

Bridging edition and corpus: a review of P. S. Post Scriptum: A Digital Archive of Ordinary Writing (Early Modern Portugal and Spain)

P. S. Post Scriptum, CLUL (ed.), 2014. <http://ps.clul.ul.pt> (Last Accessed: 22.01.2019). Reviewed by Ulrike Henny-Krahmer (Universität Würzburg), [ulrike.henny \(at\) uni-wuerzburg.de](mailto:ulrike.henny@uni-wuerzburg.de).



Abstract

The project *P. S. Post Scriptum* (2012–2017) was aimed at the collection, edition, and analysis of Spanish and Portuguese private letters from the 16th to the early 19th century. As a result, a digital text collection and edition of the letters is presented on the project website. *P. S. Post Scriptum* can be classified both as a scholarly edition and a diachronic linguistic corpus and can be considered pioneer work in combining procedures from textual criticism together with the linguistic preparation of the corpus, serving the needs of both historians and linguists. Valuable basic work has been done in encoding approximately 5,000 letters which can be downloaded in different formats on the website. Only an open licence for reuse of the data remains desirable and at times the project could have benefitted from an even more advantageous presentation of the results on the website.

1 In the *P. S. Post Scriptum* project, private letters written in Spain and Portugal between the 16th and the early 19th century were collected, edited, and analyzed. Almost all of the letters were previously unpublished, so a digital collection and edition of

them was one of the main goals of the project. Another priority was the historical and linguistic study of the letters in this diachronic corpus. The *P. S. Post Scriptum* project was funded with an ERC Grant between 2012 and 2017 and was hosted at the faculty of arts and humanities of the University of Lisbon.¹ This review is about the digital text collection and edition of the letters resulting from *P. S. Post Scriptum* and published at <http://ps.clul.ul.pt/>.²

Background

en vicio en el espejo
de la amistad en el espejo

yo estoy tan maravillada de la poca quenta que v. md haze de my no se
por que lo haze tan mal eno me ader espito dos años de q si vna carta
en q me yn dia ba adozir q se den die sen las casar yo senoz
tes por de lue go q yo haria lo q me yn dia ba amandor no desto
senoz otra carta qora alio que senoz me ayedis de la mucha
peña que dezis que teneyis de la ad fozia mya no se parece en
ader me besitado con vras cartas siquira para fa ver de la salud
pidos senoz por amor de no senoz que no lo agays an si siquira
por quen bos senoz soys y el aycho de las jentes q no les parezian
senoz yo an daba en ben der las casar como me ayedis tes y vino a
my ayris bar vero q el tenya carta quem se ven dusea y por esto no las
exendiado hasta que se bea dia ca y an si estare hachm dando paralo hazer
lo que por ella viere pidos senoz por md que con el pre mio men safero
me yn die adisar de todo por que yo no hare otra cosa mas de lo q mandare
senoz vna puda ader se llegado asta que adnpue blo quatro leguas
de aqui y all platicaramos lo que se adia de hazer y dar hoz den
en la vida qasi lo ager otras personas que andan por alla de esa
manera se me dize as besa las manos y dize q tan dion anda
como v. m. q su negocio mia se acaba y q se mira villa mucho
de que no nos aybe que se el supuse q no dize ba que el ay
vria no desers senoz de le ayedir por q en lo dard que tiene mucho
de seo de saber de sus alnd y verle aqui asta do estos dias en su casa
que no salia de my senoz hos to fader q estoy muy flaca y on esto triaba
fo en fa ber de vna salud de los dineros de el pnvsa no los cobrado
no ten go mas que le supli con se no que lo haga an si como asta aqui
de ba lly apostero de mayo ano 1553 y maria de
de castilla vna ayris alio que se mandare
pu mter q mas q asile quere

Fig. 1: Letter written by María de Espinosa to her husband in 1553. <http://ps.clul.ul.pt/index.php?action=file&cid=PS7004>. Accessed June 1, 2018.

2 The letters edited in *P. S. Post Scriptum* are private correspondence written by people of different ages, gender, occupation, and from different social backgrounds. Most of them survived because they were used as evidence in civil, military, ecclesiastical, and inquisitorial courts and preserved in the corresponding archives. The letters served

as proofs to convict or absolve their senders, recipients, or persons mentioned in them, and were often accompanied by transcriptions of hearings of witnesses and defendants. Both kinds of materials are of special documental value for the study of everyday history and the history of language, as many everyday issues are covered in the letters and transcripts and presented in a register³ close to spoken language.⁴ Fig. 1 shows an example of a letter sent from Valladolid to Toledo in 1553, written by a wife to her husband complaining that she hasn't heard from him over the last years. The letter is kept in the National Historical Archive in Spain and was used in a legal procedure where the husband was accused of bigamy.⁵

3 The collection and edition of private letters of this kind has been at the heart of two predecessor projects closely related to *P. S. Post Scriptum*. The project *CARDS Unknown Letters* started in 2007 and was aimed at the online edition and study of Portuguese private letters written between 1500 and 1900.⁶ The second related project started in 2010 and is called *FLY Forgotten Letters, Years 1900-1974*, or *FLY Cartas Esquecidas* in Portuguese. It has an online presence at <http://fly.clul.ul.pt/>.⁷ The letters of FLY were written in the 20th century, in the context of war, migration, imprisonment, and exile. Both the predecessor projects and *P. S. Post Scriptum* were carried out at the Linguistics Department of the University of Lisbon, the latter starting in 2012. While dealing with letters from different periods of time and contexts, and transmitted in different ways, the general goals of the projects CARDS and FLY were the same and also the methodology was shared. The approaches developed in the precursor projects were in principle continued in *P. S. Post Scriptum*, but with a different linguistic and regional scope. In *P. S. Post Scriptum*, the corpus was extended to also include letters from Spain from the Early Modern period. In addition, it was announced that further types of linguistic annotation were to be added.⁸ In this review, only the results of the *P. S. Post Scriptum* project are discussed.⁹

4 The goals of *P. S. Post Scriptum* are described on the homepage of the project website as follows:

Within the P. S. (Post Scriptum) Project, systematic research will be developed, along with the publishing and historical-linguistic study of private letters written in Portugal and Spain along the Early Modern Ages. These documents are almost all unpublished epistolary writings made by authors from different social backgrounds. [...] These textual resources often present an (almost) oral rhetoric, treating everyday issues of past

centuries in a register that hasn't been easy to study, apart from brief examples. Not only does the P. S. Project present a wide collection of private letters, but it also makes it available as a scholarly digital edition and as an annotated corpus.

("Project P. S. Post Scriptum." CLUL 2014. [https://web.archive.org/web/20180602124555/http://ps.clul.ul.pt/.](https://web.archive.org/web/20180602124555/http://ps.clul.ul.pt/))

5 What is highlighted here is the relevance of collecting and publishing historical sources that were previously inaccessible to a wider audience, but also their value for the diachronic study of language registers that are rarely documented otherwise. Especially noteworthy is the double goal set for this digital text collection: it is conceived as a scholarly digital edition, but also as a linguistic corpus. This broad scope is reflected in the conception of *P. S. Post Scriptum* as an inherently interdisciplinary project with a team of linguists and historians from Spain and Portugal. The "Credits" page, which lists members of *P. S. Post Scriptum* as well as of the CARDS project¹⁰, shows the range of fields involved: historical and computational linguists, a teaching researcher, members responsible for linguistic, historiographic, and paleographic survey, experts in digital scholarly editing, conceptionists, and programmers.¹¹

6 Because of its double goal, *P. S. Post Scriptum* should be reviewed as both a scholarly digital edition and a linguistic corpus. This project is one of many that don't easily fit into a taxonomy of scholarly digital resources.¹² It is very laudable that the letters in *P. S. Post Scriptum* are edited critically in order to lay the foundation for research in various disciplines. The project members themselves reflect their multidisciplinary approach:

The advantages that the new technologies offer to the humanities cannot be denied, even if there is a lack of creation of digital tools that are useful for different scholarly disciplines. In general, the editions designed for the philologist and the historian are not exploited by the linguist in the same manner that an annotated corpus would be; and vice versa, the linguistic corpora, conceived first and foremost to obtain statistics and word concordances, are resources of little utility for historical investigation and textual criticism.

In this work we present an already finished research project which incorporates the

methodologies of the digital humanities and corpus linguistics to offer an integrative treatment of sources that can be of interest in several fields of study.

(Vaamonde 2018: 140. Translated from Spanish by the reviewer.)

7 Vaamonde sees the new technologies as a type of catalyst fostering the development of tools that are useful to various humanities disciplines and at the same time making a lack of such tools visible. A similar spirit of disciplinary reunification emerging from the digital humanities is formulated by Marquilhas and Hendrickx:

Admittedly, the division of the philological method in the past primarily served as an excuse to define the differences between the modalities of edition which the three areas [linguistics, literary history, and textual criticism] respectively demanded. Now the adoption of digital methods and the creation of resources led to a situation where the “walls” between these areas began to fall, in an exemplary dynamic of interdisciplinarity.

(Marquilhas and Hendrickx 2016. Translated from Portuguese by the reviewer. Marquilhas and Hendrickx base their observation on McCarty 2005.)

8 Taking these statements of the project members into account, *P. S. Post Scriptum* aims to provide fundamental research for the digital humanities as a whole and not just to serve the needs of a single discipline. In this review, *P. S. Post Scriptum* is thus examined from various perspectives. First, the general structure and contents of the project website are discussed. Following this, the general parameters of *P. S. Post Scriptum* as a digital text collection are reviewed: the principles of selection of documents and texts, the organization and accessibility of the materials, available metadata and documentation. *P. S. Post Scriptum* is then assessed as a scholarly digital edition on the one hand and as a linguistic corpus on the other. Finally, the long-term prospects of the entire resource are evaluated.

Website

9 The website of *P. S. Post Scriptum* is available in three languages: Portuguese, English – the default, and Spanish. What is translated are the main menu, headings and introductory texts on single pages, and search options, except the XPath options and predefined queries for a syntactic search which are only presented in English. Even the metadata for the letters are partially translated. The metadata categories are available in

the three languages. For the metadata values, the language of the letter seems to determine their main language.¹⁵ The letter written from María de Espinosa to her husband, for example, is in Spanish, so the summary, context information, etc. are also given in Spanish. That the metadata is in different languages may complicate a machine-based reuse of the data. Still, the summary and context information of each letter can be switched to English.¹⁶ In some cases, even the text of the letters themselves is translated. Where this is not the case, a hint is given on how to proceed for automatic translation elsewhere.¹⁷ There seems to be no way to search for translated letters, so they can only be found by chance and it is also not possible to find out how many of them have been translated.¹⁸ The translations of the letter metadata and text are not directly linked to the main language option of the website, so it is for example possible to have the website in Portuguese and read a letter in English. It is comfortable that the main language can be switched on every subpage of the site and not only from the homepage. All in all, much effort has been put into translations to make the site accessible to readers from Portugal and Spain as well as to an international audience.

Post Scriptum
PT | EN | ES
Main Menu

- ▶ Home
- ▶ Search
- ▶ Participants
- ▶ Map
- ▶ Tree Search
- ▶ Papers
- ▶ Credits
- ▶ Downloads
- ▶ Related Projects
- ▶ Login

Powered by TEITOK
© Maarten Janssen, 2014-

Project P.S. Post Scriptum

P.S. Post Scriptum
A Digital Archive of Ordinary Writing
(Early Modern Portugal and Spain)

European Research Council (7FP/ERC Advanced Grant - GA 295562)

Within the P.S. (Post Scriptum) Project, systematic research will be developed, along with the publishing and historical-linguistic study of private letters written in Portugal and Spain along the Early Modern Ages. These documents are almost all unpublished epistolary writings made by authors from different social backgrounds. They could be either masters or servants, adults or children, men or women, thieves, soldiers, artisans, priests, political activists, among other kinds of social agents. Their epistolarity survived by chance, in most cases, when their paths met the persecution means used by the Inquisition, the ecclesiastical, the military and the civil courts, those institutions that used private correspondence as criminal evidence. In other, much less frequent cases, the letters were preserved in a non-criminal context, although they also belong to the realm of backstage interaction and are easy to locate in situational terms. These textual resources often present an (almost) oral rhetoric, treating everyday issues of past centuries in a register that hasn't been easy to study, apart from brief examples. Not only does the P.S. Project present a wide collection of private letters, but it also makes it available as a scholarly digital edition and as an annotated corpus.

