fully processable format. Three years later, part of the collection was made available online and later migrated to the data aggregator repository CIDEHUSDigital (www.cidehusdigital.uevora.pt), where the texts can be searched, almost always in 18th-century spelling (with some interventions by transcribers/proofreaders). A search engine is also available, enabling more subtle searches (Santos et al. (2020)).

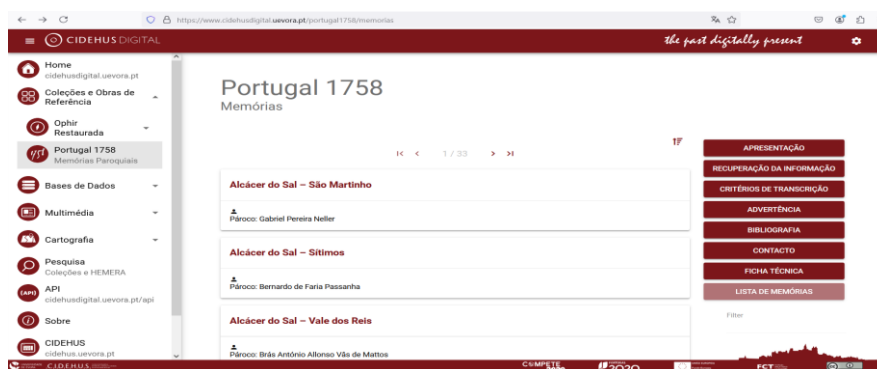


Figure 2: Collection *Portugal 1758*, at [CIDEHUSDigital](https://www.cidehusdigital.uevora.pt/portugal1758/memorias)

2. Brief notes from the lexical *corpus* of *Parish Memories-Alentejo*

In 2020-2022, a multidisciplinary team from CIDEHUS-University of Évora created a lexical corpus with the 366 Memórias do Alentejo, relating to the parishes that belong to the current districts of Portalegre, Évora and Beja. The texts were obtained from the transcriptions available on CIDEHUSDigital. When transcribing, fidelity to the original was preserved whenever possible, maintaining the spelling and punctuation of the handwritten originals. Even so, there were some interventions, namely: some transcribers updated “u” with a consonant value, registering it as “v”, abbreviations were developed, some capital letters that were inserted in the middle of the word were updated, and the long “s”.

This *corpus* consists of texts of various dimensions, written by priests of various ages with various backgrounds and geographic and social origins. These factors make this *corpus* representative of the Portuguese language in this region, in the mid-1700s, and unique in the linguistic panorama of the 18th century in Portugal.

The *corpus* from the 366 Memories mentioned, non-lemmatized and non-standardized, contains 34,181 simple words and 632,498 occurrences, processed using the *AntConc* tool (version 4.0.3). An alphabetical list and a list ordered by descending frequency were prepared. In this article, we indicate the number of occurrences for each word cited in parentheses.

Excluding stop words, some of the most frequently used words are “villa” (4621), “freguezia” (2737), “senhora”(2476), “igreja” (2254), “senhor” (1683), “legoas” (1558), “cidade” (1469), “anno” (1342), “santo” (1208), “rio” (1206), “serra” (1150), “dista” (1141), “altar” (994). These words reveal the purpose of the surveys and reflect the parish priests' vision of the region, with a high incidence of words linked to religious experience.

Concerning linguistic characteristics, graphic variation is a characteristic of this period of the Portuguese language, already abundantly described in the literature (Roberts and Kato (1993), Kemmler (2001), Gonçalves (2003), Castro (2006), Cardeira and Mateus (2008), Martins (1995, 2016, ...), Banza and Gonçalves (2018), Galves, Brito and Paixão de Sousa (2005), Galves, Kato and Roberts (2019), among others equally relevant). The *corpus* of Memórias Paroquiais-Alentejo highlights, in an expressive way, the graphic variation, not only as a characteristic of the 18th-century period, but also due to the plural constitution of the *corpus* itself, Cameron (2023).

