

SÉBASTIEN MOUREAU

*FNRS/UCLouvain*

TOWARDS EASIER  
AND FASTER COLLATION,  
EDITION AND VISUALIZATION  
OF TEXTUAL VARIANCE?  
THE CHRYSOCOLLATE PROGRAM\*

*Introduction*

The collation, critical editing, and visualization of textual variance have always been a rignarole. Comparing multiple witnesses, analyzing variants, and structuring a coherent edition requires meticulous effort and precision. Over the past few decades, the integration of digital tools into textual scholarship has prompted reflections on how established methods interact with new technological possibilities. In recent years, various digital tools have been developed to assist scholars in different aspects of this work, improving efficiency and accuracy. ChrysoCollate is one such tool, distinguished by features that were not previously available in a single platform. It is a free software tool for collating and editing texts in any language using Unicode characters, as well as visualize textual variations.<sup>1</sup> Designed to facilitate and accelerate the editorial process, it integrates collation, visualization, and edition within a workflow that, fundamentally, remains in line with the established scholarly methods. Developed by myself, it was officially released in December 2021.

The following pages take the form of an essay more than a formal paper. After describing ChrysoCollate and its features, I offer the reader reflections on the program, the advantages and limitations of this type of software, and a discussion of the impact of such tools on editorial practices.

\* I am grateful to Marion Dapsens, Daria Kohler (Kondakova) and the other editors, as well the two anonymous reviewers of this paper, for their suggestions and corrections.

<sup>1</sup> <https://cental.uclouvain.be/chrysocollate> (accessed: August 2025).

### 1. *The Origins of ChrysoCollate: Why Was ChrysoCollate Developed?*

The process that led to the creation of ChrysoCollate was extensive and multifaceted. It stemmed from my combined experiences as a text editor and a professor of paleography and textual criticism for Arabic and Latin texts. For my first critical edition, during my PhD, I employed a collation system similar to the one I had been taught. This involved copying the entirety of a reference witness and indicating next to it, in lines or columns, the divergent readings of the other witnesses, all within a table akin to those produced in Excel from Microsoft or Calc from Apache OpenOffice or LibreOffice (Fig. 1). The goal of this approach was to collate the text quickly and legibly, without the need to transcribe all the witnesses entirely, by typing only the divergent readings. However, this method had significant drawbacks, one of which was very serious: having only the reference witness fully transcribed while the other witnesses were represented solely by columns or lines containing only the divergent variant readings meant that the editor was often influenced, even unconsciously, by the text collated first, whether positively or negatively. This cognitive bias could affect the objectivity of textual analysis. This observation was confirmed during my teaching, where I noticed its effects in many students, regardless of their skills and experience.

	A	B	C	D	E	F	G	H	I	J
1	<b>Celsi</b>	<b>O</b>	<b>H</b>	<b>L</b>	<b>C</b>	<b>F</b>	<b>V</b>	<b>D2</b>		
24622	solutio		---	---						Solutio
24623	autem		---	---		---		---		---
24624	plumbi		---	---						
24625	---		---	---		aut		uel		aut
24626	alterius		---	---						
24627	corporis,		---	---						
24628	non		---	---				nihil		
24629	est		---	---				aliud		
24630	aliud		---	---				est		
24631	nisi		---	---						
24632	diuide,	diuidere	---	---		diuidere	---	diuidere		diuidere
24633	&	---	---	---						---
24634	cetera	---	---	---						---
24635	corpora		---	---						
24636	quae		---	---		qui				qui
24637	se		---	---						
24638	coniunxerunt	coniunserunt	---	---						
24639	in		---	---						
24640	sublimatione,		---	---						sublimacione.

FIG. 1  
Example of a collation using a table in Microsoft Excel.

Later, I had the opportunity to learn computer programming, which inspired me to create a program that would automatically fill in the readings identical to the first collated witness and highlight similar and divergent readings through automatic coloring: seeing the texts of the witnesses aligned and transcribed in full, and having attention drawn to discrepancies and similarities thanks to the colors, would make it possible to bypass the main problem of the Excel collation table. This initial idea of an improved visualization of the collation table led to the development of ChrysoCollate. Indeed, I quickly saw how these few lines of code could be expanded and fix other important drawbacks of the Excel table. What started as a simple script transforming an Excel collation table into a local web page soon revealed a much broader potential: a fully fledged application capable of automating a wide range of tasks, significantly reducing the workload for researchers and minimizing copying errors.

