Visualizing *Mouvance*: Toward a visual analysis of variant medieval text traditions

Stefan Jänicke
Institut für Informatik, Leipzig University, Germany

David Joseph Wrisley
Digital Humanities, New York University Abu Dhabi, UAE

Abstract

Medieval literary traditions provide a particularly challenging test case for textual alignment and the visualization of variance. Whereas the editors of medieval traditions working with the printed page struggle to illustrate the complex phenomena of textual instability, research in screen-based visualization has made significant progress, allowing for complex textual situations to be captured at the micro- and the macro-level. This article uses visualization and a computational approach to identifying variance to allow the analysis of different medieval poetic works using the transcriptions of how they are found in particular manuscripts. It introduces the notion of a meso-level visualization, a visual representation of aligned text providing for comparative reading on the screen, all the while assembling non-contradictory, intuitive solutions for the visual exploration of multi-scalar variance. Building upon the literary notion of *mouvance*, it delves into medieval French literature and, in particular, different visualizations of three versions of the *Chanson de Roland* (the Oxford, the Châteauroux, and the Venice 4 manuscripts). The article presents experimental prototypes for such meso-level visualization and explores how they can advance our understanding of formulaically rich medieval poetry.

Correspondence:
David Joseph Wrisley, New York University Abu Dhabi
PO Box 129188
Saadiyat Island
Abu Dhabi, United Arab Emirates. **E-mail:**
djw12@nyu.edu

1 Introduction

Poetry in the Middle Ages changed as it was recited and passed along, but also as it was written down and recopied. *Mouvance* is a term used by the late Swiss medievalist Paul Zumthor to designate the high degree of instability in medieval text traditions. Zumthor qualifies this instability as an 'interplay between variant readings and reworkings', balancing both the textual, literary elements of written works with oral, performative ones (Zumthor, 1992; p. 44). Cerquiglini has argued that, faced with such variance, the medievalist's 'analysis must be comparative, not archeological' (Cerquiglini, 1999; p. 44). Synoptic editions provided readers of the printed page with a visual frame for close, comparative reading of variant texts, inviting exploration and giving insight into processes of textual change. These page layouts provided, however, far less than the panoptic view required to grasp the larger phenomenon at hand. This article leverages the conceptual progress of visualization in facilitating hybrid, on-screen, multi-scalar reading of computationally aligned text and begins to explore what it can tell us about orally inflected poetry of the middle ages. Such computational alignment reveals what is

Digital Scholarship in the Humanities, Vol. 32, Supplement 2, 2017.

doi:10.1093/llc/fqx033 Advance Access published on 13 September 2017

both stable and variant about texts, and when quantified, reuse of textual strings also sheds new light on the use of formulaic language. Visualizing *mouvance* in a computational environment, therefore, balances the discovery of both instability and repetition in poetic language.

2 The Complexity of Medieval Textual Data

The medievalist in the age of digital archives does not only read critical editions but also compares different individual manuscript versions, on site or on screen, as access permits, using both human memory and deep content expertise to arrive at a larger picture of the life of texts. Nichols argues that what in textual criticism has always been labeled the variants, far from being rejected to the footnotes, should not stand as 'de-textualized tokens', but with Cerquiglini, he pleads for access to the 'dynamic life of medieval works' (Nichols, 2016; pp. 108, 110). The codicological turn, as it might be called, of the 1990s, argued for refocusing critical energy on situating the manuscript within its social and cultural context of textual production and transmission. This shift toward the material also corresponded to the general rise of the availability of digitized versions of manuscripts. The synoptic close reading of manuscripts indeed expands the frame for interpretation; yet the function of the digital for Nichols serves mainly to provide a human reader access to a wider variety of manuscript versions. In the absence of computationally tractable text, however, such a critical approach is restricted to the 'single reader's moving memory' (Bowers, 2016).

The use of computational environments to understand medieval textual complexity goes back some 50 years; yet such approaches are not devoid of assumptions taken from the field of textual criticism. At the risk of oversimplification, we can say that textual critics of different languages and periods have not been of the same opinion about what to do with multiple extant archival documents containing versions of texts that resemble each other to a lesser or greater extent. Some text traditions are complex because of the number of variants, that is the ways that documents diverge from one another, and divergence in words has been used to deduce and even visualize the hypothetical relationship between extant versions and to make heuristic claims about a more 'authentic' common ancestor text. This approach to textual criticism was developed by the students of Karl Lachmann. Robinson's two articles on the 'machine collation of forty-four manuscripts of two Old Norse poems' with the goal of automatically generating variant tables against a 'notional master' reproduce somewhat the Lachmannian paradigm (Robinson, 1989). The contemporary Web-based software, Juxta Commons, allows one 'to compare and collate versions of the same textual work' echoing similar assumption about texts (Wheeles and Jensen, 2013). An opposition to stemmatics in textual criticism arose in the field of medieval French studies with Joseph Bédier who opted for using the best copy of a text as a base text for editing, regardless of its age or geographic provenance, and developed a set of attendant critical procedures for doing so. Pierazzo has argued that much has changed in the world of scholarly editing with the rise of digital critical editions. One is no longer influenced by the economics of print-based publishing, and the question in newer forms of Web-based editing is not which text to publish, but rather where to stop in the representation of textual information (Pierazzo, 2011). In this article we take a look at medieval vernacular text traditions, without attempting to edit them, but rather to compare their variance using innovative visualizations. We explore the high degree of variance of these text traditions that, instead of being seen as a fault or contamination, might just be 'elevated into an organizing principle' (Richards, 1987).