Linguistic variation in the *corpus* has several types (Cameron, 2023). Here, we present some examples of this variation, which is not always predictable and often incoherent. At the level of double consonants without linguistic criteria, forms with and without a doubled consonant coexist, with different numbers of occurrences of each of the variants. In the *corpus*, we have “abbade”(19), but there are no occurrences without the double consonant. Nevertheless, the priests wrote “abbobada”(1), “abobada”(45), “abobeda”(220) and, in plural, “abobadas”(5) and “abobedas”(25). In this last example, the variant with the highest number of occurrences does not correspond to the current word, singular or plural.

Often, the same word contains registers of variation of different phenomena, with variants in terms of doubled consonants and the register of the nasal diphthong, for example in “cappelam”(6), “cappellam”(7), “cappellao”(1), “cappelão”(10), “cappellão”(26), “cappellám”(1), “cappellã”(2), “cappelã”(1), and, in plural, “cappellaens”(1), “cappellaenz”(1), “cappellaes”(3), “cappellães”(1), “cappeloens”(1).

There are many variants with a single occurrence, almost half the number of simple words in the *corpus*. These singularly occurring variants often demonstrate lapses or errors on the part of the writers or even some lesser knowledge of the language. We note some of them: “addemiravens”(1) (for admiráveis), “apouzentado”(1) (for aposentado), “Bultiau”(1) (referring the lexicographer Bluteau), “coditiana”(1) (for quotidiana), “domecilio”(1) (for domicílio), “espisial”(1) (for especial), “espriencias”(1) (for experiências), “exziste”(1) (for existe), “hipicondriacos”(1) (for hipocondríacos), “obestujiaõ”(1) (for obstrução), among many others that are equally interesting.

The *corpus* also contains numerous Latin words, whether expressions of religious use, such as “sede vacante”, “authoritate ordinarii”, or even Latin expressions, such as “ostiatim”, “in perpetuum”, among others (See, in this regard, about the *Memories* of Vila Viçosa, Olival (2023)).

We morphologically annotated the *corpus* using the *Lx Tagger tool*, available in [Portulan Clarin](#). Although the *corpus* was not previously normalised, the tool recognised and correctly labelled most of the words. We used the labels that were pre-defined by the tool.

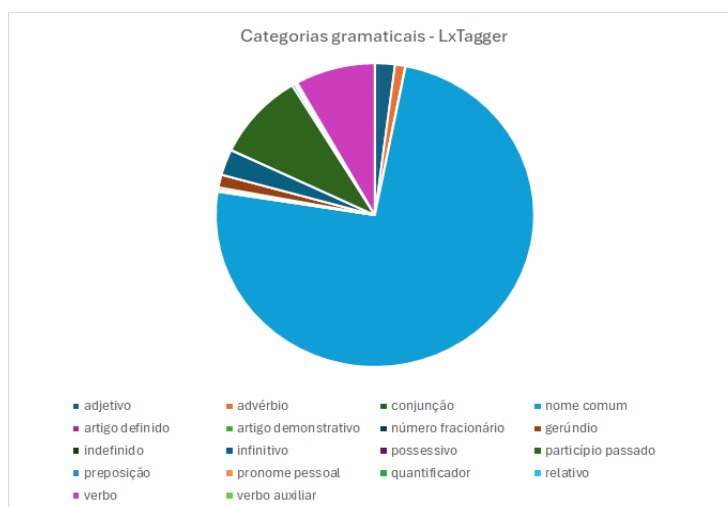


Figure 3: Breakdown of grammatical categories in the non-lemmatized and non-standardized *corpus*

Most of the *corpus* comprises nouns, followed by past participles and, thirdly, verbs. It should be noted that the *corpus* was not previously lemmatised, so the results obtained must be seen from this perspective.

3. Some lexical diachrony marks

The *corpus* of *Memórias Paroquiais-Alentejo*, in addition to being an “x-ray” of the Portuguese language from the 18th century, also constitutes a privileged historical guide to the lexicon relating to Portuguese society, especially in Alentejo, its occupations, uses and habits. Some lexical expressions reveal this past experience and are unknown to current speakers, as they have unusual meanings in the 21st century. Thus, in a second phase of the *corpus* study, the researchers read the texts, highlighting terms that had historically limited meanings or no longer used today. Once the terms were highlighted, they were observed in the *AntConc* tool, using the KWIC (keywords in context) function, allowing the word to be seen in context and giving immediate access to the entire original text.