Two years later, in early December 2021, the first version of ChrysoCollate was released online. Since then, the program has been continuously updated, with new features added on a regular basis, further enhancing its efficiency and usability. Currently, the program is maintained by me alone, and not by a team of IT specialists, but it has been part of and benefited from the research project *Incentive grant for scientific research* (MIS) of the FNRS '*Prisca alchimia*. The Beginnings of Arabic Alchemy in the East (eighth-tenth centuries) and its Transmission to al-Andalus and to the Latin World'.<sup>2</sup>

## 2. *What is ChrysoCollate?*

ChrysoCollate is an application software designed to assist with the collation, edition and production of critical editions in multiple formats, available for free and compatible with Windows, macOS and Linux. It features two modes: a collation mode and an edition mode.

The collation mode allows users to collate and compare the different witnesses of a text. In ChrysoCollate, aligned texts are

<sup>2</sup> F.4534.22, PI Sébastien Moureau, from 01/2022 to 12/2024.

segmented into units of variance or invariance. These units can be variant locations, that is, words or sequences of words where a variation is observed in the textual tradition, or invariant locations, where all witnesses agree on the same reading. For instance, given the text *arma uirumque cano Troiae* in A and *armis uirisque cano Troia* in B, we can divide it into three locations: a variant location *arma uirumque/armis uirisque*, an invariant location *cano*, and a variant location *Troiae/Troia*. The segmentation is managed by the editor, not by the program, as it can be performed in various ways according to different criteria, such as meaning, space saving, or others.<sup>3</sup> For each variant/invariant location, the program compares the content of the witnesses and assigns specific colors to the different readings, making it easy to quickly identify similar and divergent readings across the witnesses (Fig. 2).

The collation is semi-automatic rather than fully automatic, that is, ChrysoCollate accelerates the insertion of full transcriptions of witnesses by copying existing readings and adjusting them. Indeed, after extensive testing, I found that the fastest method for collating texts without an HTR program (on this, see below p. 56) involved a combination of manual collation and automated pre-filling of readings: once the first witness has been entered (by either typing it in ChrysoCollate or pasting it from another source), the program automatically suggests readings for subsequent witnesses at each variant/invariant location. In concrete terms, each time the user presses the tab key, ChrysoCollate pre-fills the reading for the variant/invariant location that follows the currently selected one by copying the reading from another witness that the user has chosen beforehand; then the copied reading can be modified if needed. This eliminates the need for repetitive typing and allows for easy modification if a reading differs.<sup>4</sup>

<sup>3</sup> On the concept of variant location, see Andrews 2020.

<sup>4</sup> The order of the witnesses can be modified at any time, so that one can choose which witness appears first and how the witnesses sequence, which facilitates the visualization of the textual tradition and the classification of the witnesses.

Variant/invariant locations can be merged, split, added or deleted as needed. The user can also adjust the program's sensitivity to specific types of variations (for example, choosing whether or not to distinguish between *a* and *e* for medieval Latin, or take the *hamza* into account in Arabic).<sup>5</sup> Additionally, editors can insert personal working notes, or comments that will appear in the apparatus criticus (such as *supra lineam*, *in margine*, etc.). A folio or page management system, along with an image viewer, allows users to easily navigate through the witnesses within the ChrysoCollate file. Users can mark the beginnings of folios or pages and link image files or webpages to them. As they navigate through the text, the program displays the image or webpage corresponding to the text currently appearing on the screen, which simplifies the collation and the back-and-forth between the collation table and the witnesses.

The edition mode enables users to create a critical edition based on the collation table. With simple clicks, users can select the readings, and the program automatically generates the edited text along with the notes for the apparatus (Fig. 3). If needed, the editors can add their own readings as well, and comment on them (see Fig. 3, variant location 62). Apparatus can be positive or negative, and the user can switch them from one to the other at any time, ChrysoCollate updating them automatically.<sup>6</sup> Users can have more than one apparatus, in order to isolate certain kinds of variant readings (orthographic apparatus, etc.). The expressions used in the apparatus language, such as those marking omissions or additions, can be modified so that any language can be used, the default one being Latin. A variety of functions allows editors to tailor the apparatus entries according to their needs:<sup>7</sup>

<sup>5</sup> Script- and language-based features are added as users request them. For the time being, they concern Latin, Spanish, Greek and Arabic, but these kinds of options can be extended to all languages and all scripts.

<sup>6</sup> A positive critical apparatus is an apparatus in which all witnesses are mentioned in each entry, while in a negative apparatus, only witnesses containing a divergent reading are mentioned. Example of an entry in a positive apparatus: 'arma uirumque A B C : armis uirisque D E'. Example of the same entry in a negative apparatus: 'arma uirumque] armis uirisque D E'.

<sup>7</sup> These modifications do not stop the automation process from running, and the entries are updated automatically every time a reading is modified, unlike full customization of entries, which is also possible, see below, p. 60.