Digital medievalists interested in computer-assisted analysis of the complex phenomenon of *mouvance* need to take all the known versions of the textual record into consideration. In the case of medieval French, we have only a few fully transcribed textual traditions full. Archives of digitized manuscripts, such as the *Roman de la Rose Digital Library* spearheaded by Nichols, exist and they provide important access to manuscript traditions, but they have been relatively unsuccessful in creating transcriptions of medieval manuscripts (Roman de

la Rose, 2017). At the time of writing this article, fewer than five of such transcriptions of the *Rose* exist. An example of a completed early digital textual archive in medieval French is the Princeton Charrette Project that produced transcriptions of the eight known manuscripts of Chrétien de Troyes' *Chevalier de la Charrette* (c. 1180) accompanying a critical edition (Uitti *et al.*, 2006). Chrétien de Troyes' romance was popular, as the number of manuscript copies attest, but the variants in it are relatively limited compared with those mentioned by Richards above, not a true test case for Zumthor's performance-rich notion of *mouvance*. Suffice it to say that not all genres nor all texts exhibit this kind of textual and performative instability.

3 Visualizing Variant Text Traditions

In this article we focus on two famous text traditions in medieval French literature that exhibit a significant degree of textual and performative *mouvance*: first, the *fabliaux*, a genre of bawdy, short verse tales, and second, the epic tradition known as the *chansons de geste*. First, there is the thirteenth-century *Du chevalier qui fit les cons parler*, extant in seven manuscripts, six in continental French and one in the French of England (Brun, 2016a). Second, there is the *Chanson de Roland*, known to be transmitted in six major versions, the first from the early twelfth century (Oxford, Venice 4, Châteauroux-Venise 7, Paris, Cambridge, and Lyons) as well as in three fragments (Lavergne, Bogdanow, Michelant) (Brun, 2016b). It is possible for us to visualize these text traditions in comparative views precisely because their individual manuscripts have been fully transcribed in modern editions (Duggan, 2005; Moffat, 2014).

In both cases, the difference between the versions—in length, content, and structure—is significant enough that text editors have chosen to publish different versions of the text together. The four manuscript versions of the octosyllabic *fabliau* in question, edited by Rychner in a synoptic edition, vary notably in the number of lines: 602 (ms. E), 609 (ms. C), 615 (ms. B), 750 (ms. I) and 292 (ms. M) (Rychner, 1960). An edition of the *Roland* presents each of the single manuscripts versions edited serially in a multivolume publication (except the Châteauroux-Venise 7 group represented by two manuscripts). The six major versions of *Roland* vary dramatically in the number of lines they contain: 4,002 (Oxford), 6,002 (Venice 4), 8,395/8,201 (Châteauroux-Venise 7), 6,828 (Paris), 5,695 (Cambridge), and 2,392 (Lyon). Whereas the different manuscripts in which we find this *fabliau* present a significant amount of rearrangement and interpolation, the *Roland* tradition is much more complicated, as it exhibits more variance than stability, so much so that that the various versions of the epic legend commonly called the *Chanson de Roland* by literary scholars challenge the very distinction of text and document articulated by traditional textual criticism. In the case of the *Chanson de Roland*, one might go so far as to call them the *Songs of Roland*. To the problem of variable length, one must add the orthographic instability of medieval French as an obstacle to an approach combining computational alignment and text reuse. Our main question is how to combine automatic alignment and visual analysis to optimize the exploration of *mouvance*.

4 From Variants to Variance

Existing methods for text alignment in digital environments, generally speaking, favor relatively stable texts with a small number of variant readings. The Versioning Machine accepts texts encoded 'according to TEI's Parallel Segmentation method' and 'interprets the encoding, parsing out the text into its constituent parts' (Versioning Machine, 2015). The authors of the Versioning Machine offer a sample alignment of a middle English 'Prophecy of Merlin' (Fig. 1). The verse-to-verse alignment has been manually encoded by the textual scholar. Similar lines are visually connected using customary mouse behavior; however, variance within the line or across lines cannot be visualized.

Another environment for the automatic collation and visualization of textual difference from plain

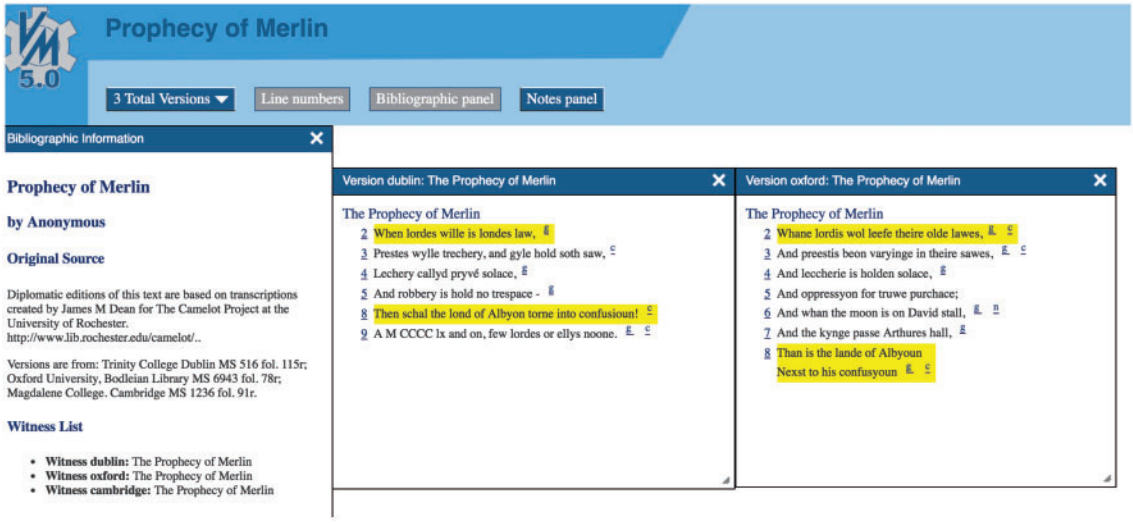

Fig. 1 Two versions of the middle English Prophecy of Merlin visualized with Versioning Machine (reproduced with permission)

text files is Juxta Commons (Wheeles and Jensen, 2013). When complex text traditions containing more than just variant readings, but also interpolations, half-line reuse, or a significant amount of orthographic variance are collated automatically in Juxta Commons, their results are unfortunately nearly illegible.