Likewise, lexicographic existences of some of the terms were searched in previous or contemporary dictionaries: in the dictionaries of Jerónimo Cardoso (16th century), Bento Pereira (7th edition, from the end of the 17th century), Rafael Bluteau (from the first quarter of the 18th century). We also consulted António de Morais Silva's dictionary, in its 10th edition, from the mid-20th century, in 12 volumes. In 21st century dictionaries, we used the Houaiss dictionary, in its first Brazilian edition, 2001.

The *corpus* is rich in very interesting examples. We chose three of them, which clarify this diachrony, revealing characteristics of society in this region at that time:

3.1. “Herdade de cavalaria” [TRAD: Cavalry homestead]

The expression “cavalry estate” is curious. The meaning used has nothing to do with the etymological root, horse, but refers to an economic occupation of the territory beginning to appear in this region at that time. In the entire *corpus*, it appears in only three *Memories*, which we reproduce:

“[...] mas como nesta freguesia andão muitas **herdades de cavalaria**, ha muytos cazeiros, e como cada cazeiro, /p. 1183/ está em costume de pagar ao Prior hum alqueire de trigo, sempre o Prior fas certos quatro mojos de trigo”.

Monforte-Nossa Senhora da Graça, ANTT, *Memórias Paroquiais*, vol. 24, nº 179, p.1182-1183

The second Memory clarifies in detail the meaning of the term:

“[...] 27º Constando esta freguezia de cincoenta herdades, somente tem dezouto lavradores; e as outras trinta e duas herdades andão **de cavalaria**, e destas se achão treze herdades com as cazas, ou moradias derribadas, a fim de que algum lavrador não pretenda orar nellas; algumas forão destruidas no tempo da guerra, e outras tem sido derribadas pelos mesmos, que as cultivão **de cavalaria**, em grave damno, e prejuizo da Republica, e do bem commum; porquanto os lavradores, que trazem as **herdades da cavalaria**, não as cultivão, nem lavrão, como erão cultivadas, e lavradas naquelle tempo, em que em cada huma herdade morava seo lavrador, e lavrava as suas folhas enteyras; o que agora não fazem, nem cultivão, ou lavrão, a fim de terem pastajes largas para criar ovelhas, e venderem muntas láas aos estrangeiros, que as levão para o Norte.

Vila Viçosa - Nossa Senhora das Ciladas, ANTT, *Memórias Paroquiais*, vol. 11, nº 326, p. 2234.

In the third Memory, the parish priest warns of the dangers of this type of economic occupation:

“[...] e podendo recolher muytos mais mantimentos os não recolhem por se utelizarem das terras, para pastajem dos gados, por cujo interesse tem deyxado cahir a mayor parte dos montes, que há nesta freguezia, que tendo legoa, e meya de comprimento, e outro tanto de largura; e trinta e sinco herdades, se achão só /p. 434/ habitando nella, sette lavradores, e as mais andão **de cavalaria** por cujo motivo falta o culto Devino, mantimentos no Reyno, sustento aos Parochos;”

Elvas - Senhora da Ajuda, ANTT, *Memórias paroquiais*, vol. 1, nº 62, pp. 433-434

The lexicographic register of this expression is not systematic. Bluteau, contemporary of the parish priests who wrote the *Memories*, already notes this meaning:

“cavallaria [...] Andar de Cavallaria. No Alentejo quando arrendão as herdades, huma das clausulas he, que não andem de cavallaria, querem dizer, que sejaõ habitadas pelos arrendadores, porque sem habitadores se perdem, e cahem as casas.”

Bluteau, R. (1712-1728), *Vocabulario*, vol. 9.

In the 20th century, Morais (1949-1959) registers the expression in more detail and it is closer to the meaning used in the *corpus* of the Memories.

