



HAL
open science

Open Stemmata: A Digital Collection of Textual Genealogies

Jean-Baptiste Camps, Simon Gabay, Gustavo Fernández Riva

► To cite this version:

Jean-Baptiste Camps, Simon Gabay, Gustavo Fernández Riva. Open Stemmata: A Digital Collection of Textual Genealogies. EADH2021: Interdisciplinary Perspectives on Data, 2nd International Conference of the European Association for Digital Humanities, Krasnoyarsk, 2021, 2021, Krasnoyarsk, Russia. halshs-03260086

HAL Id: halshs-03260086

<https://shs.hal.science/halshs-03260086v1>

Submitted on 14 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Open Stemmata: A Digital Collection of Textual Genealogies

Jean-Baptiste Camps¹, Simon Gabay², and Gustavo Fernández Riva³

¹École nationale des chartes | PSL (France)

²Université de Genève (Switzerland)

³Universität Heidelberg (Germany)

1 Gathering stemmata

Stemma codicum is the genealogical tree of the manuscripts of a given text. More precisely, it is a tool of textual criticism that represents the relationships between all the witnesses of a specific work (Duval 2015; Roelli 2020) under the form of a tree, or, in the case of contamination, a directed acyclic graph (Andrews and Macé 2012). The first stemmata were drawn in the 1830’s, with at least one antecedent in the 18th century, while the method to build them progressively took form during the 19th century and is best called the “common errors” method (Camps and Cafiero 2014).

In a stemma, the relationships between the witnesses and the hypothetical necessary (lost) nodes are represented with a tree-like structure (cf. fig. 1a). The original purpose of the stemma is to allow for the reconstruction of the archetype or the original text, as conceived by its author, but stemmata are also used to study the transmission and the reception of works over centuries (Marshall and Leighton Durham 1998), though this has sometimes spurred some debate (Croenen 2010; Varvaro 2010).

Historically, (personal) collections of stemmata have been used in epistemological debates on the common error methods. Joseph Bedier reports that he took the habit of tracing the stemmata he encountered, and built a collection of 110 stemmata, out of which he construed his remarks on the bifidity, from which he derived his criticism of the common error methods (Bédier 1928). To falsify or confirm Bédier’s claims, other scholars have replicated his endeavour, building their own collection (Castellani 1957; Haugen 2015; Shepard 1930) but, as far as we noticed, there is yet no publicly available digital collection of stemmata.

A publicly available collection of stemmata would have a great interest, because controversies on their shape are at the heart of the great philological

debate of the 20th century, at least for Romance Philology, where Bédier claims led some philologists to renounce the common error method (e.g. in France), while it gave other philological schools the necessary impulse to try and refine it (e.g., in Italy, see Trovato 2014)¹. The development of computational philology has been a good opportunity to think anew this old art (Hoenen 2020).

Moreover, study on the shapes of stemmata could prove interesting in many kinds of analysis, be it on the dissemination, reception and history of texts, or to be compared to more theoretical models.

All of this triggers the need for a digital collections of stemmata, for mining and testing purposes. This paper presents an attempt to answer this need. We therefore propose to collect all available stemmata and encode them (cf. fig. 1). Chronological and linguistic boundaries are very open: at least any European language is accepted, from every time period, even if we will focus first on Western Medieval languages (esp. French, Occitan, Italian, German, Spanish and of course Latin).

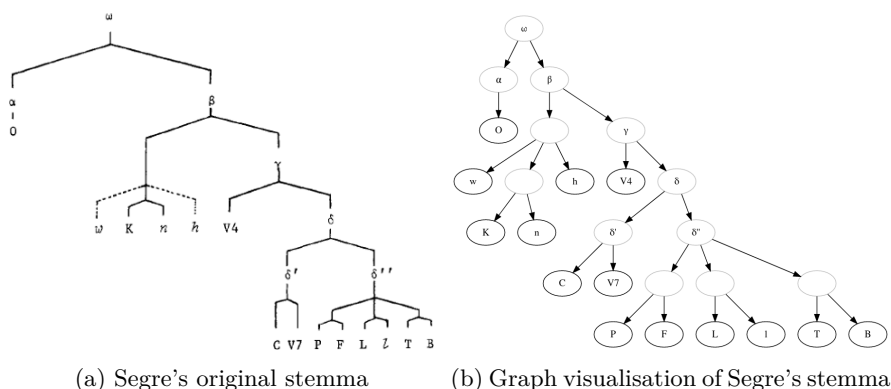


Figure 1: The stemma of the *Chanson de Roland* and its Graphviz representation. Each node represents an extant witness or hypothetical lost manuscript, needed to explain the relations between the witnesses. As is frequently the case in stemmatology, the Greek letters stand for hypothetical lost nodes (archetypes, sub-archetypes,...), and roman letters for extant witnesses (capitals for complete manuscripts and minuscules for fragments, or, here, medieval translations). Some of the hypothetical nodes are not named. In the Graphviz representation, all hypothetical nodes are circled in grey.