Figure 2 shows two versions of the aforementioned *fabliau* visualized in Juxta Commons; the visual alignment achieved by line segments in the central column, however, does an insufficient job at expressing the complex instability of the medieval text tradition. The same concept of an 'alignment column' sitting between two text columns is deployed by the TransVis prototype for the alignment of Shakespearean translations in German (TransVis, 2016). The structure of theatrical texts, essentially individual character monologues, renders the alignment significantly easier. Medieval *mouvance*, on the other hand by contrast, is not restricted to discrete blocks of text.

More dynamic visualization strategies for historical text reuse vary according to the scale of the phenomenon and the nature of the texts involved (Franzini *et al.*, 2015). Sophisticated visualizations for alignment exist at the micro-level, that is at the level of the word such as the graph visualizations of TRAViz (Jänicke *et al.*, 2015).[1] They facilitate comparative readings of variants within a specific unit of text, and the text in such alignments is fully legible. TRAViz implements the Gothenburg model for textual variance.[2] After input texts are tokenized and normalized, a brute force approach (Jänicke *et al.*, 2014) generates a Variant Graph that reflects similarities and differences on word level among the given text editions. A visualization for such graphs is tailored, so that typical features, e.g. according patterns, synonymity, or transposed words, are highlighted. A clean example of this can be found in the TRAViz alignments of English translations of the Bible (Fig. 3), a textual use case, like the Shakespearean plays mentioned above, in which units and subunits of text are already commonly agreed upon by tradition (e.g. book, chapter, and verse).

On the other hand, solutions exist for macro-level text reuse, such as fingerprinting techniques (Jänicke and Geßner, 2015), creating abstract visual patterns of textual similarity (Fig. 4).

In Fig. 3 with the micro-level visualization, variance across different translations of a single Biblical verse is human readable, with multicolored stream

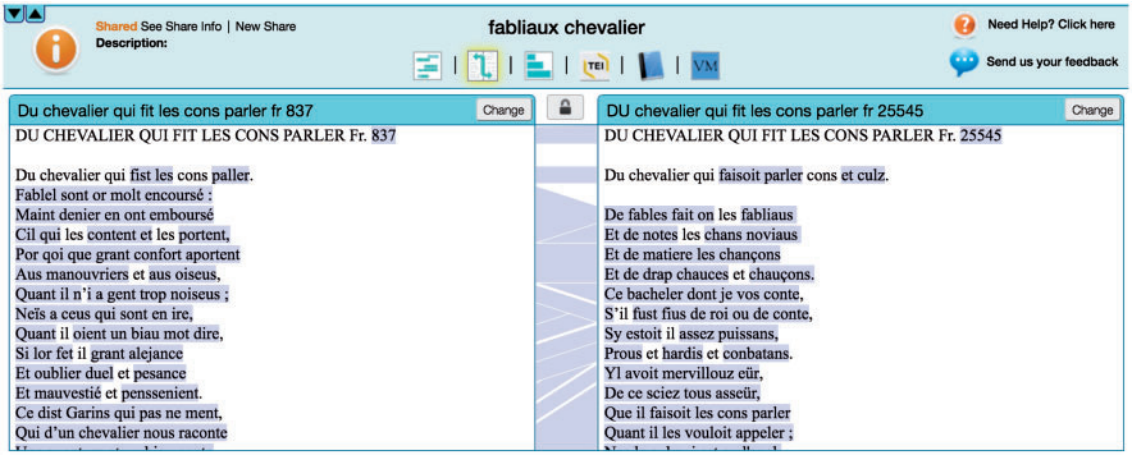

Fig. 2 Two versions of a *fabliau* visualized in Juxta Commons

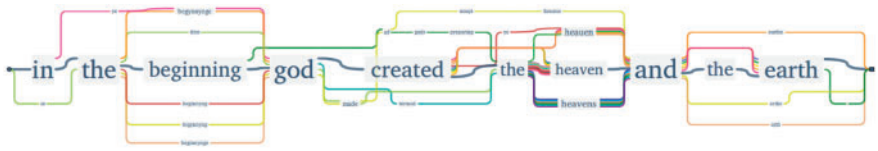

Fig. 3 Micro-level alignment of twenty-four English translations of Genesis 1:1 (Jänicke *et al.*, 2015)

graphs guiding the eye through a reading of the verse. Textual reworking at multiple scales as we find in medieval *mouvance* (shifting of whole blocks of text, variant subline strings, transposition of lines, and reuse in multiple zones of text) cannot, however, be accounted for with the use of TRAViz alignments alone. In Fig. 4, the macro-level completely eclipses the text itself in favor of an abstract representation of computed text reuse. Between these two scales of visualization, the micro- and macro-level, lies an obvious meso-level of visualization that has not been explored by the visualization community. It will be interesting to develop meso-level visualization strategies that can bridge the gap between close and distant computer-assisted reading practices, between micro-views that show only a small subset of text lines and macro-views that provide an abstract, overview of the text. Meso-level visual alignments are desirable for complex textual scholarship scenarios in as much as they allow one to visualize complex patterns of text reuse, preserving text within the visualization itself. They need, however, to stress both legibility and human interaction in the visualization.

6 Meso-level Reading on the Screen

Our design for sufficient representation of *mouvance* in the case of the manuscript versions of the single *fabliau* is shown in Fig. 5 below. The design of our proposed meso-level visualization is inspired by layered graph drawings (Sugiyama *et al.*, 1981). Each edition—given as a plain text file with line-separated verse lines—is drawn as a vertical path. The manually compiled alignment table is provided as an edge list, and each edge represents an alignment of two specific lines belonging to distinct editions. Aggregating these relations forms horizontal