**ChrysoCollate 1.3.1** Dummy file - Virgil, Aeneid... 63 VG Collation Edition

Save  Load  Import  Transl.  Viewer  Help  +C  +N  Add wit.  Wit. settings  Add text  +L/A  +F  Export wit.  Delete v.l.  Merge v.l.

Nb.	←548→	←558→	←568→	←578→	←588→	←598→	←608→	←618→	←628→	←638→	←648→	←658→	←668→	←678→	←688→	←698→	←708→	←718→
A	tantaene	animis caelestibus irae	civitas	antiqua fuit	Syrii	tenuere coloni	carthago	Italiam contra	Tiberinaque	longinque	ostia dives	opera	belli	qua				
B	Tantaene	animis caelestibus irae	Urbs	antiqua fuit	Tyrii	tenuere coloni	karthago	Italiam contra	Tiberinaque	longe	ostia dives	opum	bellorum	qua				
C	num tantae	animis caelestibus irae	urbs	antiqua fuit	Sirii	tenuere coloni	karthago	Italiam contra	Tiberinaque	longe	ostia dives	opum	studisque	asperima	bellum	qua		
D	Tantaene	animis caelestibus irae	Urbs	antiqua fuit	tirii	tenuere coloni	yarthago	Italiam contra	Tiberinoque	longe	ostia dives	opera	studisque	asperima	belli	qua		
E	Tantaene	animis caelestis	irae	antiqua fuit	tirii	tenuere coloni	yarthago	Italiam contra	Tiberinoque	longe	ostia dives	opera	studisque	asperima	belli	qua		

FIG. 2  
Collation mode (fictitious file).



witnesses can be discarded for certain variant/invariant locations if needed, readings of various witnesses can be assimilated if necessary for the readability of the apparatus,<sup>8</sup> and one can switch certain variant/invariant locations of a negative apparatus to positive apparatus notes. An automated pre-editing function is available, which can automatically pre-select readings in certain specific cases (see below p. 62). There is also a lacuna management system, which automatically inserts the beginning and end of lacunae in the critical apparatus and removes lacunar witnesses from the passages they do not contain.

Several export options are available: ODT files (OpenOffice, LibreOffice, Microsoft Word), CTE files (Classical Text Editor), TEX files (XeLaTeX `reledmac`). Additional formats, such as LaTeX `ekdosis` and XML/TEI, will be available soon. When using a negative critical apparatus, a *tradio textus* can be automatically generated in Classical Text Editor files.<sup>9</sup> ChrysoCollate also includes a basic statistical tool for comparing the variant readings and a *stemma codicum* checker to identify inconsistencies, based on the predefined rules outlined in the program's help menu.

Recently, a variant graph visualization feature has been added, in both collation and edition mode (Fig. 4). The variant graph allows the text to be visualized in the form of a path, each step of which is the juxtaposition of the different readings of a variant/invariant location, these readings being linked by segments labelled with the names of the witnesses containing the readings they link.<sup>10</sup>

<sup>8</sup> For instance, in Arabic, editors might want that variations of diacritical dots do not always appear in the apparatus, such as `تععمل` and `يععمل`; in Latin, a user might want to assimilate *Laviniaque* and *Lavinia que* to avoid overloading the apparatus.

<sup>9</sup> A *tradio textus* is an apparatus displaying the available witnesses for each page of the edition, usually including beginnings and endings of witnesses, as well as lacunae.

<sup>10</sup> On variant graphs, see Andrews 2023b and van Zundert 2020, pp. 193–201. Other links are often used in variant graphs: links between the readings of a same variant location which, in this case, indicate the type of variation. In addition to what is said in these papers, I add that, in my personal experience, the variant graph is particularly useful when correcting editions made by other people.