Transcription

The abbreviations are expanded, though signalled, and the word boundaries are standardized (except for the enclitic forms), as is standardized the distribution of «i», «j», «u» y «v». A color code highlights the abbreviation expansions, the conjectures, the difficult readings, and the manuscript erasures and additions.

Editorial Procedures

Standardization for the purpose of corpus annotation: the tokens spelling is standardized and

Fig. 2: Homepage of the *P. S. Post Scriptum* project website. <http://ps.clul.ul.pt/>. Accessed June 1, 2018.

10 The structure of the website is straightforward. On the homepage, an introductory text about the project, its goals and the methodologies used, is given (see Fig. 2). Links

to documentation are also placed on the homepage and there is a suggestion for how to cite the whole resource, but there are no citation suggestions for individual subpages. The main menu of the website is placed on the left, is visible on all subpages, and has a single level of entries, making it easy to stay oriented on the website. The first half of the menu entries comprises content-related aspects: “Search”, “Participants” and “Map” are all general access points to the letters, followed by the entry “Tree Search” which constitutes an advanced option. The second half of the menu is devoted to project-related information such as “Credits” and “Related Projects”, and to project outcomes other than the presentation of the letters on the website (“Papers” and “Downloads”). The website seems to have a backend¹⁹, as there is a link to a login. As this is not explained further, it is not clear for whom this option has been designed and how it can be used. Regarding the structure of the website, there are a few subpages that cannot be reached directly from the main menu. These are the letters themselves (discussed further below), a “Word Distribution” page²⁰ which is linked from the introductory text on the homepage, and a “Raw corpus search” page²¹ which only seems to be linked to from the help section of the general search page.

11 On the page “Related Projects”, a list of links is given which is divided into four categories: “Tools”, “Digital Editions and Historical Corpora”, “Guidelines and tutorials”, and “Networks and Associations”.²² While this is an interesting list, some explanation would have been helpful to know in what way the listed “projects” are related to *P. S. Post Scriptum*. The mentioned tools (TEITOK, eDICTOR, CorpusSearch) have probably been used in the project. It is likely that the project is a member of the associations listed (AHDig and REDaiep). But why is the *Digital Archive of Letters written by Flemish authors and composers* listed under “Guidelines and tutorials” and in what way is the tutorial “Text Processing with Linux” relevant for users of the *P. S. Post Scriptum* website? Also for the “Digital Editions and Historical Corpora”, the relationship between *P. S. Post Scriptum* and these other projects is not clear. The predecessor project FLY is listed, but also for example the Mark Twain Project which has probably inspired the methodology of *P. S. Post Scriptum* in terms of a digital scholarly edition of letters. Another project mentioned is *“En el ojo del huracán”: Cartas de Ultramar a España, 1823 - Edición digital*. The homepage of that project explains that a cooperation between *P. S. Post Scriptum* and TEITOK has been established and that the 65 letters of the edition have a second user interface at the Linguistics Centre of the University of Lisbon.²³ As can be seen, there are various reasons for projects to appear on the “Related Projects” page and a few subheadings or sentences would have helped to clarify the relationships.

12 The “Papers” subpage lists publications by type (Books, Guidelines, Book Sections, Journal Articles, Conferences) and by year (2012-2017).²⁴ The number of publications produced by the *P. S. Post Scriptum* team during the project’s funding period is impressive. There are two edited and one authored book, four detailed manuals, 25 book sections, 22 journal articles, and 64 conference papers contributing to various fields such as general linguistics, corpus linguistics, language history, discourse analysis, cultural history, gender studies, and DH.

P. S. Post Scriptum as a digital text collection

Selection of documents and texts

Constituição da amostra

Recolheram-se 3.700 cartas provenientes de 16 arquivos públicos e de 130 coleções privadas e selecionaram-se 2.000 cartas. No caso das correspondências volumosas, não se ultrapassou, para cada par "autor-destinatário", o limite de 20 cartas escolhidas aleatoriamente. Assim se garantiu um tipo de representatividade expressa nestes números: - a dimensão do corpus é de c. 700 mil palavras, escritas por 572 autores de cartas; a média é de 3,5 cartas por autor e de 6 cartas por destinatário. O desequilíbrio entre 3,5 e 6 tem a ver sobretudo com o facto de o corpus integrar, ao lado de 1.860 cartas em registo informal – dirigidas a familiares, amigos, amantes, conhecidos – outras 140 cartas escritas por privados, sobretudo emigrados, e dirigidas ao Secretariado Nacional de Informação (SNI). Entre autores e destinatários, houve 870 indivíduos envolvidos nesta correspondência: – 572 autores e autoras, 338 destinatários e destinatárias. Eram sobretudo os homens quem ia preso, combatia, ia exilado, emigrava... e aprendia a escrever. A coleção Fly tem, portanto, 78% de cartas de autores masculinos, contra 22% de cartas de mulheres. Mas quase metade das cartas (i.e. 47%) foi recebida por mulheres.

Nas Tabelas 1, 2, 3 e 4 pode observar-se a distribuição da amostra em função da sua cronologia, da geografia e dos contextos de escrita. As décadas mais representadas são as de 1960 e 1970 e a região mais representada é África, facto que decorre de o contexto mais bem representado ser também o da Guerra Colonial.

Tabela 1: Cronologia das cartas

Intervalos de tempo	Número de cartas	Porcentagem
1901-1910	62	3,1%
1911-1920	306	15,3%
1921-1930	33	1,6%
1931-1940	50	2,5%
1941-1950	14	0,7%
1951-1960	101	5,0%
1961-1970	921	46,05%
1971-1974	511	25,55%

Tabela 2: Distribuição geográfica das cartas de Portugal

Origem da carta	Total	Porcentagem
Norte	375	18,75%
Centro	191	9,55%
Sul	266	13,3%
Ilhas	63	3,15%
Portugal: s/l	10	0,5%
Portugal: total	905	45,25%

Tabela 3: Distribuição geográfica das cartas de Fora de Portugal

Origem da carta	Total	Porcentagem
África	417	20,85%
Ásia	14	0,7%
América do Sul	271	13,55%
América do Norte	40	2%
Europa	271	13,55%
Fora de Portugal: Total	1013	50,65%
S.l.	82	4,1%

Tabela 4: Contexto de escrita das cartas

Contexto	Número de cartas	Porcentagem
I Guerra Mundial	147	7,35%
Emigração	609	30,45%
Prisão	305	15,25%
Exílio	71	3,55%
Guerra Colonial	868	43,4%
Total	2.000	100%

Fig. 3: Homepage of the FLY project website with corpus statistics. <http://fly.clul.ul.pt/>.

Accessed June 1, 2018.

13 *P. S. Post Scriptum* can be classified as a primary source collection because it includes descriptions and digital facsimiles of the documents, and as a diachronic corpus because it brings together letters from Spain and Portugal from the 16th up to the

19th century. Unfortunately, there is no information on the website about the overall number of letters included in the archive and of the distribution of the letters over time. On the word distribution page, some statistics are given in numbers of words (by language, gender, and social status). The number of words is almost balanced for letters in Spanish and Portuguese (984,830 and 974,391 words, respectively).²⁵ The linguistic focus of the project is illustrated by the fact that the size is measured in “words” instead of “documents” or “texts”. A way to get to the number of letters is to use the search functions: there are 2,369 letters in Portuguese and 2,446 in Spanish which makes a total of 4,815 letters. A search by year also reveals that the earliest letter in the collection dates from 1500 and the latest letters from the 1830s. It would have been easy to provide some general numbers on the homepage or another page of the website to provide the user with a first impression about the size and composition of the corpus. Despite the word counts on the word distribution page, no overview of the temporal distribution of letters is given which would be an important information for a diachronic corpus. Interestingly, the homepage of the predecessor project FLY includes exactly this: a detailed overview of the sample in the form of a descriptive text and tables (see [Fig. 3](#)). A similar summary is desirable for *P. S. Post Scriptum*.

14 On the website itself, no information is given about the selection criteria for the letters. More information can be found in some of the publications presenting the project, but this is not ideal, for several reasons. First, a user of the website is forced to go through the list of publications and look for those which are concerned with the presentation of the project as a whole. In the case of *P. S. Post Scriptum*, the project descriptions tend to focus either on the Portuguese or on the Spanish subcorpus, depending on the context of the publication, which makes it harder to get information about the whole collection.²⁶ Then, these publications were written at different points in time so it is hard to know which ones are in line with the current state of the text collection as it is presented online. Also, not all of the publications are Open Access and directly available.²⁷ In the opinion of this reviewer the publications are an important form of scholarly output, but they cannot replace background and context information published directly with the digital text collection. If an external publication is used to document central aspects of the digital resource and can be considered up-to-date, this should be indicated on the website. This is done, for example, for the guidelines describing the data model, the editorial and the annotation procedures, which are all linked to from the homepage, but unfortunately not for the criteria of selection and composition of the corpus.

Siglo	Cartas en español	Cartas en portugués
XVI	527	307
XVII	1127	1016
XVIII	1584	1247
XIX	651	791
Total	3889	3361

Fig. 4: Distribution of letters found for different centuries. Cf. Vaamonde 2018: 147.

15 Following the latest publication about the project (Vaamonde 2018), it seems that the primary motivation for the creation of the text collection was the scarcity of sources of documenting spoken language for historical linguistics. Archives of courts were chosen as an excellent opportunity to get access to private correspondence, characterized by an immediate, communicative style and not so much influenced by traditional rhetoric. An overview of the consulted archives in Spain shows that most of the letters could be localized in the National Historical Archive. Vaamonde observes that the localization of letters in the archives is a procedure strongly influenced by chance, as systematic catalogues or even databases are rare. In some cases, the process was initiated with random samples. Regarding the chronological distribution of the letters, the author notes that it was hardest to get letters for the 16th century (see [Fig. 4](#) for an overview). Comparing these numbers with the search results above, approximately 70 % of the located letters have been edited in *P. S. Post Scriptum*.

Data model

16 To assess the data model used in *P. S. Post Scriptum*, an impressively detailed manual can be consulted following a link on the homepage. The “Guía para la Edición Digital de Textos en *P. S. Post Scriptum*”²⁸, more than a hundred pages long, was authored by Gael Vaamonde and is also available in an English version²⁹. It is complemented by the “Manual de Edición y Anotación en TEITOK de los Materiales de *P. S. Post Scriptum*”³⁰ authored by Vaamonde and Catarina Magro, this one only available in Spanish and partly in Portuguese. The data in *P. S. Post Scriptum* is modelled in XML-TEI and in the first of the two manuals, the entire TEI model is documented, including the TEI header, transcription and editing conventions, changes to the TEI needed for the linguistic annotation, the relationship of the project’s TEI model to TEI-P5, and the separate TEI model for a bibliographic database containing information about the participants of the letters. In the second manual, it is explained how the texts

have been treated further as a preparation for linguistic annotation (e.g. modernization and sentence splitting) and how they have been annotated morpho-syntactically. The transcription and editing conventions as well as the linguistic annotations will be reviewed in the corresponding sections on *P. S. Post Scriptum* as a scholarly digital edition and as a linguistic corpus below. Here, the general data model is discussed.

17 The data model of *P. S. Post Scriptum* – at least the project internal model – is not a straightforward and plain TEI model for several reasons. As Vaamonde explains, when the project started in 2012, the TEI had not decided to include specific elements for the encoding of correspondence yet.³¹ It was therefore decided to follow the model of another project concerned with the edition of letters, the *Digital Archive of Letters by Flemish Authors and Composers from the 19th & 20th century* (DALF)³². Because the schema used in the DALF project was based on TEI-P4, *P. S. Post Scriptum* inherited an already outdated version of the standard. In a way, this is an ironic twist: the intention was to take up the ongoing development of a dedicated module for the encoding of correspondence and the project ended up with an older version of the chosen XML standard. A good way out of this dilemma was found, though: the decision to offer a TEI-P5 version as an export format and to use the model adopted from DALF as an internal one. The version of the data model conforming to TEI-P5 was influenced by the schema of the CHARTA (*Corpus Hispánico y Americano en la Red: Textos Antiguos*) network.³³ It is to be valued positively that the *P. S. Post Scriptum* project looked for models elsewhere in order to learn from the experiences made in other projects, but especially in the case of the DALF's TEI-P4 model, the question arises why its innovations in terms of correspondence description were not integrated directly into an own TEI-P5 model.