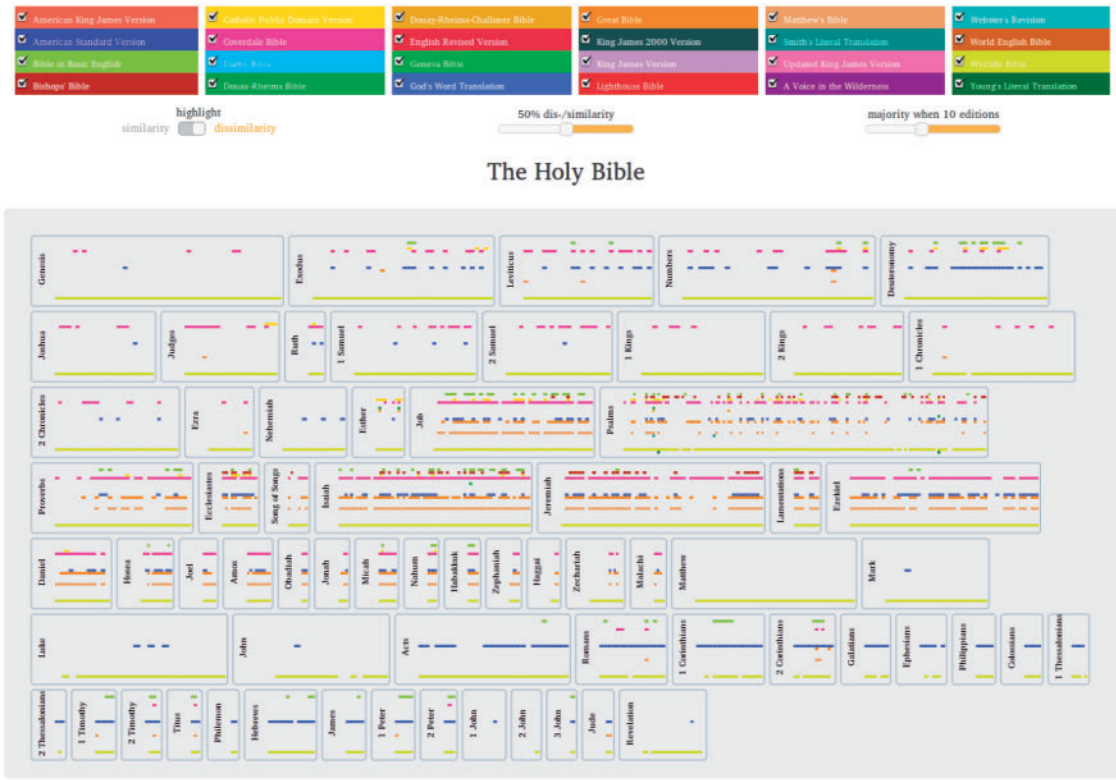

Fig. 4 Macro-level fingerprint illustrating similarity between twenty-four English Bible translations (Jänicke and Geßner, 2015)

alignment paths—visualized as stream graphs (Byron and Wattenberg, 2008)—that illustrate how specific verse lines are construed across the editions. We ordered the editions the way that crossings of streams and occlusions between streams and text are minimal. With the given design, the meso-level visualization supports the visual analysis of manually aligned poetic verse lines and illustrates the instability of one *fabliau* across the four different versions accordingly (Rychner, 1960). The different versions are juxtaposed in columns to minimize edge crossings; in other words, we order the editions according to their similarity. This places the least similar version, London, British Library manuscript, *Harley 2253* at far right. This is for reasons of legibility; the stream graph does not imply any particular textual evolution from one version to another, or from left to right, as Lachmannian critics might privilege the oldest version as the starting point of comparison.

Clicking on a specific verse line in the visualization produces a TRAViz micro-view of the line-level variance (Jänicke *et al.*, 2015), whereas the larger meso-view of this portion of the *fabliau* allows patterns between and across verse lines to be ascertained.[3] Variance in the genre of *fabliaux* generally maintains prosody and avoids hypermeter; *mouvance* is characterized here by the interpolation of larger narrative multi-line blocks of text. The exception to this general rule is the abovementioned *Harley 2253* in the column at far right where the narrative is reconstituted almost completely around sparse line reuse, perhaps illustrating what Rychner calls in the subtitle of his book 'deteriorization' [*dégradation*] (Rychner, 1960). Visualization, in the

Fig. 5 Four versions of a *fabliau* in our visual design

case of *Harley 2253*, shows verse line reprises that seem to preserve a narrative sequence from the other versions, pointing perhaps to the Anglo-Norman redactor's knowledge of the structure of the continental tale in the performative recasting of it. It is worth noting, however, that the streams we have implemented draw our attention to patterns of text reuse for later scholarly interpretation, and yet, the visualization does not assert any inherent spatialized hierarchy of meaning among the versions. Across all of the three versions at left, *mouvance* occurs in multi-line blocks, visible in highly legible streams of text reuse. In a later version of the visualization, it may be desirable to aggregate these for

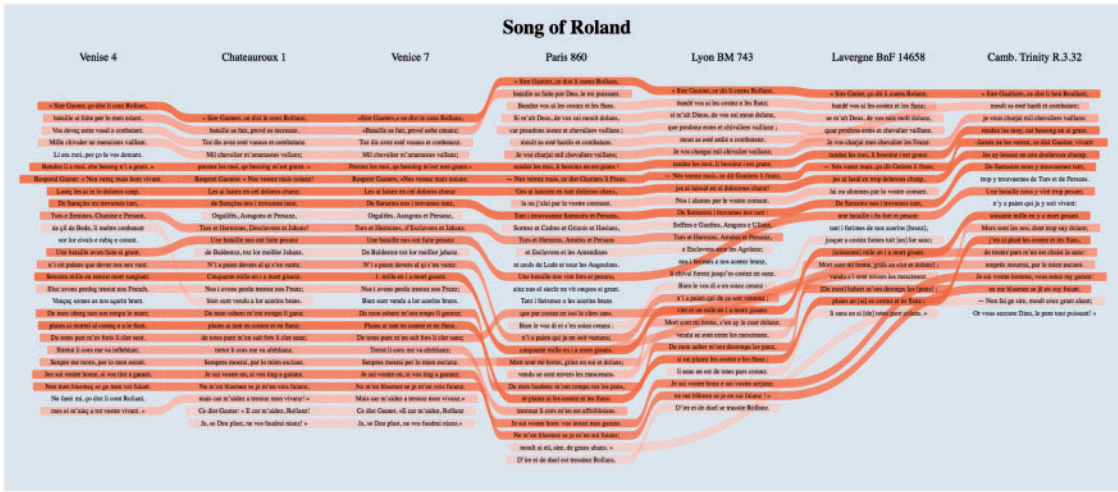

Fig. 6 Seven versions of *Laisse* 1 of the Lavergne fragment of the *Chanson de Roland*, in our visual design

legibility. We see such textual transformation that is not easy to detect in Rychner's hand-aligned synoptic edition where the text is printed in blocks with large blank spaces on the page left in the case of some versions. Another example of what we are calling meso-level textual alignment is the visualization of one stanza from the *Chanson de Roland* tradition contained in seven manuscripts. Figure 6 illustrates *laisse* 1 of the Lavergne fragment, absent from Oxford *Roland*.[4]