“[...] *herdade de cavalaria*, herdade que não tem cultura ou se cultivava só uma pequena parte || Pôr, trazer a herdade de cavalaria, cultivar a menor parte e deixar o resto a monte [...]”

Morais Silva, A. (1949-1959) *Grande Dicionário da Língua Portuguesa*, vol II, p. 1043.

Houaiss (2001) marks the expression as being a Lusitanianism, only used in Portugal:

“cavalaria [...] P. Herdade que não tem cultura ou que cultivava apenas uma parte de suas terras.

Houaiss (2001), *Dicionário Houaiss da Língua Portuguesa*, p. 659.

The meaning was not found in other contemporary dictionaries.

3.2. “Polícia” [TRAD: Police]

The term “polícia” has two occurrences in the *corpus* under consideration, both in meanings very distant from contemporary common usage. The two occurrences are included in the same Memory, which means the same person wrote both forms. However, there seem to be some differences between them:

“[...] Fora das grades do Cruzeiro sta a capella de Sancto Antonio, que com o quadro do mesmo Sancto mandou fazer o Excelentíssimo Dom Balthasar de Faria de marmores finissimos de varias cores primorisamente lavrados, onde se acha o nome de seu Excelentíssimo author gravado em huma pedra em letras inniciaes douradas enlaçadas com muita arte, e **policia**;”

Elvas-Sé, ANTT, *Memórias Paroquiais*, vol. 13, nº (E) 14, p. 76

“[...] Correndo os annos se lhe ajuntarão companheiros com mais **policia** e noticia de letras, que se ordenarão em presbíteros.”

Elvas-Sé, ANTT, *Memórias Paroquiais*, vol. 13, nº (E) 14, p. 71 a 106, p. 98

Morais (1949-1959) includes two meanings that seem to coincide with the uses given by the parish priest of the Cathedral of Elvas:

[...] apuro, esmero, cuidado ou correção da linguagem [...] || Delicadeza, aprumo, correção, polidez, civilidade [...] || Conjunto de preceitos técnicos destinados a um dado fim [...]

Morais Silva, A. (1949-1959), *Grande Dicionário da Língua Portuguesa*, vol. VIII, p. 444

The first occurrence seems equivalent to a set of technical precepts while the second may be equivalent to “care”.

We did not find, in contemporary dictionaries, meanings coinciding with the 18th century use found in the *Parish Memories*. However, this use of the term “policia” has a lexicographic record prior to the 18th century, in main bilingual dictionaries: Jerónimo Cardoso, Bento Pereira, and Bluteau:

“policia. policies(ei). politura(ae).”

Cardoso, Jerónimo (1562), *Dictionarium ex lusitanico in latinum*.

“Policia. Cicilitas, atis,. Urbanitas, atis. Elegantia, ae.”

Pereira, Bento (1647) *Thesouro da Língua Portuguesa*.

“[...] Policia no trato, na conversaçãõ, nos costumes, etc. Urbanitas, ou comitas, atis. Fem. Cic. Morum elegantia, ae. Fem. Tacit. {(A honra he a fonte de todo o bom ensino, Policia, procedimento, etc. Lobo, Corte na Aldea, Dialog. 15. pag. 306.)} {(A brandura, & Policia de vossas palavras. Idem. Primavera, 3. part. 194.)} § Policia, tambem se toma pela boa graça nas ações, & gestos do corpo, etc. Vid. Garbo, Graça, etc. {(Vereis alguns que fallaõ às pancadas, & se acharem hü pulpito diante, o farãõ em pedaços, como se a Policia podera soffrer o desassossego, & inquietaçãõ da sua esgrima. Lobo, Corte na Aldea, Dialog. 8. pag. 172.)} § Policia, algumas vezes val o mesmo, que Aceyo, Limpeza, Alinho, etc. Vid. nos seus lugares. {(A brandura no conversar, a Policia no vestir, a cortezania no tratar, Lobo, Dialog. 5. mihi pag. 114.)} {(Lhe servissem com grande Policia ministros escolhidos as melhores iguarias. [...])”

Bluteau, R. (1712-1728) *Vocabulario*, vol.6.