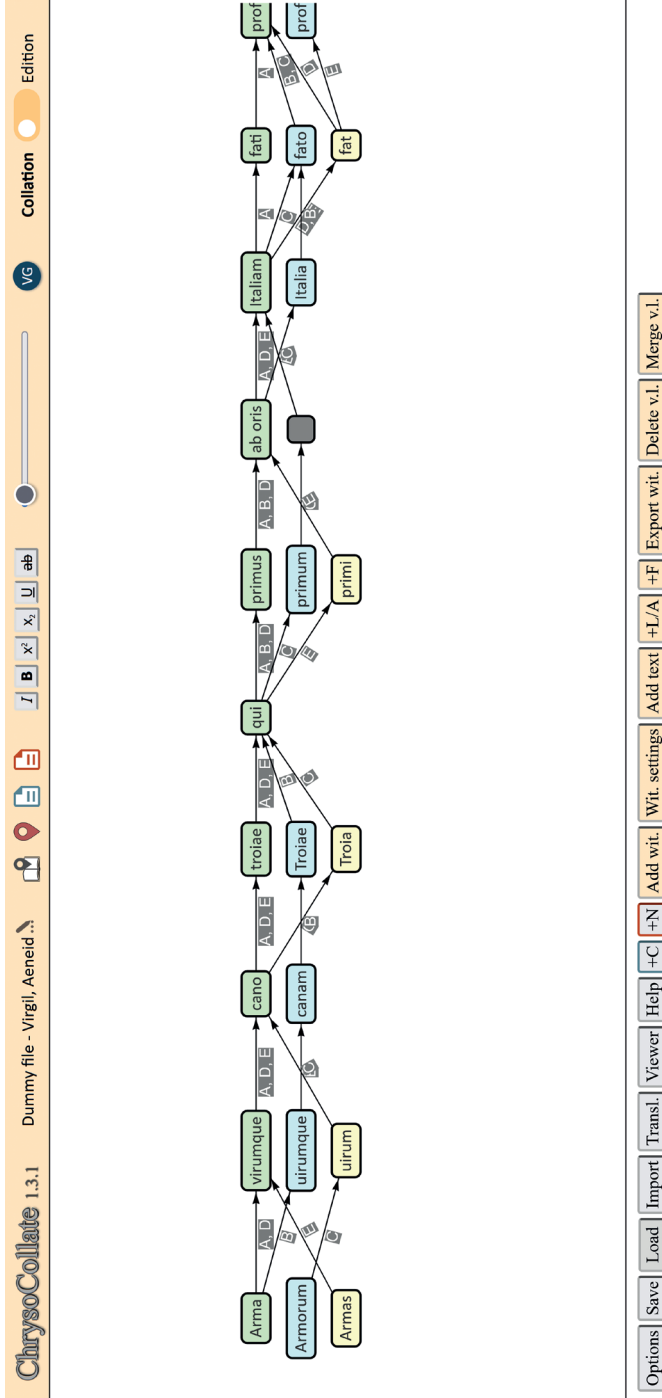


FIG. 4  
Visualization of a collated text with a variant graph (fictitious file).

In addition to the collation table and the critical edition, ChrysoCollate includes a tool for writing and synchronizing the translation of the text. This feature allows users to align the translation with the text through synchronization points manually inserted in the translation, and then to automatically export them (aligned) in various formats.

Technically, ChrysoCollate is programmed as a local web application, without a server, in HTML and JavaScript. It opens in any recent web browser and does not need access to the internet in order to run, being a local page. Variant graphs and stemmata codicum are created using the vis.js JavaScript library. Compression in certain processes is carried out using the JSZip JavaScript library.

### 3. *What is ChrysoCollate Not?*

ChrysoCollate is not designed for typesetting critical editions. It produces files in formats that are compatible with existing layout programs: ODT, CTE, TEX (and soon XML/TEI). The aim of ChrysoCollate is to assist editors with the steps that precede the layout process, not to reinvent the wheel. Many excellent programs are available for typesetting. Basic layouts for critical editions can be made with Microsoft Word, Apache OpenOffice, LibreOffice, or other well-known programs (using the ODT format). The CTE format is to be used with Classical Text Editor, a program developed by Stefan Hagel specifically for formatting and typesetting critical editions. It remains one of the most accessible tools for this purpose, offering a user-friendly, ‘what you see is what you get’ (WYSIWYG) experience. However, it is becoming quite outdated, suffering from screen lag, having Unicode compatibility issues with recent Windows versions and problems when formatting right-to-left languages, and being available only on PC.<sup>11</sup>

ChrysoCollate is not a handwritten text recognition (HTR) program. HTR programs are tools designed to transcribe images of handwritten documents. Current HTR software, which relies

<sup>11</sup> <https://cte.oeaw.ac.at/> (accessed: February 2025).

on complex machine learning techniques, such as Transkribus, eScriptorium or Calfa, can generate transcriptions of manuscripts but do not assist with collation itself.<sup>12</sup> However, the output of these programs can be used with tools for automatic collation.

ChrysoCollate is neither a computer metalanguage nor a markup format. While the software processes information in the form of a structured database using its own JSON format, this format is intended for machine readability (although not unreadable to human eye). The program acts as an interface between the structured data and the user. That said, ChrysoCollate allows users to export data in a variety of widely used formats.

#### 4. *What Other Tools Can Serve the Same Purpose?*

There is currently no other tool that offers the same features as ChrysoCollate. While some software and platforms provide functionalities that overlap with some of ChrysoCollate, none replicate them entirely.