As the number of versions of the text grows, obviously our design relies upon a larger screen to view the entire tradition, but we also want to signal frequency of text reuse across those versions. Larger traditions such as *La vie de sainte Marie l'Egyptienne* or the *Evangile de Nicodème* would pose particular challenges in this regard (Dembowski, 1977; Ford, 1973). To indicate how often a line recurs across the whole manuscript tradition, we use colored streams of varying saturation. Highly saturated colors indicate frequently repeated passages, whereas less saturated colors indicate less repeated ones. Such a feature allows for a 'consensus' visualization of the tradition in this meso-view. It is easy to see the more complex, transpositional variance of lines in the *Chanson de Roland*, compared with the abovementioned *fabliau*. This compositional feature of French epic visualized here needs to be studied across the entire corpus of seven manuscript versions and three fragments, and perhaps among other famous epic cycles. As in Fig. 5, we ordered the editions in Fig. 6 to reduce the number of crossing streams, thereby maximizing the legibility of the visualization. This is potentially at odds with readers who expect to see temporality of manuscript dating represented along the horizontal axis of the visualization. In such a case, a more legible order that produces less clutter might be hard to determine.

We began with the alignment of a single *fabliau* and the known versions of the *Chanson de Roland*, since they are traditions where editors have not only published the different textual versions but they have indexed alignment manually, either using page layout or other visual cues appropriate for the printed page. Ultimately, these forms of approximate alignment do not give a full picture of medieval *mouvance*. As the examples shown in Fig. 7 illustrates, variance found in medieval French texts very often occurs below the level of the line, exhibiting different kinds of syntactic, morphological, or lexical variance.

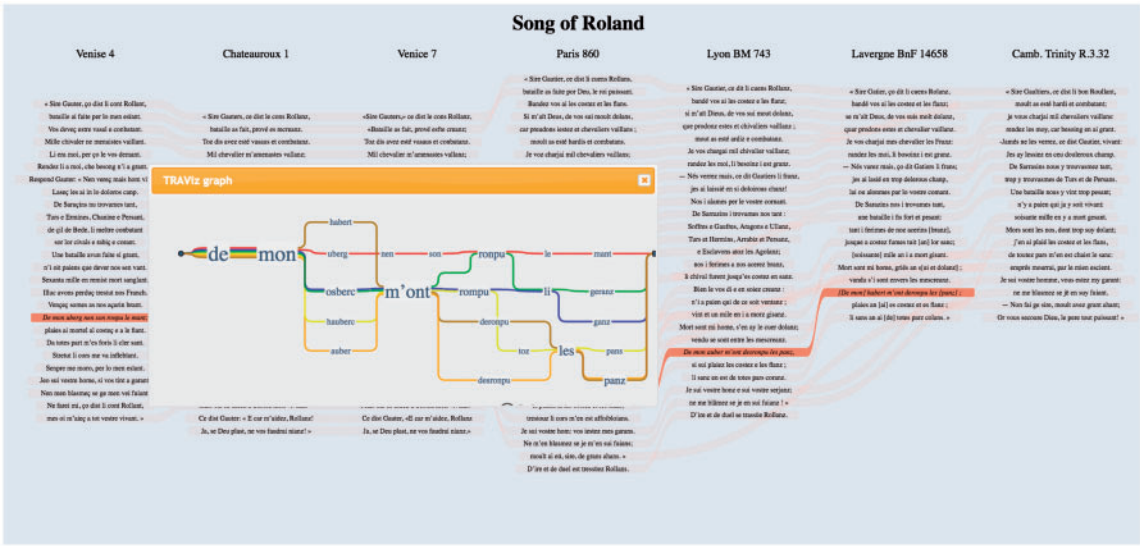

Fig. 7 A TRAViz pop-up visualization of a single aligned verse among seven *Chanson de Roland* edition, illustrating variance below the level of the line

7 From Human to Computational Alignment of the *Chanson de Roland*

The *Chanson de Roland* occupies an important place in the literary historiography of France. The oldest known manuscript is that of Oxford, and it has been edited multiple times. The entire corpus of seven manuscripts and three fragments of the *Chanson de Roland* has only been edited twice in its entirety (Mortier, 1940–44; Duggan, 2005). Both times the different versions were printed serially in a multi-volume work, ostensibly because a visually illustrative edition, as we mentioned above in the case of Rychner's *fabliaux*, would be too complex to execute. It is important to note, however, that Mortier's edition provides a crude alignment of the poems. The traditional stanzas of the Roland tradition, known as *laisses*, are numbered and Mortier created a basic laisse to laisse alignment, showing critical interest in indexing the interrelated narratives of the tradition.

Duggan published a concordance of the Oxford version of the poem with the help of a computer, and completed a monographic study a few years later concerning formulaic language in the *Chanson de Roland* (Duggan, 1969, Duggan, 1973). The concordance allows Duggan to operationalize Parry's notion of the formula (a group of words used in the same metrical conditions to express the same basic idea) and to turn it into a computable, quantifiable feature of text (Parry, 1971). This procedure in medieval French is non-trivial, since there are vast spelling differences and performative elements in the poetry that interfere with the computational detection of the seemingly simple notion of 'a group of words' in and between manuscripts versions.

