

How do you Compose your Literary Corpus or Literary Collection?

The goal of this survey is to know more about how researchers in literary studies design their research object (corpus or collection) when studying literary texts.

For example, a **project analyzing French novels written by French female writers in the 19th century** uses information about:

- the **author** (gender, nationality)
- and the **work** (original language, genre, year of first publication)

More specifically, we aim to know about which categories are used as criteria for the composition of both digital and non-digital collections of texts.

In this survey, we understand *research objects of literary texts*, *literary corpus*, and *literary collection* as broad synonyms.

Why do we do this?

First, to obtain a clearer picture about the praxis of composition of the research object in (digital) literary studies. Second, to make research infrastructure more useful for researchers. A more exact picture of how literary scholars compose their research object (either printed or digital) can help infrastructure institutions (libraries, institutions curating authority files) and projects (ERICs, NFDI) to know how to develop their tools, repositories and catalogs in a way that meets the community necessities.

When resources have the metadata that the community uses as criteria, the task to create new research objects of literary texts (corpora or collections) will be easier.

Who is doing this survey?

This survey has been prepared and launched by the Text+ Consortium, the priority program Computational Literary Studies (SPP-CLS) and the European project Computational Literary Studies Infrastructure (CLS INFRA). In its preparation, the feedback of several other scholars unrelated to the previously mentioned projects has been considered.



GDPR Statement

The survey results will not contain any personally identifiable information. Your answers will be stored securely on a server in Europe, in compliance with EU GDPR Regulations (<https://eur-lex.europa.eu/eli/reg/2016/679/oj>). Once you have completed the survey, you will have the opportunity to contact us to be included in follow-up activities. Survey responses remain internal to the project team. For more information

about your rights in relation to the GDPR, please visit the website of the Data Protection Commission (<https://www.dataprotection.ie/en/individuals/rights-individuals-under-general-data-protection-regulation>).

There are 33 questions in this survey.

Criteria relating to the author

In this and the following two sections, we ask how important the following categories have been or could be to you as composition criteria for building literary collections or corpora.

First we ask for categories related to the author, then to the work and finally to the text.

With regard to the author, this criterion was or is very important

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|---|-----------------------|---------------------------|-----------------------------|-----------------------|-----------------------|
| Authorship (for example, analyzing only the texts of one or few specific authors) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Multiple authorship (for example, including or excluding texts written by more than one author) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| First language (for example, analyzing the works of authors with a specific first language or mother tongue) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|--|--------------------------|-------------------------------|---------------------------------|-----------------------|---------------------------|
| Nationality (for example, analyzing only the texts of authors having a specific nationality) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Residence (for example, analyzing only texts of authors who were exiled or who lived for an important period of time somewhere) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Place of birth (for example, analyzing only the texts of authors born in a specific area of a country) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Date of birth or death (for example, considering only texts by authors who were born in certain decades) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|---|--------------------------|-------------------------------|---------------------------------|-----------------------|---------------------------|
| Gender (for example, in order to compose a gender balanced corpus or collection) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Ethnicity (for example, considering texts from authors from specific ethnical groups) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Sexual orientation (for example, considering only texts from gay or lesbian authors) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Religious traits (for example, considering texts from authors who are considered Protestants or Catholics) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Political traits (for example, considering texts from authors affiliated or associated to specific parties or political movements) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|---|-----------------------|---------------------------|-----------------------------|-----------------------|-----------------------|
| Belonging to an artistic period (for example, analyzing texts by authors from the romanticism or baroque period) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Main written genre (for example, analyzing texts by authors who are considered playwrights, poets or novelists) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Other criteria about the author: Have you used, or would you like to use, other criteria relating to the author that we have not covered in this part of the survey? Please separate multiple additional criteria with semicolon (;).

Please write your answer here:

Criteria relating to the work

Here are some questions about the literary work, regardless of the specific edition being analyzed. In the next section, we will ask about specific characteristics of the analyzed edition (such as its length).

