

Graphs vs trees: encoding stemmata in TEI

Simon Gabay¹, Jean-Baptiste Camps², and Gustavo Fernandez Riva³

¹Université de Genève (Switzerland)

²École nationale des chartes | PSL (France)

³Universität Heidelberg (Germany)

To understand textual traditions of more than one witness it is necessary to establish the relationships between the surviving manuscripts and to express them efficiently. To this end, philologists have developed a genealogical method known as stemmatology. For almost two centuries, they have used tree-shaped diagrams, called *stemma codicum* (plur. *stemmata*, cf. fig. 1), to offer a visual representation of the filiation. This structure is sometimes more complex, because of phenomena like lateral transmission or multiple ancestry (contamination). That two manuscripts from two different families are related would mean, following the arboreal metaphor, that two specific leaves of two different tree branches are connected to one another (cf. fig. 2) – a problem that forces us to rethink the nature of *stemmata*.

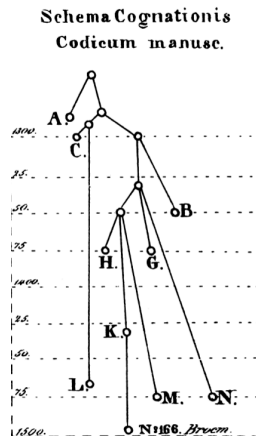


Figure 1: First known stemma, in Schlyter and Collin 1827

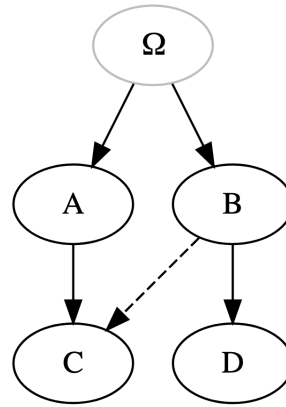


Figure 2: *stemma codicum* with a contamination $B \rightarrow C$

Because the shape containing nodes with more than one parent does not fit

the graph theory definition of a “tree”, it would therefore certainly be more appropriate to talk about “directed acyclic graphs” (Hoenen 2020). Such a decision is not only terminological, but could have an impact on some suggestions of the TEI guidelines, since these latter promote a tree-like encoding using the `eTree` element (TEI Consortium 2021, cf. fig. 3a), which is unable to express efficiently abnormal configurations such as contaminations, contrary to the `graph` element (cf. fig. 3b), in combination with `node` and `arc`. These elements were already associated with the attributes we required, with the exception of `arc/type`, that we added on our own namespace. In addition, the structure of the `graph` element was in line with standard graph encoding formats such as DOT or GraphML (Gansner 2002; GraphML Team 2002), that deal separately with the description of nodes and edges.

| | |
|---|---|
| | <pre> <graph type="directed"> <node xml:id="omega" type="hypothetical" inDegree="0" outDegree="2"> <label>Ω</label> </node> <node xml:id="A" type="witness" inDegree="1" outDegree="1"> <label>A</label> </node> <node xml:id="B" type="witness" inDegree="1" outDegree="2"> <label>B</label> </node> <node xml:id="C" type="witness" inDegree="2" outDegree="0"> <label>C</label> </node> <node xml:id="D" type="witness" inDegree="1" outDegree="0"> <label>D</label> </node> <arc cert="unknown" from="#omega" to="#A" od:type="filiation"/> <arc cert="unknown" from="#omega" to="#B" od:type="filiation"/> <arc cert="unknown" from="#A" to="#C" od:type="filiation"/> <arc cert="unknown" from="#B" to="#D" od:type="filiation"/> <arc cert="unknown" from="#B" to="#D" od:type="contamination"/> </graph> </pre> |
| <pre> <eTree type="hypothetical"> <label>Ω</label> <eTree type="extant"> <label>A</label> <eLeaf type="extant" xml:id="C"> <label>C</label> </eLeaf> </eTree> <eTree type="extant"> <label>B</label> <eLeaf type="extant"> <label>D</label> </eLeaf> <ptr type="contamination" target="#C"/> </eTree> </eTree> </pre> | |

(a) Encoding as embedded trees (b) Encoding as a graph

Figure 3: Competing TEI encodings of the stemma presented in fig. 2.

With the emergence of computational philology, the use of stemmata is slowly drifting from “simple” ecdotal purposes to broader questions regarding textual variation (Andrews and Macé 2013) or the modelling of textual transmission (Camps and Randon-Furling 2018). Using cases contained in the newly created digital collection of textual genealogies, *Open Stemmata* (Camps, Gabay, and Riva 2021), we propose to discuss competing encoding choices of *stemmata* and the underlying philological reasons behind the different possible options. The presentation will include a collection of examples.

References

- Andrews, T. L. and C. Macé** (2013). “Beyond the tree of texts: Building an empirical model of scribal variation through graph analysis of texts and stemmata”. In: *Literary and Linguistic Computing* 28.4, pp. 504–521. ISSN: 0268-1145. DOI: 10.1093/llc/fqt032. URL: <https://doi.org/10.1093/llc/fqt032> (visited on 07/27/2021).
- Camps, J.-B., S. Gabay, and G. F. Riva** (2021). “Open Stemmata: A Digital Collection of Textual Genealogies”. en. In: *EADH2021: Interdisciplinary Perspectives on Data, 2nd International Conference of the European Association for Digital Humanities*. Krasnoyarsk. URL: <https://halshs.archives-ouvertes.fr/halshs-03260086> (visited on 10/15/2021).
- Camps, J.-B. and J. Randon-Furling** (2018). “A Dynamic Model of Manuscript Transmission”. In: *Workshop on Computational Methods in the Humanities (COMHUM 2018)*. Lausanne.
- Gansner, E. R.** (2002). *The DOT language*. URL: www.research.att.com/~erg/%20graphviz/info/lang.html/.
- GraphML Team** (2002). *The GraphML File Format*. URL: <http://graphml.graphdrawing.org/> (visited on 03/24/2014).
- Hoenen, A.** (2020). “The stemma as a computational model”. In: *Handbook of Stemmatology. History, Methodology, Digital Approaches*. De Gruyter Reference. Berlin/Boston: De Gruyter, pp. 226–241. URL: <https://doi.org/10.1515/9783110684384>.
- Schlyter, C. J. and H. S. Collin**, eds. (1827). *Corpus juris Sueo-Gotorum antiqui*. 13 vols. Stockholm: Z. Haeggström.
- TEI Consortium** (2021). *TEI Guidelines, 4.2.1*. <https://tei-c.org/release/doc/tei-p5-doc/en/html/GD.html>.