18 The elements specific for the encoding of correspondence were not the only necessary customization of the schema. To add linguistic annotations to the texts, the tool TEITOK (*A Tokenized TEI environment*) was used, a web-based platform developed by Maarten Janssen in which both rich textual markup and linguistic annotations can be used in combination in single TEI files.³⁴ The TEI format used by TEITOK includes custom elements for the tokenization and annotation of the text. A third component of the family of data models in *P. S. Post Scriptum* is a TEI-P4 model inherited from the CARDS project and used for biographical information about the letters' authors and addressees.³⁵

19 On the website, each letter can be downloaded in the “Pure TEI P5 XML” export format and in the “TEITOK XML” version based on the internal TEI-P4 schema. In the P5 version, no schema is declared explicitly, so supposedly the encoding follows the TEI_all schema. In the TEITOK version, the internal DTD is linked. No schema file is offered directly on the website and as the XML of the biographic database is not accessible, its schema cannot be reviewed either.

20 All in all, the combination of different schemas and customizations is a bit confusing for someone not familiar with the project. No explanations are given directly on the website. However, the detailed manual helps the user to follow the history of the project and the complexities of the data model are layed out transparently in these guidelines. Access to all the schema files would have been good to facilitate a formal overview of the structure of the data and especially for reuse scenarios to make consistency checks and schema adaptations possible.

21 However, apart from how the information on the data model is presented, the will to follow best practices developed in similar projects, the orientation towards standards like the TEI, the thorough documentation of the data model and the possibility to view the XML of the letters on the website are all highly commendable.

Metadata

22 Each letter edited in *P. S. Post Scriptum* includes the general administrative and descriptive metadata that are usually part of the TEI header. In addition, there are metadata which are specific for correspondence. The following example shows the correspondence description of the letter PS7004 including details about the sending and reception (names of sender and recipient, locations, and dating), in the TEI-P5 version:

```
<correspDesc>
  <correspAction type="sent">
    <persName key="CDD.xml#ME1" xml:id="ME1" cert="high">
      <name>María de Espinosa</name>
    </persName>
    <placeName confidence="1" type="origin">España, Valladolid
      <location><geo>41.6522510 -4.7245321</geo></location>
    </placeName>
    <date confidence="1" when="1552-05-30" from="1552" to="1552" when-
```

```

custom="1552.05.30"/>
</correspAction>
<correspAction type="received">
  <persName key="CDD.xml#FL6" cert="high">
    <name>Francisco de Leguizamo</name>
  </persName>
  <placeName confidence="1" type="destination">España, Toledo
    <location><geo/></location>
  </placeName>
</correspAction>
</correspDesc>

```

Code 1

23 The situational context of the letters is given in the setting description and a summary of their content in the source description. If there are translations into English, they are stored in the revision description as part of a change element. One would not expect translations there, but it might also have been difficult to add the translations in situ: the setting description could have been doubled, for example, but the summary element is not allowed to occur more than once in the TEI schema. Furthermore, the text is classified according to a project specific taxonomy as can be seen in the following example, also taken from PS7004:

```

<textClass>
  <catRef target="#pragmatics"/>
  <catRef target="#type"/>
  <catRef target="#linguisticSource"/>
  <catRef target="#socioHistoricalSource"/>
  <catRef target="#balancedSelection"/>
</textClass>

```

Code 2

24 The corresponding taxonomy is specified inside the encoding description:

```

<taxonomy xml:id="Tip-PS">
  <bibl>Tipología PS</bibl>

```

```

<category xml:id="pragmatics">
  <catDesc>petición</catDesc>
</category>
<category xml:id="type">
  <catDesc>amor</catDesc>
</category>
<category xml:id="linguisticSource">
  <catDesc/>
</category>
<category xml:id="socioHistoricalSource">
  <catDesc>
    <term n="1" key="KW.xml#K141" />
    <term n="2" key="KW.xml#K193" />
    <term n="3" key="KW.xml#K6" />
    <term n="4" key="KW.xml#K48" />
    Matrimonio, Reproches, Administración, Compraventa
  </catDesc>
</category>
<category xml:id="balancedSelection">
  <catDesc>1</catDesc>
</category>
</taxonomy>

```

Code 3

25 Apparently, the “catRef” elements in the text classification point to the type of category declared in the taxonomy, where the category value is given in the description of the type of category. This seems a bit odd, because a category with the same XML-ID can then have a different category description or value in another letter file (for example, if the type is not “pedición”, but a different pragmatic category. Also, it would be good to have insight into the entire taxonomy.

26 It can furthermore be noticed that some fields are empty (in this example, / location/geo for the destination as well as the category description for “linguisticSource”). It is not clear how many of the metadata fields are empty in the overall collection. In order to know whether empty fields are due to work-in-progress or because a value is not

mandatory, the manual has to be consulted as there is no formal schema for the TEI-P5 files available. On the website, no report is given on the state of the XML files.

27 Besides the metadata for the letters, biographic metadata is collected in the separate demographic database. In the edition manual, it is described in detail how the data about correspondence participants is modelled (for example, the role, sex, and age of a person are given, person names, and a description of the events related to a person). In the letters, there are pointers from the senders and addressees to the external biographic data, but this data is not available for insight or download.

28 The metadata in *P. S. Post Scriptum* is rich and conforms to the standards set by the TEI. Points of criticism which can be raised are the missing access to some external metadata (the biographic database and the complete taxonomy) and a report on the completeness of the data.

Organization, accessibility, and usage possibilities

Biographical data

	Name	Male or Female	Language	Letter Date	Social Status (narrow sense)	Social Status (broad sense)	Residence	Secondary Residence	Birth Place	Events
ISF1	Isaac Semah Ferro, alias Crispim de Gaspar Ferro	m	PT	1672	cônsul	ordinary	Holanda, Middelburg			as suas atividades enquanto cônsul eram controladas pelo cônsul-geral em Haia, Senior Henriquez; assistente na Zelândia, província dos Países Baixos (cf. PSCR1471).
JLSB1	José Luis Soares de Barbosa	m	PT	1757	juiz de fora de Setúbal, ouvidor, advogado	ordinary	Portugal, Setúbal		Portugal, Setúbal	estudou em Coimbra, onde frequentou a Faculdade de Cânones; em 1757, passou por algumas dificuldades financeiras, tendo-se visto obrigado a contrair empréstimos junto do seu círculo de amigos; casou em 1758
AN540	AN540	m	PT	1619		unknown				estava preso no cárcere da Inquisição; recebia bilhetes de outros presos que circulavam escondidos nas panelas da cozinha.
AFM7	Aaron de Fonseca Mesa	m	ES	1672				Suriname		
AA21	Aarão Ailion	m	PT	1783			Caribe, Antilhas Holandesas, Curaçao			
AJ6	Aarão Jessurun	m	PT	1758		ordinary	Holanda, Amesterdão			primo de Jacob de David Jessurun
AMH2	Aarão Moreno Henriques	m	PT	[1758]			Holanda, Amesterdão			irmão de Aarão de [Benjamim] Moreno Henriques e de José Moreno Henriques

Fig. 5: First page of the biographical data table. See <http://ps.clul.ul.pt/index.php?action=cdd>. Accessed June 1, 2018.

29 Access to the letters and additional data related to the correspondence is offered in a number of ways on the *P. S. Post Scriptum* website. The biographical data, for example, can be assessed via the “Participants” entry in the main menu. It is organized in tabular format with one row per person (see Fig. 5). Overall, there are 5,071 rows which can be clicked through in steps of 100 entries. A possibility to sort the data would have been nice, for example by a selected column. As it is, it is not clear if the data is ordered

at all and on what principle. By clicking on one of the names, a detailed view of the entry is shown. The only additional information in the detailed view is a link to the letters written by the person, so the biographic data is a possible starting point to get to the letters.

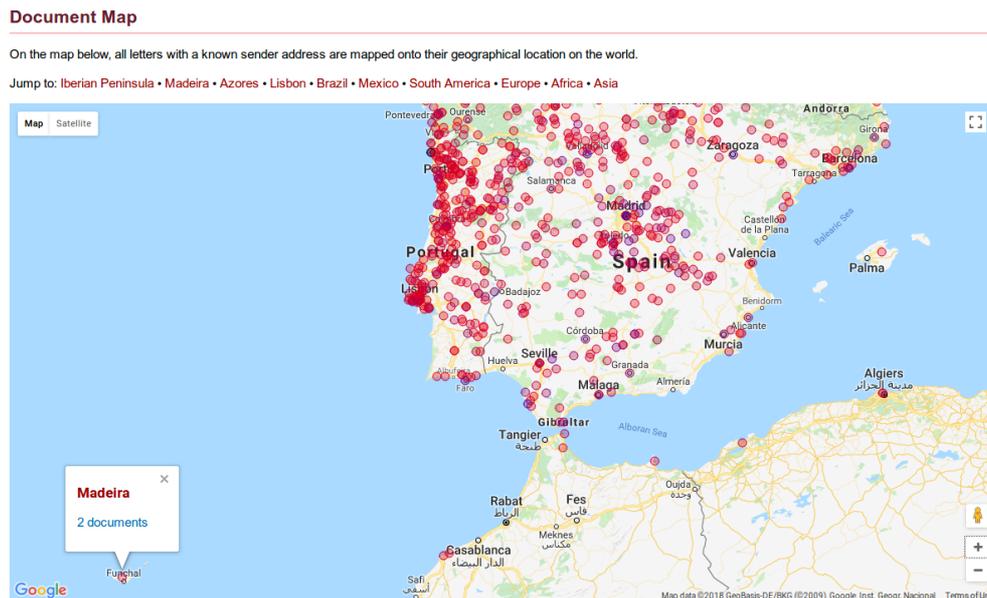


Fig. 6: Document map showing sender addresses. See <http://ps.clul.ul.pt/index.php?action=geomap>. Accessed June 1, 2018.

30 The menu entry “Map” leads to a Google Map showing all the places for letters with a known sender address (see Fig. 6). The documents can be looked up by clicking on the icons for the places.

Document Search

Language	<input type="text" value="[select]"/>
Archive	<input type="text"/>
Year	<input type="text"/> - <input type="text"/>
Letter code	<input type="text"/>
Origin	<input type="text"/>
Destination	<input type="text"/>
Keywords	<input type="text" value="[select]"/>

Author

Author name	<input type="text"/>
Social Status	<input type="text" value="[select]"/>
Male or Female	<input type="text" value="[select]"/>

Addressee

Addressee name	<input type="text"/>
----------------	----------------------

Discourse Search

Discursive type of sentence	<input type="text" value="[select]"/>
Letter type	<input type="text" value="[select]"/>

Fig. 7: Document search. See <http://ps.clul.ul.pt/index.php?action=cqp&act=advanced> and <http://ps.clul.ul.pt/index.php?action=psdx&act=query>. Accessed June 1, 2018.

31 Besides these two possibilities to browse the text collection, the main points of access are the search functions. There are two different searches: the “Corpus Search” (simply named “Search” in the menu) and “Syntactic Trees” (called “Tree Search” in the menu). The corpus search is divided into the sections “Text Search”, “Document Search”, and “Discourse Search”. The tree search is also divided into sections two of which are specific for syntactic searches (both the “XPath Search” and the “Predefined Queries” allow to search for specific syntactic units, e.g. subordinate clauses or clitics) while the third one is a general “Document Search”. The searches specific for the edited text and the syntactic annotation will be discussed below in the section on *P. S. Post Scriptum* as a scholarly digital edition and as a linguistic corpus, respectively. The document search is identical in the two search interfaces and is shown in [Fig. 7](#).

32 The document search allows to look for letters by language, archive, year (or a range of years), letter code, origin, destination, or by a list of thematic keywords (e.g. “Agriculture”, “Masonry”, “Church”). Only one of the keywords can be chosen. Documents can also be searched for with the names of the author or addressee and the social status and gender of the author. A help section with further information about the search options is available. The search is well-documented and easy to use. When the document search is performed, the results are displayed in tabular format. The query is automatically transformed into a *Corpus Query Processor* (CQP) string which is shown together with the search results. The following string, for example, represents a search for letters in Portuguese written by female authors up to the year 1600, yielding 55 results:

```
<text> [] :: match.text_lang = "PT" & int(match.text_year) <= 1600 &
match.text_gender = "f"
```

Code 4

33 Beyond the powerful search functions, more browsing options could have been offered. The only ways to browse the collection are the biographic database and the geographic map, but this archive of letters would have had much more to offer. Full indexes of the letters, for example by identifier, year, place, etc., could have been implemented without much additional effort. The search is especially helpful for someone who already knows what she or he is looking for, but for a user interested in the collection as a whole more browsing options would have been helpful to get an overview of what is there. Examples of scholarly digital editions of letters consistently offering both ways of access are *Letters of 1916*³⁶, where letters can be browsed by category, month, and author’s gender (see [Fig. 8](#)) or the *Alfred Escher-Briefedition*³⁷ access the letters.