CollateX is a text alignment program that can automatically collate witnesses of a text. However, a major drawback is that each witness must be fully transcribed, which can significantly slow down the collation process unless done automatically using an HTR program with a high accuracy.<sup>13</sup> CollateX operates via a command prompt, requiring some technical knowledge. It is now possible to import CollateX output files (JSON format) into ChrysoCollate, integrating its alignment capabilities into a user-friendly workflow.

Two tools that specialize in textual alignment and visualizing similarities and differences between witnesses are available: Variance Viewer and LERA.<sup>14</sup> While these tools share with Chryso-

<sup>12</sup> Respectively <https://www.transkribus.org/>, <https://escriptorium.inria.fr/> and <https://calfa.fr/> (accessed: February 2025).

<sup>13</sup> <https://collatex.net/> (accessed: February 2025). On automatic collation, see Andrews 2020. On CollateX, see also Haentjens Dekker et al. 2015.

<sup>14</sup> <https://github.com/cs6-uniwue/Variance-Viewer>, <https://lera.uzi.uni-halle.de/> (accessed: February 2025). Juxta was providing similar features but is no longer available. On Juxta, Variance Viewer and LERA, see also Roeder 2019. On the visualization of collated texts, see also van Zundert 2020.

Collate the idea of helping users to visualize the data, they do not support the creation of critical editions, nor help with the entire collation process.

Another common approach to encoding critical editions is to work directly in specialized markup languages without relying on a facilitator tool like ChrysoCollate. The most widely used formats are XML/TEI and LaTeX (with the `reledmac` or `ekdosis` packages).<sup>15</sup> However, encoding texts in these formats requires prior expertise and does not provide an intuitive visualization unless additional software is used or data is exported to other tools.

Classical Text Editor is another option, but since it is primarily designed for formatting traditional printed editions, it is poorly suited for managing collation.<sup>16</sup> Using a prior collation table is essential when working with this program.

There are also visualization programs that display pre-encoded data but do not assist with the initial encoding process. For example, EVT 1 and 2 (Edition Visualization Technology) allow users to visualize XML files of digital editions.<sup>17</sup> Alternatively, files exported as PDFs from `reledmac`, `ekdosis`, or Classical Text Editor provide a visualization similar to a printed edition, with the edited text and the critical apparatus.

Finally, some digital work environments enable text editing, such as Ediarum or Cadmus. However, these platforms require IT specialists to set them up and are primarily intended for large, well-funded research projects.<sup>18</sup>

### 5. *What Are the Advantages of ChrysoCollate?*

ChrysoCollate offers several advantages over traditional collation systems consisting of manually filled collation tables (paper, XLS, DOCX, or other formats), which only display variant readings

<sup>15</sup> Respectively <https://tei-c.org/>, <https://ctan.org/pkg/reledmac> and <http://www.ekdosis.org/fr/index.html> (Accessed: February 2025).

<sup>16</sup> <https://cte.oeaw.ac.at/> (accessed: February 2025).

<sup>17</sup> <http://evt.labcd.unipi.it/> (accessed: February 2025).

<sup>18</sup> Never having used these two tools, I cannot speak about them in detail: <https://www.ediarum.org/> and <http://fusisoft.it/apps/cadmus-show/#/intro>.

diverging from the first collated witness. It also surpasses editions written entirely in word-processing programs, which require constant updates and manual formatting.

The primary advantage of ChrysoCollate is speed. By automating key steps in the editing process, it allows editors to focus on collation and edition rather than the practical constraints of writing and updating the critical apparatus. Another major advantage is data visualization. As explained above, and in addition to variant graphs, ChrysoCollate displays all witnesses in full, highlights both similarities and differences with colors, and enables users to adjust the sensitivity of textual comparison of the program. This facilitates the visualization of the links between the witnesses, improves their comparative analysis and classification, and frees the editor from the influence of the first collated witness. Additionally, the automatic generation of the critical apparatus ensures formatting consistency, prevents transcription errors when copying readings from the collation table, and reduces attribution of readings to the wrong witnesses. Since ChrysoCollate stores collation tables and editions as structured databases, it is relatively easy to develop new export formats or integrate additional tools (e.g. for phylogenetic analysis or other textual studies). Finally, ChrysoCollate is free, regularly updated, and continuously improved on the basis of user feedback or personal ideas. Bugs are corrected, and new features are frequently added, ensuring that the software evolves to meet the needs of the academic community.

### *6. What Are the Disadvantages of ChrysoCollate?*

Like any tool, ChrysoCollate has its limitations. Some of these are technical constraints that can be addressed, while others are inherent to its design and more difficult to overcome.