2 Information, modelling and production

Open Stemmata is a collaborative project, where researchers can participate by sharing stemmata that they encode. Guidelines have been published to help

¹This even led some philologists to suggest alternatives to the tree-like structure, such as Venn diagram, unrooted graphs... (Carsten-Peust 2012). In the case of some very complex traditions, some editors renounce to draw a stemma (Perugi 2000; comp. to Zufferey 2007).

volunteers (OpenStemmata 2021), who are required to provide three documents for each stemma:

- Metadata as a `.txt` file (such as source of the stemma and researcher responsible for the encoding).
- A stemma encoded in the DOT format (Graphviz 2021), as a `.gv` file
- A picture of the stemma, as a cropped PNG.

Metadata are produced with an online form² to ensure the homogeneity and the completeness of the data. Minimal standardisation is offered via VIAF (OCLC 2020) for author and work, and ORCID (ORCID 2021) for submission contributor.

A stemma being a graph, we have decided to encode data with the DOT language (Graphviz 2021) (fig. 2) which provide all the subtlety required and has a simple syntax that enables easy manual encoding. Hypothetical nodes are identified with “grey” as the value of the color attribute, contamination as dashed lines. We suggest the use of an online graphviz editor (*Edotor* n.d.), but many other options are available.

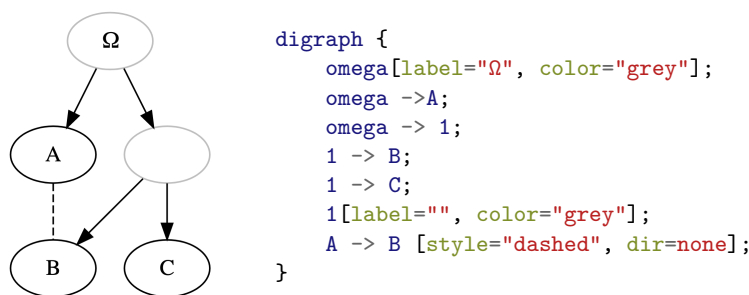


Figure 2: *Example of graph and its encoding in DOT language*

On top of the online form for the metadata, additional steps are in place to ensure data quality, documentation and re-usability.

- On the one hand, data is shared, published and distributed via GitHub. It is compulsory to add stemmata with pull requests, which have to be validated by a member of the organisation. Data structure is then controlled through continuous integration and tests, but also by human inspection and quality control.
- On the other hand, information is stored in a masterfile encoded according to the TEI Guidelines (TEI Consortium 2020): metadata in the `teiHeader` and the graph in the `body`. It allows basic interoperability and sustainability of our digital collection, but also a good documentation of the encoding choices (with a dedicated ODD file and schema).

²Available on: <https://openstemmata.github.io/document-your-stemma.html>.

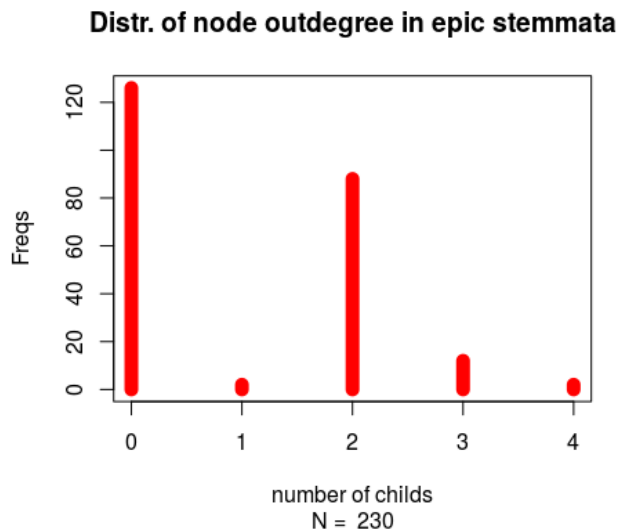


Figure 3: Distribution of nodes (witnesses and hypothetical nodes) in a collection of Old French Epic stemmata (*Chansons de geste*)

- Conversion scripts into GraphML, a common exchange format for graph and network analysis, as well as instructions on how to use them, will be available in the repository, in order to facilitate the re-use of the data for other researchers.

3 Present and Future of OpenStemmata

At present, there are roughly 50 stemmata available in the database, most of them concerning Old French traditions. In the future, we hope to have a coverage as exhaustive as possible. Such data will allow many kind of analyses on the shape of stemmata, including comparative or specific analyses for different types of traditions (fig. 3).