34 A great asset of *P. S. Post Scriptum* is the “Downloads” page.³⁸ It is not yet usual that full access to the data underlying an online publication of a text collection is offered, even if it should be expected and is desirable. The RIDE statistics on this topic are still quite disillusioning and fortunately *P. S. Post Scriptum* raises the bars on the Yes-side.³⁹ The letters can be downloaded in several formats and compositions:

- in text format, the whole corpus distributed by language, century, and format (original text, standardized text, POS annotated text, PSDX⁴⁰)

- in text format, the whole corpus distributed by gender, language, century, and format
- in text format, the whole corpus distributed by social status, language, century, and format
- in text format, a balanced⁴¹ corpus distributed by language, century, and format
- in XML-TEI-P5, the whole corpus distributed by language and century
- in XML-TEI-P5, a balanced corpus distributed by language and century

35 With these download options, the data can in principle be reused in multiple ways going beyond the usage scenarios created for the website. The data could for example be prepared for integration into the *correspSearch* environment, a metadata registry and web service for scholarly editions of letters.⁴² However, there is no rights declaration on the “Downloads” page where reuse is encouraged.⁴³ The facsimiles, on the other hand, are not offered for a bulk download and the user is asked to request the authorization of the responsible archive for reuse. Despite the download options for the various text- and XML-based formats of the letters, reuse of the data is limited by a copyright statement in the TEI files, reserving the rights on the data for the Linguistics Centre of the University of Lisbon:

```
<availability status="restricted">
  <p>Copyright 2013, CLUL</p>
</availability>
```

Code 5

36 With this general rights declaration it is not clear if the copyright applies to all the content of the TEI files or just to parts of it, e.g. the markup, metadata, or annotations which were added to the texts in the project. One would expect the original, historical content of the letters not to be subjected to copyright. With these contradictory statements other researchers are left in uncertainty: Can they download and use all the data in external tools without asking for permission? Is a publication of the plain text files elsewhere not allowed or does this just apply to the TEI files? A more detailed licensing in the TEI files and on the website would help to clarify which types of reuse are in principle allowed and for which cases a special permission is needed.

37 Finally, in terms of accessibility and usability of the text collection as a whole, a note on persistent URLs is indicated. The addresses used for the individual webpages

are technical addresses, for example <http://ps.clul.ul.pt/index.php?action=cqp&act=advanced> for the corpus search, <http://ps.clul.ul.pt/index.php?action=psdx&act=query> for the syntactic search, <http://ps.clul.ul.pt/index.php?action=cdd> for the biographical data page, and <http://ps.clul.ul.pt/index.php?action=edit&cid=PS7004>, http://ps.clul.ul.pt/index.php?action=edit&cid=Revistas/ModernizadasTeitok/anotadas_ES/PS7004.xml&tpl=long, or http://ps.clul.ul.pt/index.php?action=edit&cid=Revistas/ModernizadasTeitok/anotadas_ES/PS7004.xml&tpl=translation for the same single letter in different views. Especially for the letters, readable URLs like for example <http://ps.clul.ul.pt/letter/PS7004> or <http://ps.clul.ul.pt/letter/PS7004/translation> would be much better as they would be independent of the technical setup (not referring to a PHP script, for instance), easier to read, and easier to cite.

P. S. Post Scriptum as a scholarly digital edition

Representation of documents and texts

38 The text of the letters is transcribed and encoded in the TEI body and a detailed description of the elements used can be found in the edition manual, together with a chapter on the transcription and editing conventions. The following changes and additions are applied to the text of the manuscripts when it is transcribed:

- the punctuation is standardized (// and = are for example transcribed as a full stop)
- word boundaries are standardized
- the distribution of “i”, “j”, “u”, and “v” is standardized
- abbreviations are expanded (not silently)
- annotations on margins by external hand are not transcribed as part of the text
- page breaks and line breaks are marked
- additions, deletions, replacements, unclear text and gaps are marked
- underlined text is marked
- necessary conjectures are marked with <supplied>
- hand shifts are marked⁴⁴

39 The project members themselves describe their transcription as conservative, respecting many details of the manuscript, so as to serve the goal of *P. S. Post Scriptum* as a multidisciplinary edition:

In the case of Post Scriptum, we understand that the collected documents are interesting as a source for linguistic data, but also as a source for historical data, and even as objects representing fragments of a practice, produced manually by hundreds of persons who lived at a certain time in the Early Modern Age and who put their daily preoccupations on paper. We are definitely faced with a kind of documentation that can and must be approached from three different perspectives: as an artefact, understood as a physical object; as a text, understood as linguistic content; and as a context, understood as the entirety of historical circumstances associated with the text and the artifact. [...] our work as editors must be a meticulous work striving to preserve every detail of the manuscript.

(Vaamonde 2018: 148. Translated from Spanish by the reviewer.)

40 Another level of the text that is edited are the discursive parts of the letters. In addition to the usual opener and closer, the sentences are classified as either harangue, peroration, or narration, for example the peroration “Que o mesmo Senhor lha conserve [...]” (That the Lord himself shall save her [...]):

Presentation of documents and texts

PSCR0669

[1770-1779]. Carta de Maria Inácia para a sua irmã, Gertrudes Felizarda Cogominho Pavia e Vasconcelos Sardinha.

Author(s)

Maria Inácia

Addressee(s)

Gertrudes Felizarda Cogominho Pavia e Vasconcelos Sardinha

Summary

A autora justifica-se na demora da confeção de umas meias para a irmã, cuja encomenda no entanto agradece por ser uma maneira de lhe pagar os favores que lhe deve

► more header data • In English

View options

Text: [Transcription](#) | [Edition](#) | [Variant form](#) | [Standardization](#) | - Show: [Colors](#) | [Formatting](#) | [<lb>](#) | [Images](#) | - Tags: [Linguistic notes](#)

Minha estimada Mana do meu coração, só Deus sabe a alegria que este recebeu quando alcancei as suas notícias, que há muito as desejava e as não tinha procurado por não ter portador. E peço a VM se não descuide de me dar notícias suas e do Mano, a quem me recomendo muito saudosa. As meias, terá a Mana paciência se as não fizer com aquela brevidade que eu desejo, mas será com aquela que puder. E não é preciso pagar-me, que este dinheiro que me deu a portadora, recebi-o como esmola e não por paga, que muito mais devo eu. E, se tiver mais alguma coisa para fazer, não se descuide, que nisto me faz grande gosto. Recebi o mais que VM me mandou. E eu não quero que a Mana tire de si as coisas para me dar, porque me dá pena ver que VM não tem, e que, para me mostrar o seu amor, faz estes extremos. E eu, então, sem lho poder recompensar, mas que tão somente em me confessar agradecida e em rogar a Deus pela saúde do Mano e pela sua. Que o mesmo Senhor lha conserve para alívio desta sua verdadeira e amante irmã do coração,

Maria

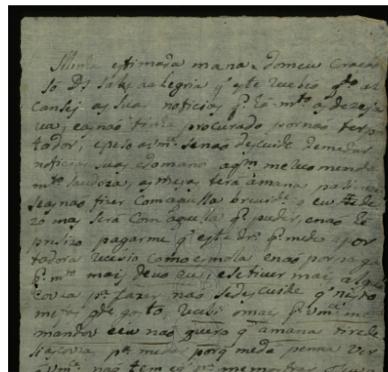


Fig. 8: Presentation of a letter on the *P. S. Post Scriptum* website. See <http://ps.clul.ul.pt/index.php?action=file&cid=PSCR0669>. Accessed June 1, 2018.

41 Each letter is presented on an individual page (see [Fig. 8](#) for an example). The metadata is presented at the top of the page, starting with the autor, addressee, and letter summary. The metadata area can be expanded for more detailed information and switched to English if translations are available. The text of the letter is given on the left side with the corresponding facsimile(s) to the right. The default text view is the diplomatic transcription, but there are several alternatives that can be chosen in the corresponding “View option” bar above the letter text: “Edition”, “Variant form”, and “Standardization”. In the edition view, abbreviations are expanded. The variant forms view shows the edited text where deletions, for example, are excluded, but historical variants (e.g. “cousa” for “coisa”) are kept. In the standardized view, the edited text is given in a modernized form. Especially the last view is very helpful for a user interested in the content of the letters without being familiar with the historical conventions of writing and historical orthography. This is also the text version that forms the basis for the linguistic annotation and is essential for that task, because a modernized, established

text version without variants is usually expected by automatic taggers. The standardized forms of single words are also shown in small boxes as a mouse-over effect.

Sentence view
[1770-1779]. Carta de Maria Inácia para a sua irmã, Gertrudes Felizarda Cogominho Pavia e Vasconcelos Sardinha.

Author(s)
Maria Inácia

Addressee(s)
Gertrudes Felizarda Cogominho Pavia e Vasconcelos Sardinha

Summary
A autora justifica-se na demora da confeção de umas meias para a irmã, cuja encomenda no entanto agradece por ser uma maneira de lhe pagar os favores que lhe deve
[▶ more header data](#) • [In English](#)

[1] Minha estimada Manna do meu coração só Ds sabe a aLegria q este recebeo qdo aLcansei as suas noticias q há mto as dezejava, e as não tinha procurado por não ter portador,

[2] e peso a vmce se não descuide de me dar noticias suas e do mano a qm me recomendo mto saudoza,

[3] as meyas terá a mana pacienci se as não fizer com aquella brevide q eu lhe dezo mas será com aquella q puder,

[4] e não hé presizo pagarme q este dro q me deo a portadora recebio como esmola e não por paga q mto mais devo eu.

[5] e se tiver mais algua couza pa fazer não se descuide q nisto me fas gde gosto,

[6] recebi o mais q vmce me mandou

[7] e eu não quero q a mana tire de si as couzas pa me dar porq me dá penna ver q vmce não tem e q pa me mostrar o seu amor fas estes estremos

[8] e eu então sem lho poder recompensar mas q tão sómte em me comfesar agradesida e em rogar a Ds pella saude do manno e pella sua

[9] q o mesmo Snr lha comserve pa alivio desta sua verdadeira e amte Irmam do coracao

[10] Ma

Text view

Fig. 9: Sentence view of the letter PSCR0669. See <http://ps.clul.ul.pt/index.php?action=block&cid=Revistas/ModernizadasTeitok/PSCR0669.xml&pageid=&jmp=&elm=s>.

Accessed June 1, 2018.

42 Besides the different text views, there are additional display options grouped into the bars “Show” and “Tags”. With the options in the “Show” bar, abbreviations, variants, and standardized forms can be highlighted, the letter text can be displayed with original line breaks, and the facsimiles can be switched on and off. In the “Tags” bar, the different options serve to activate a tokenized version of the letter, to display the part of speech and lemma of each word as well as additional notes on special linguistic characteristics, for example words with clitic doubling in Portuguese, complex syllables, diphthongs or hiatuses. At the bottom of the page, there is a small legend indicating the colors used to display editorial changes to the text. A “Sentence view” is also offered. In this view, the text is split into sentences and displayed as one sentence per line (see [Fig. 9](#)).

43 There are two download options at the bottom of each letter page, one for the download of the XML file, either in the TEI-P5 or in the TEITOK version, the other for the download of the current view in a plain text format.

44 Browsing through various letters, it becomes apparent that not all of them have the same options. Some letter texts are not accompanied by a facsimile (e.g. PSCR5513), some don't have the detailed POS, lemma, linguistic notes and sentence view options (e.g. PSCR5285), or the variant and standardization views (both PSCR5513 and PSCR5285, for example). As for other aspects mentioned before, it would have been good to have an overview of the editing status of the letters in the whole collection. Another way to make the state of progress more transparent would have been to include search options allowing to look for finalized letters.