ChrysoCollate operates as a local web application (without a server). Essentially, it functions like a web page stored on the hard drive, which must be launched via a browser. While this approach offers many advantages, such as cross-platform compatibility and access to various web libraries, it also imposes technical constraints. One notable limitation is that ChrysoCollate struggles with very large numbers of witnesses (over 100), par-

ticularly for long texts. Performance can degrade as the dataset grows. To mitigate this, an option is available to reduce the number of displayed variant/invariant locations, but this does not always entirely solve the issue.

Additionally, like traditional collation tables, ChrysoCollate has difficulty managing transpositions (i.e. changes in word or phrase order). A satisfactory system for handling this type of variation is yet to be developed in the program.

Finally, automation comes at the cost of reduced flexibility. Since ChrysoCollate automatically structures apparatus entries, they follow a predefined format, which may not always align with the editor's preferences. Users can manually fully customize some apparatus entries, but these entries can no longer be updated automatically once modified, requiring manual adjustments if changes are needed later.<sup>19</sup> Additionally, improper use of automation can lead to problematic ways of working (see below p. 63).

### *7. What Are the Future Developments for ChrysoCollate?*

Several enhancements are planned for ChrysoCollate. Beyond expanding the range of export formats (including LaTeX ekdosis and XML/TEI), new script- and language-based features for the comparison of readings are planned. These developments will be guided by feedback from specialists in various languages, such as Armenian, Georgian, and others. Additionally, new tools for analyzing textual traditions are under consideration, along with expanded export functionalities, particularly for phylogenetic analysis. These improvements aim to enhance both the analytical capabilities and interoperability of ChrysoCollate within broader scholarly workflows.

### *8. How Might ChrysoCollate Change the Way Editors Work?*

After this brief presentation of ChrysoCollate, it is worth considering the potential impact of this kind of tool on editorial

<sup>19</sup> This customization is different from the tailoring of entries mentioned above (p. 51). Rather, this refers to a full modification of the entry through direct text input.

practices. ChrysoCollate is fundamentally a tool. It does not advocate for any particular methodology and can be used within various editorial approaches, whether neo-Lachmannian, Bédierist, or others.<sup>20</sup> However, as with any tool, its use inevitably influences workflows, sometimes subtly modifying established habits. That said, we should not overestimate its impact. ChrysoCollate was not designed to revolutionize textual editing, but rather to update and improve the application of long-proven collation methods using current computing capabilities. *Nihil novi sub sole*. Nothing fundamentally new is being introduced; the software builds on techniques that have been successfully employed and refined for over a century, adapting them to a digital environment.<sup>21</sup> Additionally, it is difficult to predict its medium- and long-term effects on editorial practices.

One of the first things I noticed, and which I did not necessarily anticipate, is that ChrysoCollate has a pedagogical role, particularly for students and early-career researchers. Several users unfamiliar with collation tables have expressed surprise at how much simpler and more objective their work becomes with ChrysoCollate. Tasks that were once tedious (noting variants in footnotes and constantly modifying editions in word processors, or writing apparatus entries from collation tables) are now streamlined and more precise. Furthermore, ChrysoCollate is a valuable teaching tool in textual criticism courses. By visually highlighting textual variations, it helps students grasp the importance of collation and edition work.

ChrysoCollate's visualization features, i.e. tables with readings highlighted by means of colors and variant graphs, are not new in themselves. Previously, some editors used Excel or Calc spreadsheets with manual coloring, but this approach was time-consuming and impractical for full texts. ChrysoCollate auto-

<sup>20</sup> The neo-Lachmannian method is Karl Lachmann's editorial method (i.e. reconstructing a text basing on the genealogical kinship of witnesses) as it was refined and theorized by later scholars (Gaston Paris, Paul Maas, etc.). It is also called the genealogical method. On this, see Trovato 2014. Joseph Bédier, a pupil of Gaston Paris, criticized the Lachmannian method and proposed to edit a 'bon manuscrit' instead of producing a reconstructed text. On this, see Trovato 2013.

<sup>21</sup> On this reflection, see also Bordalejo 2018.

mates this process, the color display is very intuitive and quickly draws the editor's attention to variations (see, for example, Fig. 2, the contrast between variant/invariant locations 59 and 60 or 62, or that between variant/invariant locations 63, 64 and 65). The variant graph, now a well-known tool in textual studies, offers a different perspective on textual tradition. Instead of focusing the analysis on witnesses, it concentrates on textual variations. This visualization system has its own advantages and limitations, which have been discussed in the scientific literature.<sup>22</sup>

One of ChrysoCollate's most transformative features is automation, which can cause several profound changes in the editorial workflow.<sup>23</sup> Its most visible impact is the significant reduction in time required for the collation and the production of the edition and the apparatus. As a personal experience, the entire process, from collation to final export, is completed three times faster than with traditional methods, though this varies depending on the type of text, the edition method, and the editor's working habits. This automation has a direct impact on research projects, allowing editors to focus more on textual tradition and editorial decisions rather than spending excessive time on repetitive technical tasks.