3.3. “País” [TRAD: Country]

The word country is used in the *Memories* in two different meanings, designating a territory with a political organization, a use that is predominant today:

“[...] indo sahindo da igreja o tomarão os mouros, e captivo o levãrão para Argel, sem mais pessoa alguma deste **país**.”

Vila Nova de Milfontes - Nossa Senhora da Graça, ANTT, *Memórias Paroquiais*, vol. 23, nº 142, p. 908

and referring to a specific region, as is common today in the French language:

“[...] He o seu curso neste **País** moderado, sem mi[ni]ma precipitação, por serem planicie, os campos por onde passa.”

Elvas-Terena, ANTT, *Memórias Paroquiais*, vol. 36, nº 44, p. 277

“[...] laranja doce deste **pays** he particular no gosto em todo Alentejo [...]”

Montemor-o-novo-Santiago do Escoural, ANTT, *Memórias Paroquiais*, vol. 14, nº 55, p. 402

Houaiss (2001) includes this meaning, referring it to the French term “pays”, which has been used since the Renaissance. This meaning can be recognized in derived words such as “paisagem” (landscape), “paisagismo” (landscaping), “paisano”, among others.

Bluteau, in the middle of a definition for another purpose, uses the term in this sense, referring to “region”:

“Aquitania. Grande provincia de França, que os modernos divide em tres, a saber a primeyra Aquitania, que contem o Berry, o pays de Bourbon, a alta, & a baxa Alvernia,

o pays de Velay, de Gevedan, de Roverga, dos Albigenses, de Querey, de Limoges, & da alta, & baxa Macha. A 2. Aquitania contem o pays de Bordeos [...]
Bluteau, R. (1712-1728), *Vocabulario*, vol.1.

This meaning had already been included in the 7th edition of *Prosodia* (1697), by Bento Pereira:

“Adrana, ae, m. g. Pais de Alemanha. p. l. ex Graec. Tacit. lib. 1. ann. [...]”
Pereira, Bento, (1697) *Prosodia*, 7th ed.

4. Conclusions

The lexical *corpus* of Memórias Paroquiais-Alentejo brings together an expressive graphic variation, characteristic of the time, and also the result of the multiplicity of authors who wrote the almost four hundred texts that constitute it. It contains lexical diachrony marks, as we saw briefly. It is a true source from a linguistic point of view. Access to the lexical mass, with the support of concordance generation tools, allows each word to be seen in context, enabling the construction of lexical datasets, which can feed ontologies and, at the same time, constitute observatories of societal changes (e.g. “cavalry estate”).

Given the size of the *corpus*, this first phase of defragmenting the text into simple words, useful for studying variation, was followed by a second phase of morphological annotation of the *corpus*. Almost two-thirds of the *corpus* is filled with names. However, it is estimated that this value may decrease after lemmatization.

The third task was carried out manually, observing the texts, in order to see the words in context. The *corpus* is being annotated with Named Entities customised to the historical context (Vieira, 2021), (Santos, 2024). However, the study of these low-frequency terms cannot be carried out automatically, always requiring the human intervention of linguists and historians. Therefore, the authors did a new systematic reading of the texts, to withdraw uses and meanings with diachronic value.

The automation of historical *corpora* will greatly benefit from all the efforts to process and study the Portuguese lexicon. However, a subsequent phase of human intervention will always be necessary, to fully enhance the study and dissemination of these valuable historical lexical *corpora*. The power of the Artificial Intelligence arm is still limited.

5. References

- Banza, A. P. et Gonçalves, M. F. (2018) *Roteiro de História da Língua Portuguesa*. Évora: Cátedra UNESCO - Universidade de Évora.
- Bigotte Chorão, M. J. M. (2023) «As Memórias Paroquiais de 1758, de novo», *Notícias da Torre do Tombo*. Lisboa, N.º 1 (jul. 2023).
- Bigotte Chorão, M. J. M. (1987) “Inquéritos promovidos pela coroa no século XVIII”, *Revista de História Económica e Social*, nº 21, Lisboa, pp. 93-130.
- Bluteau, R. (1712-1728) *Vocabulario Portuguez e Latino*, Coimbra, Lisboa, Lisboa Colégio das Artes, Pascoal da Sylva, Joseph Antonio da Sylva, Patriarcal Officina da Musica.