45 Although most of the functions are self-evident, a help text explaining the meaning of the different text views, for example, would have been helpful. If the view options are not documented exhaustively on the website, a user might miss them if she or he by accident only views letters where a few of them are available. It was not clear to the reviewer how to find a letter where all options are available, if not by chance.

46 Of all the information encoded in the TEI files, the discursive parts of the letters seem to be the only aspect that is not presented. Apart from that, all the different levels of encoding can be assessed via the web interface in a user-friendly manner. In addition, the "Text Search" supports looking for words on the different levels of transcription (provisional transcription, expanded abbreviations, and standardization). Even if the parts of discourse are not taken up into the HTML view, they can still be searched for with the "Discourse Search" which allows to look for letters containing different types of sentences (opener, closer, harangue, peroration, narration, etc.) and for letter types (friendship, love, family, private business and anonymous letters).⁴⁶

P. S. Post Scriptum as a linguistic corpus

47 The second side of *P. S. Post Scriptum's* double face as a digital scholarly resource is its design and capacity as a linguistic corpus, the goal to compile "an electronic resource which facilitates the statistical exploitation and treatment of the textual data" with corpus linguistic methods as the primary approach.⁴⁷

Linguistic annotations

48 Before the several layers of linguistic annotation could be added to the letters in *P. S. Post Scriptum*, the texts had to be normalized. This was achieved through the manual preparation of the standardized, edited text, as described above. The

morphosyntactic annotation was then done with the package *FreeLing* for the Spanish texts and with *eDictor* for the Portuguese ones.⁴⁸ Vaamonde reports that it was a challenge to work with a range of tools providing different output formats that could not be easily combined, especially the information encoded in TEI with the different tagger outputs, leading to different versions of the corpus that cannot be exploited altogether.⁴⁹ This is a common problem in projects working at the junction of different disciplines in digital humanities.⁵⁰ In *P. S. Post Scriptum*, the solution was to centralize the linguistic treatment of the corpus in the tool TEITOK which allows to store and treat both rich textual markup and linguistic annotations in the same XML format.⁵¹ The tokenization, orthographic normalization, lemmatization, and morphosyntactic annotation were all realized in TEITOK after the import of the digitally edited TEI. The result is a format similar to TEI, but with some custom elements and attributes as in the following example from the letter PS7004:

```
<s id="s-2">[...]  
  <tok id="w-8" lemma="el" mfs="DA0FS0" tagsrc="corpus:3">la</tok>  
  <tok id="w-9" lemma="poco" mfs="DI0FS0" tagsrc="corpus:2">poca</tok>  
  <tok id="w-10" nform="cuenta" lemma="cuenta" mfs="NCFS000" tagsrc="corpus:  
2">quenta</tok>  
  <tok id="w-11" lemma="que" mfs="PR0CN000" tagsrc="corpus:2">que</tok>  
  <tok nform="VM" id="w-12" lemma="VM" mfs="NP00000" tagsrc="corpus:1">Vmd</  
tok>  
  <tok id="w-13" nform="hace" lemma="hacer" mfs="VMIP3S0" tagsrc="corpus:  
1">haze</tok>  
  <tok id="w-14" lemma="de" mfs="SPS00" tagsrc="corpus:1">de</tok>  
  <tok id="w-15" nform="mí" lemma="mí" mfs="PP1CS000" tagsrc="corpus:1">my</  
tok>  
  <tok nform="." id="w-16" lemma="." mfs="Fp" tagsrc="corpus:2"><ee/></tok>  
</s>
```

Code 6

49 The original form of each token is kept as the content of the “tok” element. In the “nform” attribute, a normalized version of the token is given. The other attributes serve to hold the lemma, dialectal variants, and so on.⁵² So the basic unit chosen for the XML version of a letter combining digital edition and corpus is the token or word to which

linguistic information is directly added. But the model is also able to keep specific markup from the historical-philological edition. If there is any markup inside a token such as a line break, for example, it is kept. In the case of a span of underlined words, the element to encode the highlighting surrounds several token elements.

50 From the point of view of this reviewer, it is not just a technical question whether different layers of annotation are kept inside a single file or modelled as stand-off markup. Of course a stand-off setup is more difficult to process because units like sentences or words have to be realigned when several layers of annotation are analyzed or presented together. But to find a way to directly integrate both editorial markup and linguistic annotations also contributes to build a rich resource which can be archived in its integrity. In this scenario, the linguistic annotations become part of the edition and are not just another layer which can be added repeatedly in an ad-hoc way, especially if their quality is checked manually. This is not to say that there aren't any arguments for stand-off markup, but that the editorial decision on the way to model annotations contributes to define the type of resource that is created, in this case a critically edited linguistic corpus.

51 In addition to the tokenization and lemmatization, also morphosyntactic annotations and syntactic trees are included in *P. S. Post Scriptum*. For the morphosyntactic annotations, the EAGLES tagset was used. EAGLES has been developed by an initiative of the European Commission and aims to cover the morphological features of most European languages.⁵³ In *P. S. Post Scriptum*, the tagset for Spanish has been used in a slightly modified version which is documented and linked to from the homepage – apparently for both the Spanish and the Portuguese subcorpora.⁵⁴ For the syntactic annotation, the procedures were inspired by the *Penn Parsed Corpora for Historical English*, the *Tycho Brahe Parsed Corpus of Historical Portuguese*, and the CLUL-projects *Syntax Oriented Corpus of Portuguese Dialects* (CORDIAL-SIN) and *Word Order and Word Order Change in Western European Languages* (WOCcWEL).⁵⁵ The second manual for *P. S. Post Scriptum*, the *Manual de Edición y Anotación en TEITOK de los Materiales de P. S. Post Scriptum: Edición Modernizada, Anotación Morfosintáctica (POS), Anotación Sintáctica (en portugués)*⁵⁶ the preparation of the texts for the linguistic annotations, the annotations themselves, and the necessary follow-up works. On the project's homepage, it is stated that about 25% of each balanced corpus (a corpus of one letter per author in Portuguese and Spanish, respectively) has been annotated syntactically. For the morphosyntactic annotation, it is not clear to what extent it has been completed. The results of the morphosyntactic

annotation are directly incorporated into the XML encoding, as can be seen in the above example, whereas the output formats of the syntactic annotation procedure are in principle text-based. They were converted to the XML-based format PSDX (an XML variant of the Penn Treebank format) and then aligned with the main edition file containing the philological and linguistic information by sentence and token, meaning that the syntactic annotation is kept separately, but correlated.⁵⁷ The results of both the morphosyntactic and syntactic annotations were revised manually.

Syntactic Trees

1642. Carta de António Mascarenhas para Cristóvão Leitão de Abreu, ouvidor-geral.

Author(s)
António Mascarenhas

Addressee(s)
Cristóvão Leitão de Abreu

Summary
O autor solidariza-se com o destinatário e concorda com as suas observações, transmitidas em cartas anteriores.
▶ [more header data](#)

Sentence s-1 Chegou o portador e Criado de Vm com os dois esCritos

Sentence s-2 fica aComodado E no q for Em min não lhe falarei:

Sentence s-3 , facil me he de Crer de hu frade q não tem mais de sello q o abito yrse atras de seus appetes tão mal EntemSeonados coanto Vm o demonstra Epor seu EsCritto

Sentence s-4 tãobem não ponho duvida a ser lhe adjunto João de sa porq como em natureza combina hũ com o outro E ainda podera mto bem ser q em Costumes não era rezão lhe faltasse

Sentence s-5 de João gomes de Lemos me Espanto porq o tinha em melhor predicamto dos tres mas como he falto de Algũas Aduellas podia faseimte Vir no q não era bem Viese

Sentence s-6 tãobem Vejo a Vm Estar estemuLado de Andre furtado por quem se pode dizer tem dois Libra nos Domine

Sentence s-7 convem a saber hũ deles de fidalgo de Entre ambas as sellas E o outro E o outro debaixo de semplexidade maldade

Sentence s-8 folgo tenha Vm Anto d amaral Em difirente comseitto de qualquer desses fidalgos

Sentence s-9 Eu Estou Em q lhe fas mta ventagem E q he amigo de seus amigos E fo delle não falara Em servir a Vm Em tudo pois sabe a vontade q Eu tenho tãobem de o fazer

Sentence s-10 E Como debaixo do q digo não aja outra couza Esteja Vm serto q Eu Elle obraremos Em seu servico a quem Ds gde etc

[More files](#) • [Search in this file](#) • [Download file](#)

Fig. 10: “Syntactic Trees” view of the of the letter CARDS3126. See <http://ps.clul.ul.pt/index.php?action=psdx&cid=Revistas/EdictorMerged/CARDS3126.xml&pageid=&jmp=w-195&elm=>. Accessed June 1, 2018.

52 If a syntactic annotation is available for a letter, a corresponding link is shown at the bottom of the page. Clicking on that link, one gets to another page called “Syntactic Trees” (see Fig. 10). In this view, the text is displayed in sentence rows. At the bottom, the link “More files” leads to a list of letter IDs, supposedly all the letters with syntactic annotation. Via the link “Search in this file”, a “Tree Search” can be performed which is limited to the individual file. From the syntactic trees view, the letter can be downloaded in either the Penn Treebank PSD or in the PSDX format. This download option is not available from the main letter page, so some clicks through the subpages are needed to get to it.

1642. Carta de António Mascarenhas para Cristóvão Leitão de Abreu, ouvidor-geral.

Author(s)

António Mascarenhas

Addressee(s)

Cristóvão Leitão de Abreu

Summary

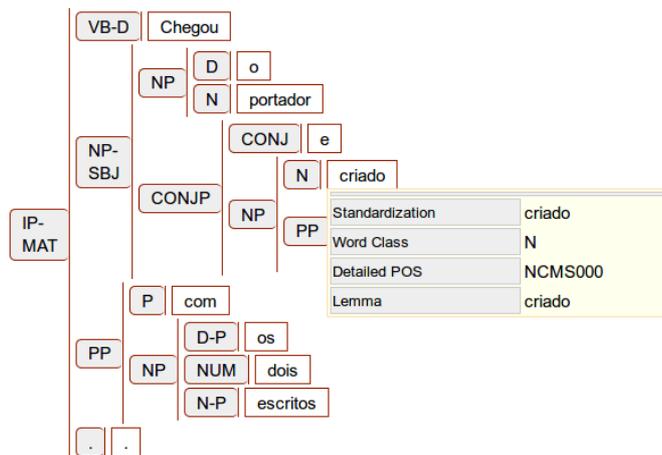
O autor solidariza-se com o destinatário e concorda com as suas observações, transmitidas em cartas anteriores.

► more header data

Tree tree-1 = Sentence s-1

Chegou o portador e Criado de Vm com os dois esCritos

Move your mouse over the leaves in the tree to get info from the corresponding word in the sentence.



sentence list • to text mode • tree style: text - sentence - brackets - table - table graph - vertical graph - svg tree • next sentence

Fig. 11: “Syntactic tree visualization of the letter CARDS3126’s first sentence. See <http://ps.clul.ul.pt/index.php?action=psdx&cid=CARDS3126&treeid=tree-1>. Accessed June 1, 2018.

53 Choosing one of the sentence links, another subpage with a visualization of the syntactic tree opens (see Fig. 11). Hovering over a word in the tree, it is highlighted in the sentence above and additional information (the standardized word form, lemma, and part of speech) is shown in a box. Various formats are offered for the visualization of the syntactic tree: a bracket notation, a tabular format, the table graph (which is the default as in Fig. 11), a vertical graph, and an SVG tree. They are nice to have, for example if one would like to use a specific visual form of a syntactic tree in a presentation or publication, but essentially they all provide the same kind of information.

Search options

Corpus Search

Text Search

Search method: CQP Word Search

Provisional transcription

Expanded abbreviation

Standardization

Word Class

Lemma

Linguistic notes

Tycho POS tag

POS source

Special characters: ~u = ũ, ~e = ê, ~i = ĩ, ~y = ÿ

Display method: KWIC Context

Context size: words

Sort on:

Matching strategy:

Fig. 12: Using the text search to look for the triphthong “uou”. See <http://ps.clul.ul.pt/index.php?action=cqp&act=advanced>. Accessed June 1, 2018.