The operations that are automated in ChrysoCollate are precisely those that, while necessary, are not at the heart of the editor's scientific work. However, certain aspects of editorial decision-making could, in theory, be automated further, particularly the selection of readings. By integrating deep learning tools trained on specific textual corpora, ChrysoCollate could theoretically assist in proposing readings to the editor. However, deep learning tools are not yet sufficiently reliable for such work; currently, they are better suited for generating new texts than for reconstructing existing ones based on preserved witnesses. For these reasons, I have limited automation to a pre-editing function that does not rely on deep learning tools. This function offers three different modes: 1) automatic selection of unique readings in in-

<sup>22</sup> See note 10.

<sup>23</sup> The literature on tools that automate tasks for critical editions is quite abundant, as are the tools. See for instance Andrews 2023a.

variant location (i.e. readings common to all witnesses); 2) automatic selection of readings from the base witness for all variant and invariant locations (which naturally requires thorough review), which can be useful when using a method that favors one witness over all others; and 3) a more complex hybrid system (that also needs a thorough review), which selects the common readings in invariant locations, the majority reading when all but the base witness agree, and defaults to the base witness in all other cases. These rules were designed for specific cases requested by users, and I have included a clear warning in the program to ensure they are applied with discernment.

Another crucial consequence of automation is the reduction, sometimes drastic (especially on days of fatigue), of copying errors. The fact that selected readings and, most importantly, critical apparatus entries are automatically written prevents any copying or misattribution errors at these stages. The only remaining potential for copying errors lies in reading and encoding the readings of the witnesses during the collation process. My experience, both in my own editions and those of my doctoral students, has shown this to be a major benefit. Additionally, consistency in the apparatus entries is ensured.

However, it is worth pointing out that the automation of some of the editor's tasks is, like any automation, double-edged. Over time, I have observed that ChrysoCollate, while addressing many of the weaknesses of manual collation, can also introduce new challenges if not used properly; ChrysoCollate can sometimes suffer from the defects of its own strengths. I have identified four difficulties that I have seen users encounter, all stemming from imperfect use of ChrysoCollate. 1) The most common issue is that automation can sometimes cause the editor to filter variations incorrectly. Since ChrysoCollate makes it extremely easy to generate exhaustive apparatus entries, some editors end up including too much information, drowning essential details in an excess of minutiae. With the understanding that a critical edition should primarily remain usable for readers, excessive precision, while often stemming from a wishful desire for scientific accuracy and fidelity to the textual tradition, can ultimately be detrimental to an edition. But this bias is not unique to ChrysoCollate. 2) Another frequent issue I have noticed, particularly among

students, is a tendency to segment variant/invariant locations too finely. Traditionally, editors tend to divide variant/invariant locations into meaningful units whenever possible and justified, a practice that comes naturally when manually writing apparatus entries. Since ChrysoCollate automates the production of these entries, some users overlook this principle and create apparatus notes that are overly fragmented, making them difficult to read.<sup>24</sup> 3) Moreover, while ChrysoCollate has clear pedagogical benefits, particularly in its visualization aspect, it also carries the risk of diminishing students' mastery of apparatus structure. Because entries are generated automatically, students who do not take the time to study them carefully may fail to develop an intuitive understanding of how they are structured. It is therefore necessary to stress this point to students and young researchers who use the program. This phenomenon is common with automation: when a process is handled automatically, it is no longer learned in the same way, or sometimes not learned at all. 4) Finally, a more subtle issue, though rarer, is the tendency of some users to work too quickly. The speed and ease of editing a variant location with a simple click can lead to a routine where editors fail to give sufficient attention to certain variant locations. However, it must be said that this is not a new problem, as there have always been people working too fast.

The question of data security and long-term software sustainability is also worth addressing. Any digital tool carries a risk of data loss, whether due to accidental overwriting, software bugs, or user errors. To mitigate this, ChrysoCollate incorporates a backup system of versioned files in which every saved project generates a new file with a timestamp. This ensures that earlier versions are always available, though it may require users to occasionally clean up the numerous backups accumulated over time. Since the software's release, only one instance of data loss has been reported (one day of collation corrections), caused by

<sup>24</sup> For example, if we imagine the following text in witnesses A, B and C 'arma uirumque A, arma uirosque B, arena uirumque C', some students will produce the entry 'arma A B: arena C; uirumque A C: uirosque B' instead of 'arma uirumque A: arma uirosque B; arena uirumque C' which is better for readability. These short cuts become very problematic when there are many witnesses.