Cameron, H.F, Olival, F., Vieira, R. (2023) Planear a normalização automática: tipologia de variação gráfica do corpus das Memórias Paroquiais (1758), *LaborHistorico*, v.9 n.1, <https://doi.org/10.24206/lh.v9i1.52234>

Capela, J. V. (1998) *Barcelos nas Memórias Paroquiais de 1758*, Câmara Municipal de Barcelos.

Cardeira, E. et Mateus, M. H. M. (2008) *Norma e Variação*. Alfragide: Editorial Caminho.

Cardoso, J. (1569-1570) *Dictionarium latinolusitanicum & vice versa lusitanicolatinum*, Coimbra, João de Barreira.

Castro, I. (2006). *Introdução à História do Português* (2ª edição revista e muito ampliada). Lisboa, Colibri.

Cosme, J. et Varandas, J. (2009) *Memórias Paroquiais (1758), volume I [Abação-Alcaria]*, Casal de Cambra, Caleidoscópio.

Galves, C.; Kato, M.; Roberts, I. (2019). *Português brasileiro. Uma segunda viagem diacrônica*. Campinas. Editora da Unicamp.

Galves, C. M. C.; Britto, H. S.; Paixão de Sousa, M. C. (2005), «The Change in clitic placement from Classical to Modern European Portuguese: Results from the Tycho Brahe Corpus». *Journal of Portuguese Linguistics*, v. 4, p. 39-67.

Gonçalves, M. F. (2003) *As ideias ortográficas em Portugal: de Madureira Feijó a Gonçalves Viana (1734-1911)*, Lisboa, Fundação Calouste Gulbenkian.

Houaiss, A. (2001) *Grande Dicionário Houaiss da Língua Portuguesa*, Rio de Janeiro, Editora Objetiva.

Kemmler, R. (2001), «Para uma História da Ortografia Portuguesa: o texto metaortográfico e a sua periodização do século XVI até à reforma ortográfica de 1911». *Lusorama* 47-48, pp. 128-319.

Olival, F., Cameron, H.F, Farrica, F., Vieira, R. (2023) As Memórias Paroquiais do atual concelho de Vila Viçosa, Callipole, n.29, pp. 85-128.

Pereira, B. (1697) *Prosodia in Dictionarium bilingue*, Évora, Évora, Tipografia da Academia, 7th edition.

Reynaert, M. W. C., Hendrickx, I. H. E., & Marquilhas, R. (2012). «Historical spelling normalization. A comparison of two statistical methods: TICCL and VARD2». In F. Mambrini, M. Passarotti, & C. Sporleder (Eds.), *Proceedings of the Second Workshop on Annotation of Corpora for Research in the Humanities (ACRH-2)*, Lisbon, pp. 87-98. Edicoes Colibri.

Roberts, I. et Kato, M. A. (1993). *Português brasileiro. Uma viagem diacrônica*. Editora Contexto.

Santos, I., Olival, F., Sequeira, O., (2020), «Excavating the data pit: the Portuguese Parish Memories (1758) as a gold standard», in *DHandNLP 2020: Digital Humanities and Natural Language Processing: Proceedings of the Workshop on Digital Humanities and Natural Language Processing (DHandNLP 2020)* co-located with International Conference on the Computational Processing of Portuguese (PROPOR 2020). ed by M. José Finatto; Renata Vieira; Senja Pollak; Saturnino Luz, Évora, Vol. 2607.

Santos, J., Cameron, H.F., Olival, F., Farrica, F., Vieira, R. (2024) Named entity recognition specialised for portuguese 18th-century history research. In: *Proceedings of the 16th International Conference on Computational Processing of Portuguese*. pp. 117–126

Vieira, R., Olival, F., Cameron, H., Santos, J., Sequeira, O., Santos, I. (2021) Enriching the 1758 portuguese Parish Memories (Alentejo) with Named Entities. *Journal of Open Humanities Data* 7, 20