Corpus Search

CQP Query: advanced

Special characters: ~u = ũ, ~e = ê, ~i = ĩ, ~y = ÿ

Standardization contains uou

10 results

Text: [Transcription](#) | [Edition](#) | [Standardization](#)

Tags: [Word Class](#) | [Detailed POS](#) | [Lemma](#) | [Linguistic notes](#)

context engano para ficarem bem. **Continuou** Sandoval, dizendo-me, q
context , e partio immediatamente. | **Continuou** Alpoim dizendo que elle tinha
context de sorte que elle me **autuou** , e deixou prezo na
context santiçima trindade- **continuou** o cura | em preguntar ao
context . | «Meu amigo, **continuou** Sandoval, não ha remedio
context fonceca Vigro gl e **continuou** os primeiros dias com notavel
context mto menos a cauza q **efeituou** os de sua prizão:
context , fiz como vm me **insinuo** de pedir licença ao
context qto á porção | infirior **suou** sangue; e temeo a
context memdes ha quaza he | tambem **yaguom** he sehou hy he

Download results as TXT - [Direct query URL](#)

Frequency Options

Use the query above to calculate:

Collocation by: | Context size: | direction:

Frequency by:

Fig. 13: Search results as KWIC. See <http://ps.clul.ul.pt/index.php?action=cqp&cql=%5Bnform%3D%22.%2Auou.%2A%22%5D+within+text>. Accessed June 1, 2018.

54 Very much effort has been put into making the linguistic annotations searchable. One way is to use the “Text Search” which is part of the general “Corpus Search”. [Fig. 12](#) shows an example of a search for words in the standardized form containing the triphthong “uou”. Per default, the results are displayed as *Keyword in Context* (KWIC) as in [Fig. 13](#). The search options are manifold: not only can the different transcription layers of the edition be searched but also word classes, lemmas, POS tags, and linguistic notes. Search queries can also directly be formulated in the corpus query language CQP.

55 The results page offers further options to display the findings: by clicking on “context”, it is possible to jump to the text of the whole letter and see the search term highlighted in it. Also, the different textual views can be activated for the search results. A “Direct query URL” is given and can be used to cite the search and its results (at this point the technical character of the URL is understandable because it is meant to reflect complex search queries).

Corpus Distribution

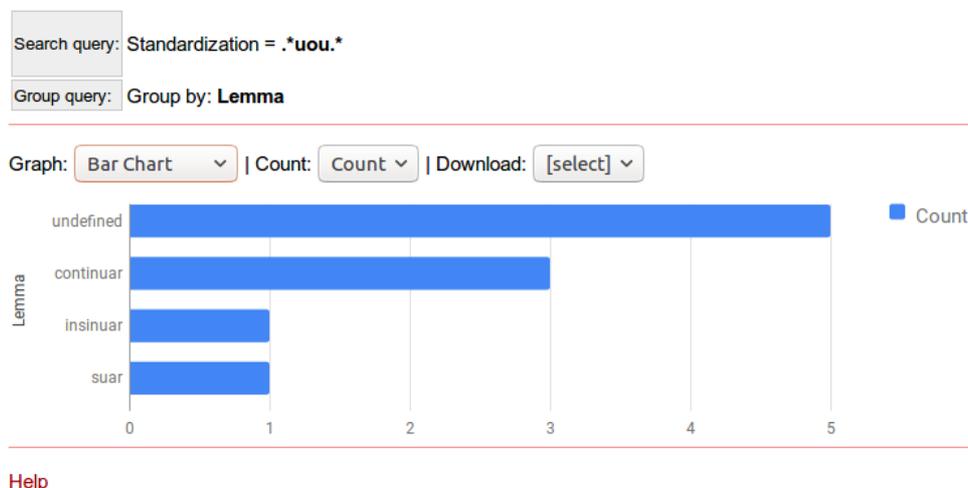


Fig. 14: Frequency of the triphthong “uou” by lemma. See <http://ps.clul.ul.pt/index.php?action=psdx&cid=CARDS3126&treeid=tree-1>. Accessed June 1, 2018.

56 Even more, the results of a query can be used to look for collocations and display frequency tables and visualizations by using the “Frequency Options” below the search results (see [Fig. 14](#) for a bar chart showing the frequency of the triphthong “uou” by lemma).⁵⁸ The frequency calculations are a useful feature and the various visualization options of their results are a convenient extra, but as for the different visualizations of syntactic trees the different types of charts are not essential for their information value, because they all present the same counts. A welcome feature is the possibility to download all the search results in text-based formats for future use.

Syntactic Trees

XPath Search

```
//eTree[contains(@Label, "IP") and ./eTree[@Label="CP-QUE"]]
```

Tree style:

Predefined Queries

Click on one of the named queries below to copy it to the search window

- MATRIX CLAUSES
 - declarative
 - imperative
 - exclamative
 - interrogative
- SUBORDINATE CLAUSES
 - 'that' clauses (subject)
 - 'that' clauses (object)
 - adverbial clauses
 - infinitive clauses (subject)
 - infinitive clauses (object)
 - indirect questions
 - relative clauses
 - ✓ nominal wh (subject)
 - ✓ nominal wh (object)
 - ✓ prepositional wh (indirect object and modifier)
 - ✓ adverbial wh
 - ✓ subject adjoined
 - ✓ object adjoined
 - ✓ other NP adjoined
 - ✓ clause adjoined
 - free relative clauses
 - reduced relative clauses
 - cleft constructions
 - ✓ clefts
 - ✓ basic pseudoclefts
 - ✓ 'e que' inverted pseudoclefts
- CLITICS
 - accusative clitics
 - dative clitics
 - impersonal 'se'
 - inherent 'se'
 - enclisis
 - proclisis
 - clitic climbing
 - clitic doubling
- NULL SUBJECT
 - in matrix clauses
 - in subordinate clauses
 - ✓ in 'that' clauses
 - ✓ in relative clauses
 - ✓ in adverbial clauses
- SECONDARY PREDICATES
 - subject oriented
 - object oriented
- WORD ORDER
 - postverbal subjects

Fig. 15: "Tree Search", a dedicated search form for syntactic annotations. See <http://ps.clul.ul.pt/index.php?action=psdx&act=query>. Accessed June 1, 2018.

Move your mouse over the leaves in the tree to get info from the corresponding word in the sentence.

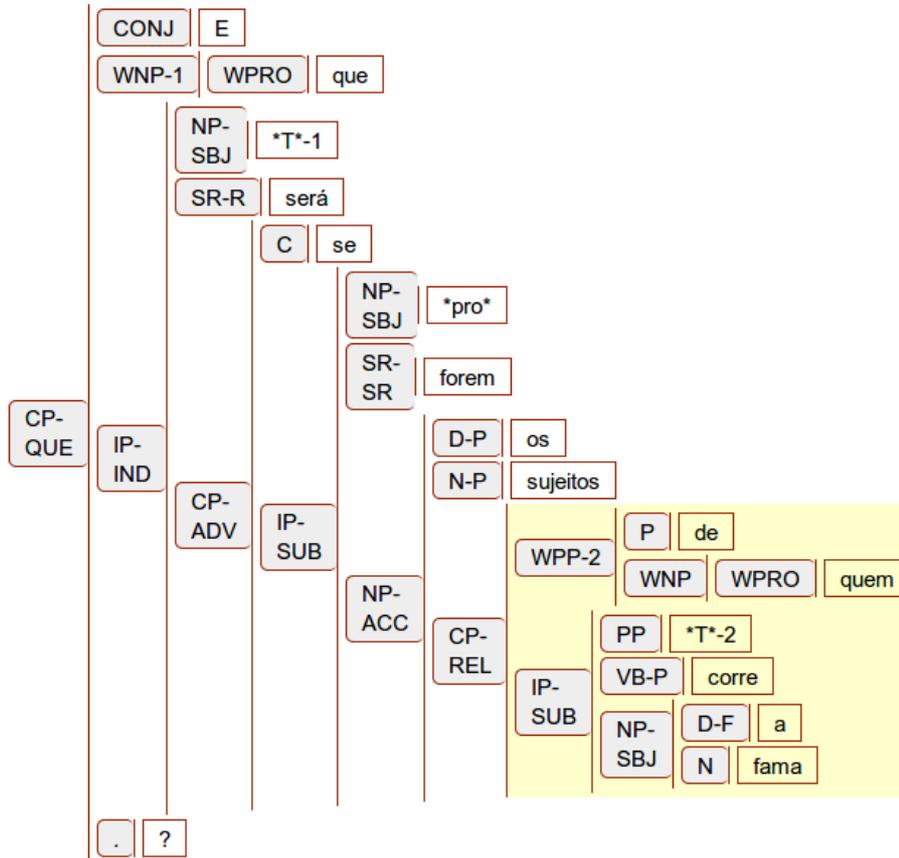


Fig. 16: A subclause with postverbal subject in the letter CARDS0094. See <http://ps.clul.ul.pt/index.php?action=psdx&cid=CARDS0094&sentence=s-17&node=node-888>.

Accessed June 1, 2018.

57 For the syntactic annotations, there is a dedicated search (“Tree Search”, see Fig. 15), with an XPath Search field and a number of predefined queries. It is possible to choose between several options for the display of the results (brackets, table graph, vertical graph, etc.). The XPath search operates on the PSDX format containing the syntactic annotations in XML. The predefined queries help to get started with the XPath expressions which can be quite complex depending on the structure of interest. A search for adverbial clauses, for example, is formulated as `//eTree[@Label="CP-ADV"]`, a search for indirect questions as `//eTree[contains(@Label, "IP") and ./eTree[@Label="CP-QUE"]]`, and for postverbal subjects `/*[eTree[contains(@Label, "NP-SBJ")]/preceding-sibling::eTree[contains(@Label, "VB")]]`. The results are displayed per sentence. A search for postverbal subjects, for example, yields 1,700 results. A tree view of one of the resulting sentences (“corre a fama” – “runs the rumour”) is given in Fig. 16.

58 Both for the general corpus search and for the syntactic search, basic help texts are offered. Not every search option is explained but the information that is given helps the user to find his or her way through the search forms when using the various access points that are offered to explore *P. S. Post Scriptum* as a linguistic corpus.

59 Altogether, a remarkable range and amount of linguistic information has been added to the edited texts of the letters and it has been integrated convincingly into the entire text collection. The many possibilities to search for specific linguistic annotations, to view them in the context of the transcribed documents and to download almost all of the generated information (the linguistic data itself, the search results, and visualizations of corpus distributions as well as of linguistic details such as single clauses) must be a great asset for a historical corpus linguist. While the rich user interface supports qualitative as well as quantitative work, the option to download all data in text- as well as XML-based formats is ideal for researchers who would like to integrate the data into their own technical workflow.

Long-term prospects

60 Officially, *P. S. Post Scriptum* is a finished project. On the website, no comments on the future of the data or the maintenance of the project website itself are found. But even though the funding period ended in 2017, the user manual on the linguistic annotations was still updated in March 2018 and this reviewer got an immediate answer and help from the project coordinator when the website was slowing down during the reviewing process. As the project was based at the Linguistics Centre of the University of Lisbon and is also technically hosted there, the prospects are good that there will be institutional support for this digital text collection also in the future. The graphical user interface is generated with the tool TEITOK, a platform that has already been in use for 14 research projects, so it can be assumed that it will also be maintained further on. Even if the website and user interface should be shut down at some point, the underlying data is held in standard formats and could be kept and used in other environments. Nothing is said on the website about external repositories for the data. It would be an option to additionally store the XML and text data in an infrastructure designed for the long-term, for example at a CLARIN centre.⁵⁹

61 In terms of documentation, it has already been said that many publications have been produced in *P. S. Post Scriptum* contributing to a whole range of disciplines. Not only has the project itself been documented meticulously in the publications and

manuals, but its results as a digital scholarly edition and historical linguistic corpus have also already been made fruitful in analytic research contributions by members of the project, securing the transfer of knowledge from the project to the relevant disciplines.

62 An open question, which all project-based digital resources face, is how to proceed with changes to the data. In *P. S. Post Scriptum*, the levels of encoding and annotation are not the same for all the data. This has probably various reasons: two different languages that need to be dealt with, data coming from preceding projects, decisions to change annotation procedures, the end of the funding period, and many more. Although the ideal would be to have completely homogenous, completed data, it is even more important to communicate the state of the data clearly to give the users orientation. In addition, an explicit contact address, a contact form, or a ticket system could enable users to report anything that they come across while using the platform. This could even be left unmonitored but would make it possible to collect feedback for those who want to reuse the data or in case there would be a follow-up project.