With regard to the work, how important has this criterion been or could be?

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|--|-----------------------|---------------------------|-----------------------------|-----------------------|-----------------------|
| Original language of the work (for example, considering only works written originally in a specific language) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Year of creation or first publication (for example, considering only works written or published for the first time between certain decades) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Place of creation or first publication (for example, considering only works published or written in a specific country or place) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|---|--------------------------|-------------------------------|---------------------------------|-----------------------|---------------------------|
| Genre and subgenre (for example, only analyzing prose, novels, erotic novels, poems, sonnets, travelogue, essay, literary blog posts...) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Belonging to an artistic period (for example, only texts belonging to the period of romanticism or baroque) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Fictional status (for example, considering only works that are fictional or non-fictional) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|---|-----------------------|---------------------------|-----------------------------|-----------------------|-----------------------|
| Translations (for example, considering only works when, at some point in history, they have been translated into other languages) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Publisher (consider only works who have been published by a specific publisher or in a specific series) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Treatment by publishers (for example, considering works with at least one or several re-prints, works with current editions, printed as some point as volume, etc.) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Treatment by scholars (for example, considering works which were reviewed by scholars, or treated in research articles or present in bibliographies written by scholars) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Literary awards (for example, only considering works awarded with specific literary awards) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Presence in databases (for example, only considering works present in databases such as VIAF, Wikidata, GND, etc.) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

| | Not important | Slightly important | Moderately important | Important | Very important |
|--|-----------------------|---------------------------|-----------------------------|-----------------------|-----------------------|
| Presence in catalogs and libraries (for example, only considering works with editions in some library catalogs) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Other criteria about the work: Have you used, or would you like to use, other criteria relating to the work that we have not covered in this survey? Please separate multiple additional criteria with semicolon (;).

Please write your answer here:

Criteria relating to the analyzed text

Here are some questions about the specific editions or text being analyzed.

With regard to the text, how important has this criterion been or could be?

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|---|-----------------------|---------------------------|-----------------------------|-----------------------|-----------------------|
| Language of the text (considering texts in a specific language) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Language of the text is the same as the original language (considering only texts when they are in their original language; for example, if your project collects texts in German and you admit translations, this criterion would not be important for you; if you would not accept translated texts, this criterion would be very important for you) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|--|-----------------------|---------------------------|-----------------------------|-----------------------|-----------------------|
| Availability of full text (considering only texts which are already available in full text) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Quality of full text (considering only texts digitized with a certain quality of the OCR output) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Type of edition (for example, considering only digital or printed editions with a critical apparatus, or editions prepared by a scholar) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Representation format (for example, considering texts which are already available in a specific digital representation format, such as PDF, HTML, TEI, plain text, tabular format...) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

If the representation format is important for you, which specific formats are specially important?

Please write your answer here:

How important this criterion **has been or could be** for you in the composition of research objects with literary texts (collection or corpora)?

Please choose the appropriate response for each item:

| | Not important | Slightly important | Moderately important | Important | Very important |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Text length (for example, considering only texts with a minimal or a maximal extension; this extension can be measured in different possible units such as pages, tokens, characters, etc.) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Structural characteristics of the text (considering only texts with or without the presence of some textual structure; for example, considering only theater plays in verse or exclude narrative texts when they contain a certain number of verses) | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Other criteria about the text: Have you used, or would you like to use, other criteria relating to the text that we have not covered in this survey? Please separate multiple additional criteria with semicolon (;).

Please write your answer here:

Questions about the project

Almost done! Now we have a few questions about your work!

Answering your questions, you may have considered specific corpora or projects in which you have worked. Please, give us some information about them.

Names of your collections or corpora: if it has any name (here are some examples of names of corpora: ELTeC, ELTeC-eng, Digital Library, Calderón Drama corpus, DisKo, Folger Shakespeare Library, etc.). For several corpora or collections, please separate the names with semicolon (;).

Please write your answer here:

Name of the project: Please, provide the name of the project or projects. For several projects, please separate the names with semicolon (;).