References

- Andrews, T. and C. Macé** (2012). “Trees of Texts: Models and methods for an updated theory of medieval text stemmatology | Digital Humanities 2012”. In: *DH 2012*. Hamburg. URL: <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/trees-of-texts-models-and-methods-for-an-updated-theory-of-medieval-text-stemmatology.1.html> (visited on 10/29/2016).
- Bédier, J.** (1928). “La tradition manuscrite du Lai de l’ombre : réflexions sur l’art d’éditer les anciens textes”. In: *Romania* 54, 161–196 et 321–356. DOI: 10.3406/roma.1928.4345. URL: https://www.persee.fr/doc/roma_0035-8029_1928_num_54_214_4345.
- Camps, J.-B. and F. Cafiero** (2014). “Genealogical variant locations and simplified stemma: a test case”. In: *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*. Ed. by **T. Andrews and C. Macé**. Lectio 1. Turnhout: Brepols, pp. 69–93. DOI: 10.1484/M.LECTIO-EB.5.102565. URL: <https://halshs.archives-ouvertes.fr/halshs-01435633/>.
- Carsten-Peust, K.** (2012). “The stemma of the story of Sinuhe. or: How to use an unrooted phylogenetic tree in textual criticism”. In: *Lingua Aegyptia* 20, pp. 209–220. DOI: 10.11588/propylaeumdok.00002543. URL: <http://archiv.ub.uni-heidelberg.de/propylaeumdok/2543/> (visited on 01/09/2021).
- Castellani, A.** (1957). *Bédier avait-il raison?: La méthode de Lachmann dans les éditions de textes du Moyen Age*. Discours universitaires, Nouvelle série = Freiburger Universitätsreden, Neue Folge. - Freiburg/Schweiz : Universitätsverlag, 1940-... 20. Fribourg: Suisse. Éditions Universitaires. 62 pp.
- Croenen, G.** (2010). “Stemmata, Philology and Textual History: A Response to Alberto Varvaro”. In: *Medioevo Romanzo* 34, pp. 422–426.
- Duval, F.** (2015). *Les mots de l’édition de textes*. Magister. Paris: École nationale des chartes. ISBN: 978-2-35723-054-5.
- Edotor* (n.d.). <https://edotor.net/>. Accessed: 2020-01-08.
- Graphviz** (2021). *The DOT Language: Documentation*. <https://www.graphviz.org/doc/info/lang.html>. Accessed: 2020-01-08.
- Haugen, O. E.** (2015). “The silva portentosa of stemmatology: Bifurcation in the recension of Old Norse manuscripts”. In: *Digital Scholarship in the Humanities* 31.3, pp. 594–610. DOI: 10.1093/llc/fqv002. URL: <https://academic.oup.com/dsh/article/31/3/594/2340338>.
- Hoenen, A.** (2020). “History of computer-assisted stemmatology”. In: *Handbook of Stemmatology: History, Methodology, Digital Approaches*. Berlin: De Gruyter, pp. 294–303. URL: <https://doi.org/10.1515/9783110684384-006>.
- Marshall, P. K. and R. Leighton Durham** (1998). *Texts and transmission: a survey of the Latin classics*. Oxford: Clarendon Press.
- OCLC** (2020). *Virtual International Authority File (VIAF)*. URL: <https://viaf.org/> (visited on 01/10/2021).

- OpenStemmata** (2021). *Guidelines*. URL: <https://openstemmata.github.io/guidelines.html> (visited on 01/10/2021).
- ORCID** (2021). *Open Researcher and Contributor ID*. URL: <https://orcid.org/> (visited on 01/10/2021).
- Perugi, M.**, ed. (2000). *La Vie de Saint Alexis*. Textes littéraires français 529. Genève: Droz. 319 pp. ISBN: 978-2-600-00456-5.
- Roelli, P.** (2020). *Handbook of Stemmata: History, Methodology, Digital Approaches*. Berlin: De Gruyter. ISBN: 978-3-11-068438-4. URL: <https://doi.org/10.1515/9783110684384> (visited on 01/09/2021).
- Shepard, W. P.** (1930). “Recent theories of textual criticism”. In: *Modern Philology* 28.2. Publisher: University of Chicago Press, pp. 129–141.
- TEI Consortium** (2020). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version Version 4.1.0. URL: <https://tei-c.org/Guidelines/>.
- Trovato, P.** (2014). “Bédier’s Contribution to the Accomplishment of Stemmatic Method: An Italian Perspective”. In: *Textual Cultures* 9.1. Publisher: JSTOR, pp. 160–176.
- Varvaro, A.** (2010). “Un nuovo studio sulla tradizione delle Chroniques di Jean Froissart”. In: *Medioevo Romano* 34, pp. 145–152.
- Zufferey, F.** (2007). “La Tradition manuscrite du Saint Alexis primitif”. In: *Romania* 125.497. Publisher: Persée - Portail des revues scientifiques en SHS, pp. 1–45. DOI: 10.3406/roma.2007.1387. URL: https://www.persee.fr/doc/roma_0035-8029_2007_num_125_497_1387 (visited on 01/09/2021).