From Duggan's lists of formulaic speech in the Oxford version of the *Chanson de Roland*, we generated a simple visualization that shows how his examples of formulaic speech are distributed throughout two manuscripts of the *Chanson de Roland*, the Oxford manuscript (left) and the Venice 7 manuscript (right).[5] In Fig. 8 below, we reproduce his taxonomy of formulaic speech (battles, persons, epithets, acts of speech, etc.). The white sliding bars in each of the macro-view columns at left allow different parts of the long poems to be visualized on demand on the screen. The user browsing through this example will find

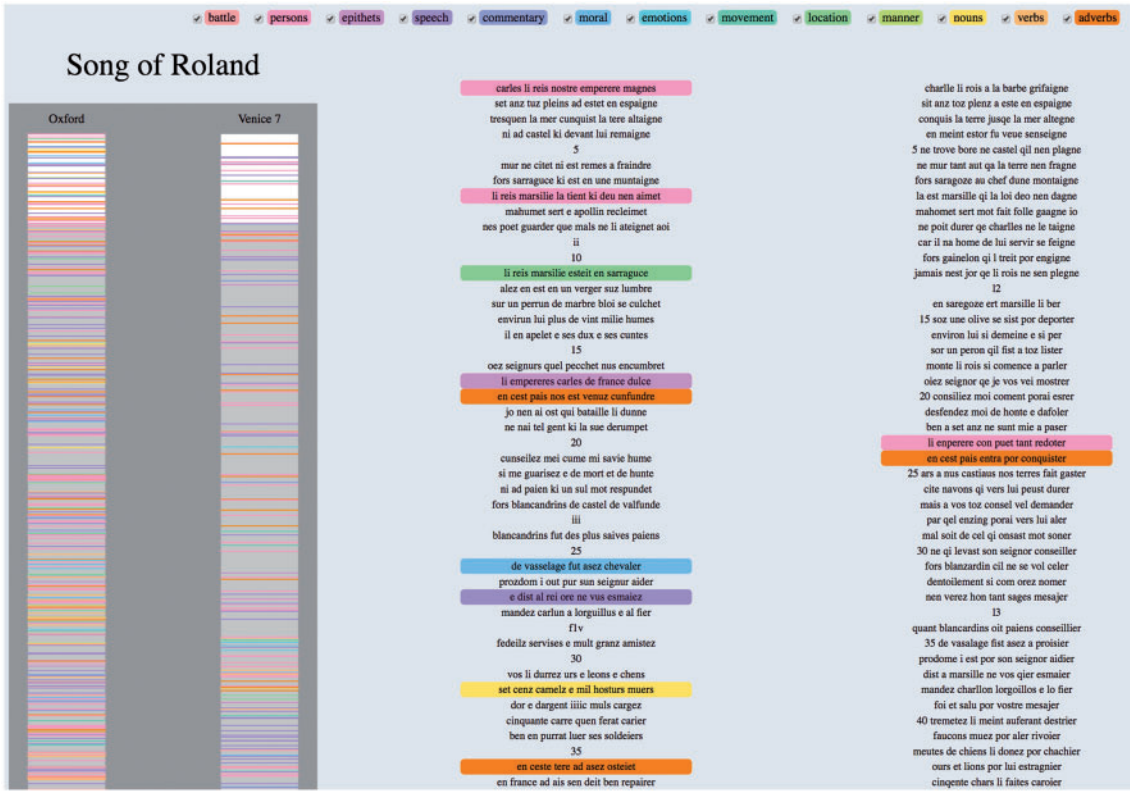

Fig. 8 A visualization of the Oxford and Châteauroux versions of the *Chanson de Roland* with color coding of lines in which different kinds of formulaic speech classified by Duggan (1973) are found

visual evidence for Duggan's claim of the formulaic nature of the *Chanson de Roland*, particularly in the Oxford version; however, formulaic homogeneity is not as clear in the Venice 7 Châteauroux version due to orthography. Duggan's work was foundational for understanding the *Chanson de Roland*, but given the rudimentary nature of the concordance in 1969, he was neither able to take into account broken *n*-grams nor automatically correct for orthographic variance, two shortcomings in the exploration of the fluidity of poetic craft. A more nuanced means of detecting repetition is needed.

Our visualization of the text tradition of the *Chanson de Roland* adopts a user-centered approach. The assumed user of the emergent visual analytics system is the literary scholar aware of the general problems of medieval text traditions. The collaboration between the literary historian and the visualization researcher here can best be called an exploration of both literary and linguistic problems in medieval French and how they might be explored using visualization. In the visualization prototypes that follow, we extend the notion of a synoptic style reading for comparative textual variance to the complex textual tradition of the *Chanson de Roland*. Our prototype is not an intervention into textual criticism in the classical sense. We assert no urtext nor do we make a judgment about the best version of the text; we simply use the available corpus of transcriptions to embrace, and explore, the complexity of the textual situation.

As we have asserted above, this article makes a plea for legible, comparative visualization of textual variance at multiple levels. We sought to produce a