63 For the long run, a more explicit and detailed licensing of the data is also important to ensure that other researchers take the opportunity to work with the results of *P. S. Post Scriptum*. As it is, the project's strategy to grant free access to the data but to publish the TEI files under copyright can not be considered an overall Open Access policy (see the [discussion of rights issues above](#)).

Conclusion

64 *P. S. Post Scriptum* can certainly be classified both as a scholarly digital edition and a diachronic linguistic corpus. It can be considered pioneer work in combining procedures from textual criticism with the linguistic preparation of the corpus — on the levels of the data model, the use of an integrated research environment, and the presentation and possibilities of interaction with the results on the website.

65 Valuable basic work has been done in collecting, transcribing, and encoding the approximately 5,000 historical private letters kept in archives all over the Iberian Peninsula and making them available to researchers and to the public. The decision to make all basic data available for download is exemplary. An open licence for would have been desirable for data reuse, but still it is very useful to have insight into the data.

66 The project has followed community standards and advice from similar editorial undertakings, thereby making its data reusable and prepared for the future. The detailed documentation makes up for a partly complex combination of data standards and tools, originating from the project's history and its double goal to edit the corpus critically and prepare it linguistically.

67 The site's usability would have benefitted from more context information directly presented on the website to guide a user not familiar with the project and also to have the latest information about the selection, amount, distribution, and state of the data in one place. It would have been good to include filters for the availability of annotation layers into the search forms (e.g. to only show letters with completed linguistic annotation). On the other hand, some of the many visualization options for statistics and syntactic trees could have been skipped. Still, the range of search options that have been implemented for the graphical user interface is impressive. It could only have been enhanced by more options to browse through the collection of letters.

68 It is to be hoped that this carefully edited and comprehensively prepared corpus of private correspondence will be maintained and used by many linguists, historians, and digital humanists to come.

Notes

1. For more information about the EU-funded project see "P. S." *CORDIS. Community Research and Development Information Service*. Last updated on November 20, 2017. https://web.archive.org/web/20180603094656/https://cordis.europa.eu/project/rcn/103300_en.html.

2. This review is written from the perspective of a digital humanist trained in Spanish and Portuguese philology, but not a specialist in the cultural history or the history of these languages in the Early Modern period. Its focus is therefore on general philological aspects of the text collection and edition as well as characteristics relevant to the field of digital humanities.

3. A form of language used for a particular purpose or in a particular social setting.

4. Cf. "Project P. S. Post Scriptum." CLUL 2014. <https://web.archive.org/web/20180602124555/http://ps.clul.ul.pt/>, and Vaamonde 2015a: 58.

5. Cf. “1553. Carta de María de Espinosa para su esposo Francisco de Leguizamo.” CLUL 2014. <https://web.archive.org/web/20180602124423/http://ps.clul.ul.pt/index.php?action=file&cid=PS7004>.

6. A description of this project, which was sponsored by the national Portuguese funding institution FCT, can be found on the website of the Linguistics Centre of the University of Lisbon (Cf. “CARDS - Unknown Letters.” *Centro de Linguística da Universidade de Lisboa*. 2018. <https://web.archive.org/web/20180602131349/http://www.clul.ulisboa.pt/en/researchers/10-research/705-cards-unknown-letters>), but an individual website for the CARDS project is not accessible anymore, since it segued into the *P. S. Post Scriptum* project (all the links that the reviewer found are redirected to the *P. S.* project website).

7. General information about the FLY project can also be found at “FLY - Forgotten Letters 1900-1974.” *Centro de Linguística da Universidade de Lisboa*. 2018. <https://web.archive.org/web/20180602151532/http://www.clul.ulisboa.pt/en/10-research/703-fly-1900-1974>. There is also an umbrella page for all three projects. See <https://web.archive.org/web/20180603133455/http://cards-fly.clul.ul.pt/>.

8. Cf. Gomes et al. 2011.

9. The results published on the FLY website will only be taken into account occasionally, where they reveal shifts in the methodology and publication strategy when compared to *P. S. Post Scriptum*.

10. For someone not familiar with the history of the projects, it can be confusing that the credits list is divided into two parts and that it mentions the CARDS project at all. It would be helpful to include a brief explanation of the relationship between the two projects into the website to make it self-contained, because currently this information cannot be found on the credits page, nor anywhere else on the *P. S. Post Scriptum* website.

11. “Credits.” CLUL 2014. <https://web.archive.org/web/20180602160832/http://ps.clul.ul.pt/index.php?action=ficha>.

12. This review therefore draws on both catalogues of criteria developed by the IDE, the *Criteria for Reviewing Scholarly Digital Editions* (<http://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/>. Accessed June 1, 2018) and the *Criteria for Reviewing Digital Text Collections* (<https://www.i-d-e.de/publikationen/weitereschriften/criteria-text-collections-version-1-0/>. Accessed June 1, 2018), the latter including criteria

for the assessment of linguistic corpora. Because the review is published in an issue on scholarly digital editions, the accompanying factsheet is only about editorial aspects, though. Interestingly, *P. S. Post Scriptum* labels itself “A Digital Archive”, avoiding the terms “edition” and “corpus” and placing the focus on the organizational, institutional, and preservational dimension instead of the methodological orientation of the selection and treatment of documents and texts. Similar observations are made for other digital resources in the editorial of RIDE 6. Cf. Henny-Krahmer and Neuber 2017.

15. No documentation was found on this aspect of the resource. In the XML files of the letters only the language of translations is marked but the metadata language is not indicated with a language attribute (@xml:lang). Only the language of translations is marked.

16. Once there, the only way to get back to the original language version is the back button of the browser. It would have been nice to also have a button to switch the metadata language back, but this is just a small inconvenience.

17. “If there is no translation for the letter itself, you may copy the text (while using the view 'Standardization') and paste it to an automatic translator of your choice.” “1553. Carta de María de Espinosa para su esposo Francisco de Leguizamo.” CLUL 2014. <http://web.archive.org/web/20180602124423/http://ps.clul.ul.pt/index.php?action=file&cid=PS7004>.

18. See “1756. Carta de José da Costa Martins para o seu pai, Luís da Costa Martins, negociante.” CLUL 2014. <https://web.archive.org/web/20180605124518/http://ps.clul.ul.pt/index.php?action=file&cid=PSCR0647&tpl=translation> for an example of a letter with a translation into English (found by typing “I have” into the Raw corpus search).

19. Here, “backend” is understood as a non-public part of the website.

20. Cf. “Word distribution.” CLUL 2014. <https://web.archive.org/web/20180603155606/http://ps.clul.ul.pt/index.php?action=cqp&act=distribute>.

21. Cf. “Raw corpus search.” CLUL 2014. <https://web.archive.org/web/20180603155850/http://ps.clul.ul.pt/en/index.php?action=rawsearch>.

22. Cf. “Related Projects.” CLUL 2014. <https://web.archive.org/web/20180603143435/http://ps.clul.ul.pt/en/index.php?action=links>.

23. See Stangl n.d. and *Ultramar*. 2014. <https://web.archive.org/web/20180603144610/http://cards-fly.clul.ul.pt/teitok/ultramar/>.
24. Cf. “Publicações / Publicaciones / Papers”. CLUL 2014. <https://web.archive.org/web/20180603145629/http://ps.clul.ul.pt/en/index.php?action=papers>.
25. Cf. “Word distribution.” CLUL 2014. <https://web.archive.org/web/20180603155606/http://ps.clul.ul.pt/index.php?action=cgp&act=distribute>.
26. These are most probably: Gomes et al. 2012, Marquilhas 2012 (about the Portuguese letters and the CARDS and FLY projects), Vaamonde et al. 2014, Vaamonde 2015a, Vaamonde 2015b, Vaamonde 2017a, Vaamonde 2018 (about *P. S. Post Scriptum* and partly with a focus on the Spanish subcorpus), plus some presentations.
27. In this case, of the ones concerned with the description of *P. S. Post Scriptum* or its predecessors, six are freely accessible online, one is not.
28. See https://web.archive.org/web/20180605131619/http://ps.clul.ul.pt/files/Manual_PS.pdf.
29. See https://web.archive.org/web/20180605133414/http://ps.clul.ul.pt/files/Manual_PS_english.pdf.
30. See https://web.archive.org/web/20180605133516/http://ps.clul.ul.pt/files/Manual_Mod_Pos_Sin.pdf.
31. This was finally achieved in 2015 with the TEI-P5 release 2.8.0. Cf. Stadler et al. 2016-2017.
32. See <http://ctb.kantl.be/project/dalf/>. Accessed June 1, 2018.
33. Cf. Vaamonde 2017c: 5f. The website of the CHARTA network is available at <http://www.redcharta.es/>. Accessed June 1, 2018.
34. See <http://www.teitok.org/>. Accessed June 1, 2018.
35. Cf. Vaamonde 2017c: 83f. and 107-118.
36. See <http://letters1916.maynoothuniversity.ie/>. Accessed June 1, 2018.
37. Cf. Jung 2015.

38. See <https://web.archive.org/web/20180606164919/http://ps.clul.ul.pt/index.php?action=downloads>.

39. See <https://web.archive.org/web/20180606170241/https://ride.i-d-e.de/data/charts/>.

40. The format suitable for the TEITOK environment.

41. The balanced corpus is created by selecting only one letter per author.

42. See <https://correspsearch.net/>. Accessed June 1, 2018.

43. “For the benefit of researchers who want to deal with our data using external tools, we offer them below in a text format.” <https://web.archive.org/web/20180606164919/http://ps.clul.ul.pt/index.php?action=downloads>.

44. See Vaamonde 2017c, 64ff. for details.

46. See <https://web.archive.org/web/20180607152620/http://ps.clul.ul.pt/index.php?action=cqp&act=advanced>.

47. Vaamonde 2018: 152. Translated from Spanish by the reviewer.

48. See Padró and Stanilovsky 2012 and Piaxão de Sousa et al. 2012.

49. Cf. Vaamonde 2018: 153.

50. Similar experiences have been made in the project *Computational Literary Genre Stylistics* (CLiGS) that the reviewer is part of. In CLiGS, it was decided to keep two versions of the corpora of literary texts: a master version with structural and inline TEI encoding and a linguistically annotated version where FreeLing output in XML format is incorporated into a basic TEI structure. Cf. Schöch et al. (forthcoming) and <https://github.com/cligs/textbox>. Accessed June 1, 2018.

51. See Janssen 2016.

52. See Vaamonde and Magro 2018 for details about the meaning of the various attributes used in the linguistic annotation.

53. Cf. <http://www.ilc.cnr.it/EAGLES/browse.html>. Accessed June 1, 2018.

54. Cf. <https://web.archive.org/web/20180608101655/http://ps.clul.ul.pt/index.php?action=tagset>.

55. See <http://www.ling.upenn.edu/hist-corpora/annotation/index.html>, <http://www.tycho.iel.unicamp.br/~tycho/corpus/en/>, <http://www.clul.ulisboa.pt/en/10-research/307-cordial-sin-project-description>, and <http://alfclul.clul.ul.pt/wochwel/>. Accessed June 1, 2018.

56. See Vaamonde and Magro 2018.

57. *Ibid.*: 72-81.

58. In this case, the lemmas for the words “autuou”, “efetuou”, and “jejuou” were not recognized. “Continuou” occurred five times in the search results, but was only lemmatized in three cases. The results in this example might have been influenced by files where the annotation has not been finished.

59. See <https://www.clarin.eu/content/repositories>. Accessed June 1, 2018.

References

CLUL (ed.). 2014. *P. S. Post Scriptum. Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época Moderna*. Accessed June 1, 2018.

<http://ps.clul.ul.pt/>.

Gomes, Mariana, Ana Rita Guilherme, Leonor Tavares, and Rita Marquilhas. 2012. “Projects CARDS and FLY: two multidisciplinary projects within Linguistics.” In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. Istanbul: ELRA, 2833-2837. Accessed June 1, 2018.

http://www.lrec-conf.org/proceedings/lrec2012/pdf/1031_Paper.pdf.

Henny-Krahmer, Ulrike and Frederike Neuber. 2017. “Editorial: Reviewing Digital Text Collections.” *RIDE* 6. doi: 10.18716/ride.a.6.0. Accessed June 1, 2018.

Janssen, Maarten. 2016. “TEITOK: Text-Faithful Annotated Corpora.” In *Proceedings of the Language Resources and Evaluation Conference (LREC 2016)*. Portoroz: ELRA, 4037-4043. Accessed June 1, 2018.

http://www.lrec-conf.org/proceedings/lrec2016/pdf/651_Paper.pdf.