a browser issue rather than the program itself. The durability of the data is guaranteed by the format of the ChrysoCollate saved files: they are simple JSON objects stored on the user's hard drive in the form of text files (Unicode), without compression or encryption. These files can be opened with any text editor. Data security is identical to that of any file on the user's computer. No collation or edition data is transmitted over the Internet, everything is local. No usage data is collected.<sup>25</sup> In terms of long-term viability, ChrysoCollate is developed in HTML and JavaScript, ensuring compatibility with future web technologies. While occasional updates may be needed to maintain browser compatibility, major obsolescence issues are unlikely. ChrysoCollate is free, but not open source for the time being. However, this option is by no means out of the question in the long term.<sup>26</sup>

### *Conclusion*

ChrysoCollate, like many applications developed to assist editors in their work, is primarily a tool. It is not intended to reform the practices of the discipline but rather to help editors in their tasks by increasing speed, reducing copying errors, and facilitating data visualization. Nevertheless, it should be acknowledged that any new tool can influence working habits and thereby present potential challenges. Consequently, it is important for editors to remain mindful of time-tested and established philological practices when using these new tools. And conversely, editors should not succumb to the temptation of rejecting new instruments for fear of losing long-standing practices. The key to successfully adopting a new tool lies in its proper use.

<sup>25</sup> Only the surname, first name, affiliation and email address are required to download the program. They are collected when the user registers on the ChrysoCollate site and are not linked to the instance of ChrysoCollate that is downloaded. No data is transmitted to the server while the program is installed or run. The registration data is used solely to send emails about updates. It is possible to request its deletion by sending an email to [chrysocollate@uclouvain.be](mailto:chrysocollate@uclouvain.be).

<sup>26</sup> I have also made arrangements in the event of my death. The code would then be placed as open source on the ChrysoCollate website.

*Abstract*

Collation, critical edition, and the visualization of textual traditions are intricate and time-intensive scholarly processes. In recent decades, various digital tools have been developed to aid researchers in these complex tasks—among them, ChrysoCollate. This paper introduces ChrysoCollate, a software application specifically designed to support the collation, critical editing, and visualization of textual traditions. It provides a comprehensive overview of the tool’s features and functionalities, and examines its relevance and impact on contemporary scholarly practices in the fields of textual criticism and critical editing.

*Bibliography*

- Andrews, T. (2020) ‘Transcription and Collation’, in P. Roelli (ed.) *Handbook of Stemmatology*. Berlin: De Gruyter, pp. 160–75.
- Andrews, T. (2023a) ‘Where are the Tools? The Landscape of Semi-Automated Text Edition’, in R. Bleier, H. W. Klug (eds) *Digitale Edition in Österreich. Digital Scholarly Edition in Austria*. Norderstedt: Books on Demand (Schriften des Instituts für Dokumentologie und Editorik, 16), pp. 3–18.
- Andrews, T. (2023b) ‘Algorithms for the Manipulation and Transformation of Text Variant Graphs’, in A. ŠeĽa, F. Jannidis, I. Romanowska (eds) *Proceedings of the Computational Humanities Research Conference 2023*. CEUR Workshop Proceedings, pp. 577–91. Available at: <https://ceur-ws.org/Vol-3558/paper6515.pdf> (accessed: May 2025).
- Bordalejo, B. (2018) ‘Digital versus Analogue Textual Scholarship or The Revolution is Just in the Title’, *Digital Philology: A Journal of Medieval Cultures*, 7(1), pp. 7–28.
- Haentjens Dekker, R., van Hulle, D., Middell, G., Neyt, V. and van Zundert, J. (2015) ‘Computer-supported Collation of Modern Manuscripts: CollateX and the Beckett Digital Manuscript Project’, *Digital Scholarship in the Humanities*, 30(3), pp. 452–70.
- Roeder, T. (2019) ‘Juxta Web Service, LERA, and Variance Viewer. Web based collation tools for TEI’, *RIDE – a review journal for digital editions and resources*, 11. DOI: 10.18716/ride.a.11.5.
- Trovato, P. (2013) ‘La tradizione manoscritta del *Lai de l’ombre*. Riflessioni sulle tecniche d’edizione primonovecentesche’, *Romania*, 131, pp. 338–80.

- Trovato, P. (2014) *Everything you Always Wanted to Know about Lachmann's Method. A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Cladistics, and Copy-Text*. Limena: Libreriauniversitaria.it edizioni (Storie e linguaggi).
- van Zundert, J. (2020) 'Data Representation', in P. Roelli (ed.) *Handbook of Stemmatology*. Berlin: De Gruyter, pp. 175–207.