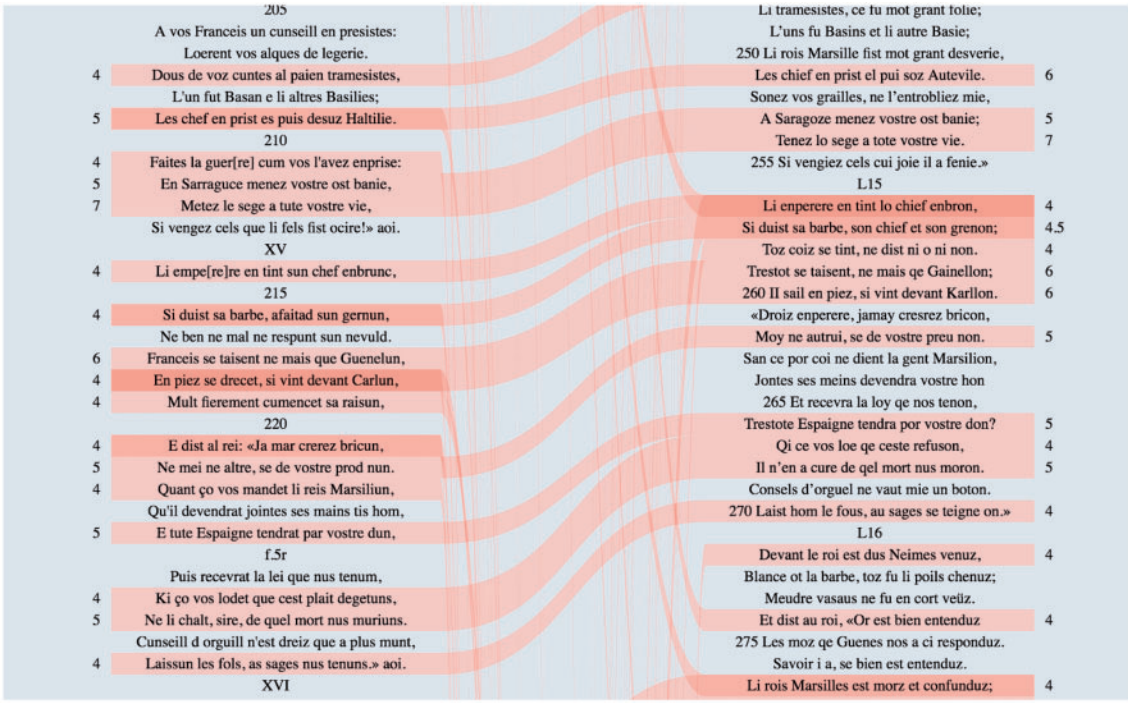

Fig. 9 A Computed alignment using RED (Jänicke *et al.*, 2015) of two versions of the *Chanson de Roland* with statistics of *n*-gram matches next to the used verse line

granular alignment of two versions of the *Roland* using computational means (the Oxford *Roland*, the first known version of the work from the second quarter of the twelfth century, and the Venice 4 *Roland* of the thirteenth century). We used the TRAViz alignment method (Jänicke *et al.*, 2014) to align the entirety of the two texts (Oxford having 4,002 and Venice 4 having 6,002 lines). In preparation, we split both texts O(xford) and V(enice) into lines $O_1,\ldots,O_{4002}$ and $V_1,\ldots,V_{6002}$. For each tuple of lines $\{O_i,V_j\}$, we computed if there is a significant overlap of words or not, computationally speaking, if both lines share an *n*-gram. An *n*-gram is a contiguous sequence of *n* words, and in our case, we require for a tupel $\{O_i,V_j\}$ to be aligned that both O_i and V_j contain the same sequence of *n* words. To take spelling variants into account, we applied a relative edit distance (RED) of strings (Jänicke *et al.*, 2015) with RED = 0.5, which we recognized as the best choice for the given medieval French orthography. One of our main concerns was to create a manageable result set. To assert an alignment, we require at least a shared 4-gram between the two lines of a tuple. Given this configuration, we obtained 1,759 automatically aligned line tuples for 24,020,004 computations. Were we to accept 3-grams, it would have generated 7,118 aligned tuples. Figure 9 above illustrates a side-by-side view of the resultant aligned texts where stream graphs indicate alignment, and saturation is an indicator of the frequency of repetition.

Since alignment is carried out automatically using RED of strings, it is important to have a visual indication of the robustness of the shared, contiguous sequences of *n*-words. Saturation has already been used to indicate the number of times a line has a shared 4-gram or greater across the text(s), so here we use a numeral placed in the margin to indicate the average *n*-gram length between matching lines.

8 Toward Multi-scalar Screen-based Reading of Variance

Unlike the example illustrated above, in Fig. 9, we visualize the entirety of the two poems at once, although this is not immediate visible from the included screen crop. By scrolling up or down in the prototype, one can access the rest of the text of the two poems. The stream graphs indicate in both the visual field of the screen as well as far above and far below it, how the *mouvance* in complex medieval texts occurs. As a computationally aligned visualization, it is significantly different from the human-aligned text fragments found above in Figs 5 and 6. Whereas we suggested that the Anglo-Norman redactor of the *fabliau* in question may have been preserving the narrative sequence of the continental version of the poem, this is perhaps an illusion created by the fact that only a small segment of the text is visualized. If we think of the screen as one would the human mind, bounded in its memory to a certain field of comparison, it is possible that this is how a text editor makes decisions about variants, from within a bounded space determined more by a narrative unit than by the overall language of the poem. As expected, there are significant echoes composed of *n*-grams across the two versions of the *Roland*, and of differing densities in different sections of the versions. Importantly, there are significant echoes within the body of a single version as well confirming existing claims about formulaic speech (Duggan, 1973). Generally speaking, these results lead us to call for an expansion of what the performative variance in texts, Zumthor's *mouvance*, might constitute when they are aligned and visualized.

Scholars of the medieval French *chanson de geste*, and the *Roland* in particular, have long studied the 'formulaic style' of the poetic composition. Critical studies of the genre have urged, even attempted, quantitative approaches to its repetitive compositional style (Rychner, 1955; Nichols, 1961; Duggan, 1973; Martin, 1987). Our computed visualization provides an innovative standpoint from which to re-examine the phenomenon of textual repetition more systematically and perhaps to begin to document *mouvance* more empirically. It could open new critical dialogue with a long-standing body of literature of the oral formula in ancient Greek or other orally inflected literatures.

The visualization of the medieval French *chanson de geste* in our design illustrates a high degree of string repetition across the full text. In the computed alignment, only *n*-grams equal to or greater than four were retained for visualization. The rationale for this is that, given the high degree of orthographic variance of the medieval French, matching of less than 4-grams might lead to a high number of false positives. Even with 4-grams, occasional false positives occurred. Figure 10 illustrates how 'il le vit a' can be matched with 'il ne vint a', whereas there is no meaningful connection between the lines. Nonetheless, a preliminary check of the alignment showed such cases to be relatively uncommon. The merging of nodes with the RED of strings is more common among sentences containing shorter (often function) words.

Given the screen limitations in representing the visual alignment of such long poems, the robust, micro-level TRAViz pop-up provides an important topological overview of how poetic fragments are stretched and deformed across a wide expanse of text. The TRAViz view shows very clearly how *n*-gram matches in the *Roland* vary in their position in the verse line. It has been claimed that where such repetitive language sits in the verse line is characteristic of the oral poet's craft, or of the more literary text-specific imitations of orality (Duggan, 1973; p. 11). Even Duggan provides a critical caveat that such text reuse does not always sit firmly in either hemistich. We find many examples of a frontal text reuse, in particular, of a discursive sort, introducing direct discourse or action as in Fig. 11 below.