- Jung, Joseph (ed.). 2015. *Digitale Briefedition Alfred Escher*. Zürich: Alfred Escher-Stiftung. Accessed: June 1, 2018.
<https://www.briefedition.alfred-escher.ch/>.
- Marquilhas, Rita, and Iris Hendrickx. 2016. "Avanços nas humanidades digitais." In *Manual de Linguística Portuguesa*, edited by A. M. Martins and Ernestina Carrilho, 252-277. Berlin: De Gruyter.
- Marquilhas, Rita. 2012. "A Historical Digital Archive of Portuguese Letters" In *Letter Writing in Late Modern Europe*, edited by Marina Dossena and Gabriella Del Lungo Camiciotti, 31-43. Amsterdam: John Benjamins.
- McCarty, Willard. 2005. *Humanities Computing*. Basingstoke: Palgrave Macmillan.
- Padró, Lluís, and Evgeny Stanislovsky. 2012. "FreeLing 3.0: Towards Wider Multilinguality." In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. Istanbul: ELRA, 2473-2479. Accessed June 1, 2018.
http://www.lrec-conf.org/proceedings/lrec2012/pdf/430_Paper.pdf.
- Paixão de Sousa, Maria Clara, Fábio Natanel Kepler, and Pablo Picasso Feliciano de Faria. 2012. "E-Dictor. Novas perspectivas na codificação e edição de corpora de textos históricos." In *Caminhos da linguística de corpus*, edited by Tania Shepherd, Tony Berber Sardinha, and Marcia Veirano Pinto, 191-224. Campinas: Mercado de Letras.
- Schöch, Christof, José Calvo Tello, Ulrike Henny-Krahmer, and Stefanie Popp. "The CLiGS textbox: Building and Using Collections of Literary Texts in Romance Languages Encoded in XML-TEI." *Journal of the Text Encoding Initiative (jTEI)* (forthcoming).
- Stangl, Werner (ed.). N.d. *"En el ojo del huracán": Cartas de Ultramar a España, 1823 - Edición digital*.
<https://web.archive.org/web/20180603144250/http://www.cartas-de-ultramar.net>.
- Stadler, Peter, Marcel Illetschko, and Sabine Seifert. 2016-2017. "Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing <correspDesc>." *Journal of the Text Encoding Initiative* 9. doi: 10.4000/jtei.1433. Accessed June 1, 2018.
<https://journals.openedition.org/jtei/1433>.
- Henny-Krahmer, Ulrike. "Bridging edition and corpus: a review of P. S. Post Scriptum: A Digital Archive of Ordinary Writing (Early Modern Portugal and Spain)." *RIDE* 10 (2019). doi: 10.18716/ride.a.10.4. Accessed: 11.08.2021.

- Vaamonde, Gael. 2018. "Escritura epistolar, edición digital y anotación de corpus." *Cuadernos del Instituto Historia de la Lengua* 11, 139-164. Accessed June 1, 2018. <https://www.cilengua.es/tienda/publicacion/cuadernos-del-instituto-historia-de-la-lengua-2018-no11>.
- Vaamonde, Gael. 2017a. "Cartas olvidadas, fuentes explotadas. Edición y anotación de un corpus histórico epistolar." In *Congreso HDH2017. Libro de resúmenes*, edited by Nuria Rodríguez Ortega. Málaga: Universidad de Málaga, 212-215. Accessed June 1, 2018. <http://hdh2017.es/wp-content/uploads/2017/10/Actas-HDH2017.pdf>.
- Vaamonde, Gael. 2017b. *Guía para la Edición Digital de Textos en P. S. Post Scriptum*. Lisboa: Centro de Linguística da Universidade de Lisboa. Accessed June 1, 2018. http://ps.clul.ul.pt/files/Manual_PS.pdf.
- Vaamonde, Gael. 2017c. *Userguide for Digital Edition of Texts in P. S. Post Scriptum*. Translated by Clara Pinto. Lisboa: Centro de Linguística da Universidade de Lisboa. Accessed June 1, 2018. http://ps.clul.ul.pt/files/Manual_PS_english.pdf.
- Vaamonde, Gael. 2015a. "P. S. Post Scriptum: Dos corpus diacrónicos de escritura cotidiana." *Procesamiento del Lenguaje Natural* 55: 57-64. Accessed June 1, 2018. <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5216/3020>.
- Vaamonde, Gael. 2015b. "Limitaciones en el uso de corpus diacrónicos del español. Nuevas aportaciones desde el proyecto de investigación Post Scriptum." *E-Aesla* 1, [1-10]. Accessed June 1, 2018. http://cvc.cervantes.es/lengua/eaesla/eaesla_01.htm.
- Vaamonde, Gael, Ana Luísa Costa, Rita Marquilhas, Clara Pinto, and Fernanda Pratas. 2014. "Post Scriptum: Archivo Digital de Escritura Cotidiana." In *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, edited by Sagrario López Poza and Nieves Pena Sueiro. Janus, Anexo 1, 473-482. Accessed June 1, 2018. <http://www.janusdigital.es/anexo.htm?id=5>.
- Vaamonde, Gael, and Catarina Magro. 2018. *Manual de Edición y Anotación en TEITOK de los Materiales de P. S. Post Scriptum: Edición Modernizada, Anotación*

Morfosintáctica (POS), Anotación Sintáctica (en portugués). Accessed June 1, 2018.

[http://ps.clul.ul.pt/files/Manual Mod Pos Sin.pdf](http://ps.clul.ul.pt/files/Manual_Mod_Pos_Sin.pdf).

Factsheet

Resource reviewed	
Title	P. S. Post Scriptum
Editors	CLUL
URI	http://ps.clul.ul.pt
Publication Date	2014
Date of last access	22.01.2019

Reviewer	
Surname	Henny-Krahmer
First Name	Ulrike
Organization	Universität Würzburg
Place	Würzburg
Email	ulrike.henny (at) uni-wuerzburg.de

Documentation		
Bibliographic description	Is it easily possible to describe the project bibliographically along the schema "responsible editors, publishing/hosting institution, year(s) of publishing"? (cf. Catalogue 1.2)	yes
Contributors	Are the contributors (editors, institutions, associates) of the project fully documented? (cf. Catalogue 1.4)	yes
Contacts	Does the project list contact persons? (cf. Catalogue 1.5)	no
Selection of materials		
Explanation	Is the selection of materials of the project explicitly documented? (cf. Catalogue 2.1)	yes
Reasonability	Is the selection by and large reasonable? (cf. Catalogue 2.1)	yes

Archiving of the data	Does the documentation include information about the long term sustainability of the basic data (archiving of the data)? (cf. Catalogue 4.16)	no
Aims	Are the aims and purposes of the project explicitly documented? (cf. Catalogue 3.1)	yes
Methods	Are the methods employed in the project explicitly documented? (cf. Catalogue 3.1)	yes
Data Model	Does the project document which data model (e.g. TEI) has been used and for what reason? (cf. Catalogue 3.7)	yes
Help	Does the project offer help texts concerning the use of the project? (cf. Catalogue 4.15)	yes
Citation	Does the project supply citation guidelines (i.e. how to cite the project or a part of it)? (cf. Catalogue 4.8)	yes
Completion	Does the edition regard itself as a completed project (i.e. not promise further modifications and additions)? (cf. Catalogue 4.16)	yes
Institutional Curation	Does the project provide information about institutional support for the curation and sustainability of the project? (cf. Catalogue 4.13)	no
Contents		
Previous Edition	Has the material been previously edited (in print or digitally)? (cf. Catalogue 2.2)	yes
Materials Used	Does the edition make use of these previous editions? (cf. Catalogue 2.2)	yes
Introduction	Does the project offer an introduction to the subject-matter (the author(s), the work, its history, the theme, etc.) of the project? (cf. Catalogue 4.15)	yes
Bibliography	Does the project offer a bibliography? (cf. Catalogue 2.3)	yes

Commentary	Does the project offer a scholarly commentary (e.g. notes on unclear passages, interpretation, etc.)? (cf. Catalogue 2.3)	yes
Contexts	Does the project include or link to external resources with contextual material? (cf. Catalogue 2.3)	yes
Images	Does the project offer images of digitised sources? (cf. Catalogue 2.3)	yes
Image quality	Does the project offer images of an acceptable quality? (cf. Catalogue 4.6)	yes
Transcriptions	Is the text fully transcribed? (cf. Catalogue 2.3)	yes
Text quality	Does the project offer texts of an acceptable quality (typos, errors, etc.)? (cf. Catalogue 4.6)	yes
Indices	Does the project feature compilations indices, registers or visualisations that offer alternative ways to access the material? (cf. Catalogue 4.5)	yes
Documents		
Types of documents	Which kinds of documents are at the basis of the project? (cf. Catalogue 1.3 and 2.1)	Collection of texts, Archival holding, Letters
Document era	What era(s) do the documents belong to? (cf. Catalogue 1.3 and 2.1)	Early Modern, Modern
Subject	Which perspective(s) do the editors take towards the edited material? How can the edition be classified in general terms? (cf. Catalogue 1.3)	Philology / Literary Studies
Presentation		
Spin-offs	Does the project offer any spin-offs? (cf. Catalogue 4.11)	none
Browse by	By which categories does the project offer to browse the contents? (cf. Catalogue 4.3)	Authors, Documents, Dates, Persons, Places
Search		
Simple	Does the project offer a simple search? (cf. Catalogue 4.4)	yes

Advanced	Does the project offer an advanced search? (cf. Catalogue 4.4)	yes
Wildcard	Does the search support the use of wildcards? (cf. Catalogue 4.4)	yes
Index	Does the search offer an index of the searched field? (cf. Catalogue 4.4)	no
Suggest functionalities	Does the search offer autocompletion or suggest functionalities? (cf. Catalogue 4.4)	no
Helptext	Does the project offer help texts for the search? (cf. Catalogue 4.4)	yes
Aim		
Audience	Who is the intended audience of the project? (cf. Catalogue 3.3)	Scholars
Typology	Which type fits best for the reviewed project? (cf. Catalogue 3.3 and 5.1)	Diplomatic Edition
Method		
Critical editing	In how far is the text critically edited? (cf. Catalogue 3.6)	Normalization, Emendation
Standards	(cf. Catalogue 3.7)	
XML	Is the data encoded in XML?	yes
Standardized data model	Is the project employing a standardized data model (e.g. TEI)?	yes
Types of text	Which kinds or forms of text are presented? (cf. Catalogue 3.5.)	Facsimiles, Diplomatic transcription, Edited text, Translations, Commentaries
Technical Accessibility		
Persistent Identification and Addressing	Are there persistent identifiers and an addressing system for the edition and/or parts/objects of it and which mechanism is used to that end? (cf. Catalogue 4.8)	none

Interfaces	Are there technical interfaces like OAI-PMH, REST etc., which allow the reuse of the data of the project in other contexts? (cf. Catalogue 4.9)	none
Open Access	Is the edition Open Access?	yes
Accessibility of the basic data	Is the basic data (e.g. the XML) of the project accessible for each part of the edition (e.g. for a page)? (cf. Catalogue 4.12)	yes
Download	Can the entire raw data of the project be downloaded (as a whole)? (cf. Catalogue 4.9)	yes
Reuse	Can you use the data with other tools useful for this kind of content? (cf. Catalogue 4.9)	yes
Rights		
Declared	Are the rights to (re)use the content declared? (cf. Catalogue 4.13)	yes
License	Under what license are the contents released? (cf. Catalogue 4.13)	No explicit license / all rights reserved
Personnel		
Editors	Rita Marquilhas	
Programmers	Maarten Janssen Iris Hendrickx Michel Génèreux Sérgio Gomes Cláudio	
Advisors	Teresa Rebelo da Silva Maria Teresa Oliveira Rosabel San Segundo José Miguel Escribano Páez Sandra Antunes Raïssa Gillier Liliana Romão Teles	

Contributors	Ana Luísa Costa Ana Rita Leitão Catarina Magro Elisa García Prieto Fernanda Pratas Guadalupe Adámez Castro Laura Martínez Martín Gael Vaamonde Víctor Pampliega Pedreira Carmen Serrano Sánchez Catarina Carvalheiro Clara Pinto Leonor Tavares Mariana Gomes Tiago Machado de Castro
---------------------	---