Please write your answer here:

Link(s) to the project(s): Please, provide link or links to the project websites. For several links, please separate the names with semicolon (;).

Please write your answer here:

Questions about the publication of the data

Now, a few questions about the publication of the data (metadata and texts).

References or Metadata: Have you published online the references to the analyzed works or editions, for example as metadata of the corpus or collection (in any format, including a list of primary sources in a bibliography, in a printed document, PDF file or Word document, but also other formats such as TEI or RDF)?

Please choose **only one** of the following:

Yes

No

If you have published references or metadata, please, provide links as examples.

Please write your answer here:

Texts: Have you published the texts (i.e. the corpus)?

Please choose **only one** of the following:

Yes

No

If you have published texts, please, provide links as examples.

Please write your answer here:

Description: Have you published articles presenting the composition criteria or describing or analyzing in any form the metadata or the texts?

Please choose **only one** of the following:

Yes

No

If you have published any article, please, provide links as examples.

Please write your answer here:

Personal questions

Last questions! The questions in this section are mainly motivated by two reasons.

- The first is that we want to be critical of our own findings and know more about our potential biases in the answers of this survey.
- The second is to be able to analyze whether certain composition criteria are associated with certain characteristics of the researchers.

Our goal is to try to know if the unavailability of certain metadata in resources hinders the work of certain researchers, and to suggest how infrastructure projects can be more sensitive to the needs of different groups.

Each question in this section is optional, feel free to answer or ignore any of them. If you have any feedback, we would be happy to hear from you.

In which university do you currently work?

Please write your answer here:

Gender

❗ Choose one of the following answers

Please choose **only one** of the following:

non-binary

female

male

Other

Which position do you currently have

❗ Choose one of the following answers

Please choose **only one** of the following:

Student

PhD Student

Post-doc

Professor

Other position within academia

Other position outside the academia

Other

Age

❗ Only numbers may be entered in this field.

Please write your answer here:

Using a broad concept of digital humanities (considering also areas such as corpus linguistics, digital editions or computational humanities), **do you consider that you work within the digital humanities?**

Please choose **only one** of the following:

- Yes
 No

In which field do you consider yourself to be in?

(These categories are based on the library classification system Basic Classification (<https://wiki.k10plus.de/pages/viewpage.action?pageId=437452809>)).

! Check all that apply

Please choose **all** that apply:

- Information and Documentation
 Philosophy
 Theology, religious studies
 History
 Linguistics
 Literature
 Arts
 Theater, film, music
 Other field in the Humanities
 Computer science
 Other engineers
 Field in the natural sciences
 Field in the social sciences
 Other
 No answer

If in the previous question you have selected any category with the word "other", Natural or Social sciences, please, specify the field or fields you consider to work in.

Please write your answer here:

Literary texts in which languages or groups of languages do you usually analyze?

(These categories are based on the library classification system Basic Classification (<https://wiki.k10plus.de/pages/viewpage.action?pageId=437452809>)).

🗨 Check all that apply

Please choose **all** that apply:

- English
- German
- Other Germanic languages
- French
- Italian
- Spanish
- Other Romance languages
- Latin
- Old Greek
- Modern Greek
- Slavic languages
- Other European languages
- Hebrew
- Arabic
- Chinese
- Japanese
- Other Asiatic languages
- African languages
- American languages
- Austronesian languages
- Other languages
- No answer

If in the previous question you have selected any category with the word "other", please, specify the language or languages that you normally analyze.

Please write your answer here:

In which country do you currently work?

Please write your answer here:

Thank you for participating in the survey and helping us make infrastructure resources more accessible to literature researchers!

If you would like to stay in touch, please click here and send us an email with "Keep me informed about the composition criteria survey" (<mailto:calvotello@sub.uni-goettingen.de?Subject=Keep%20me%20informed%20about%20the%20composition%20criteria%20survey>) in the subject line.

09.10.2023 – 09:12

Submit your survey.

Thank you for completing this survey.