Likewise, there are many equivalent examples of terminal text reuse, that is falling at the end of the verse line. More surprising to us were the equally large number of examples of medial text reuse, straddling the hemistich, a form of repetition less remarked upon by critics.

What we find in Fig. 12 below is looking at the Venice 4 *Roland* from the perspective of the Oxford *Roland*. First of all, the existence of multiple blue streams indicates all the lines in which the Venice 4 version has *n*-gram matches with line 1,254 of the Oxford *Roland*. The different verse lines in Venice 4

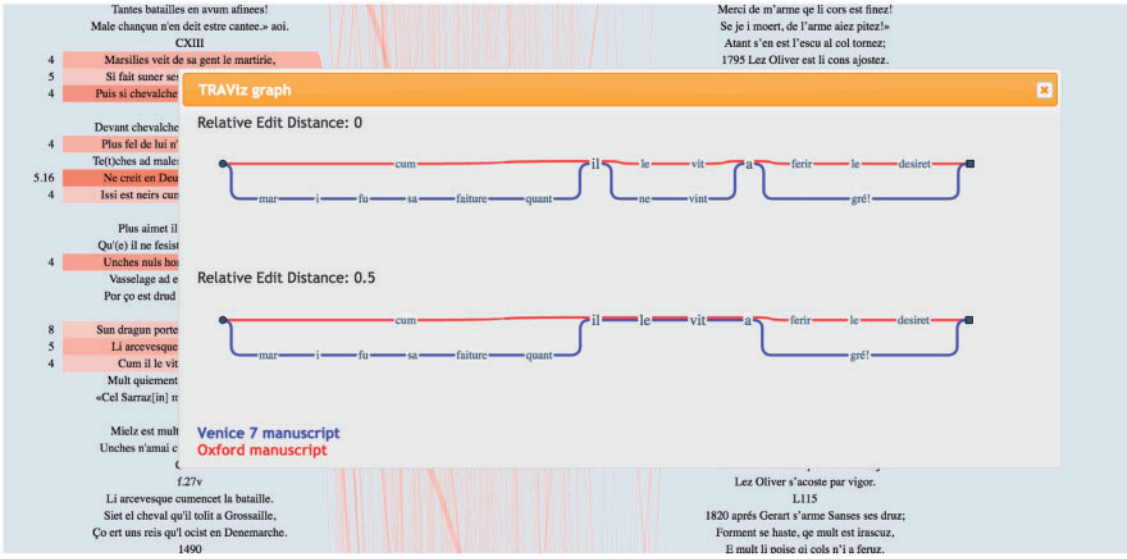

Fig. 10 A TRAViz pop-up illustrating a false positive in alignment

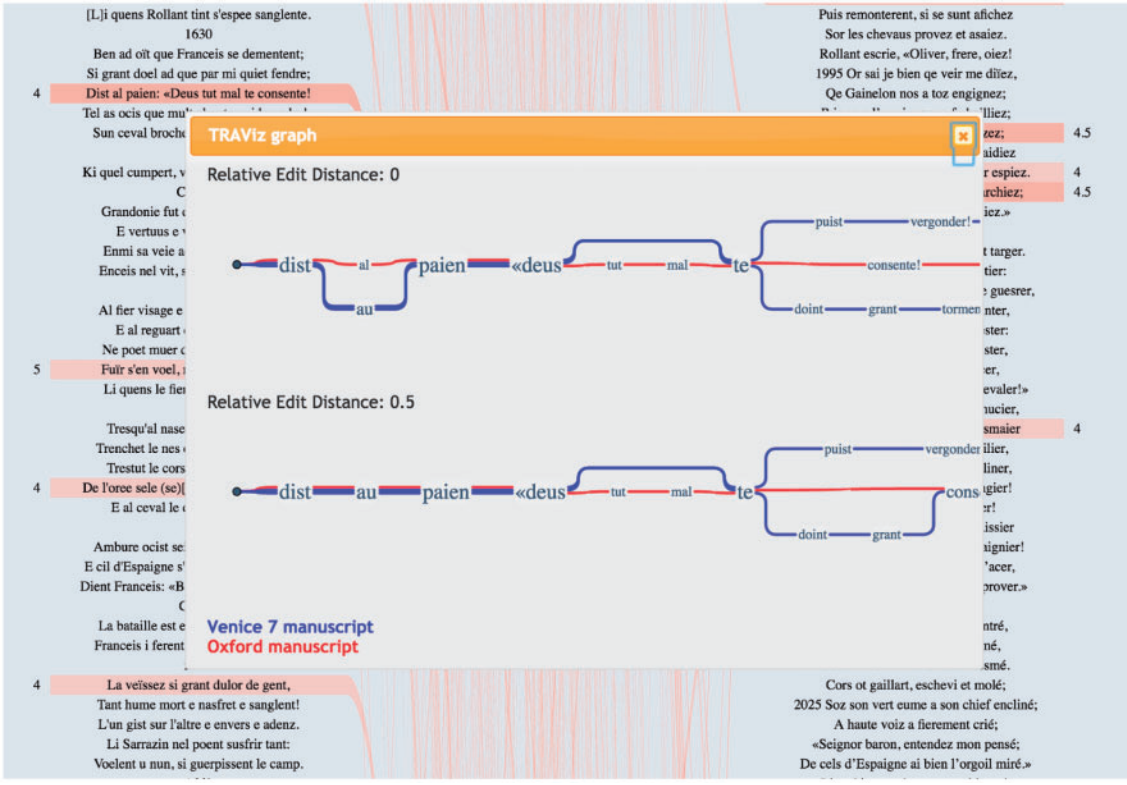

Fig. 11 A TRAViz pop-up example of frontal text reuse, line 1,565 in the Oxford Roland, introducing one of the many curses of the pagan, visualized with and without the application of RED

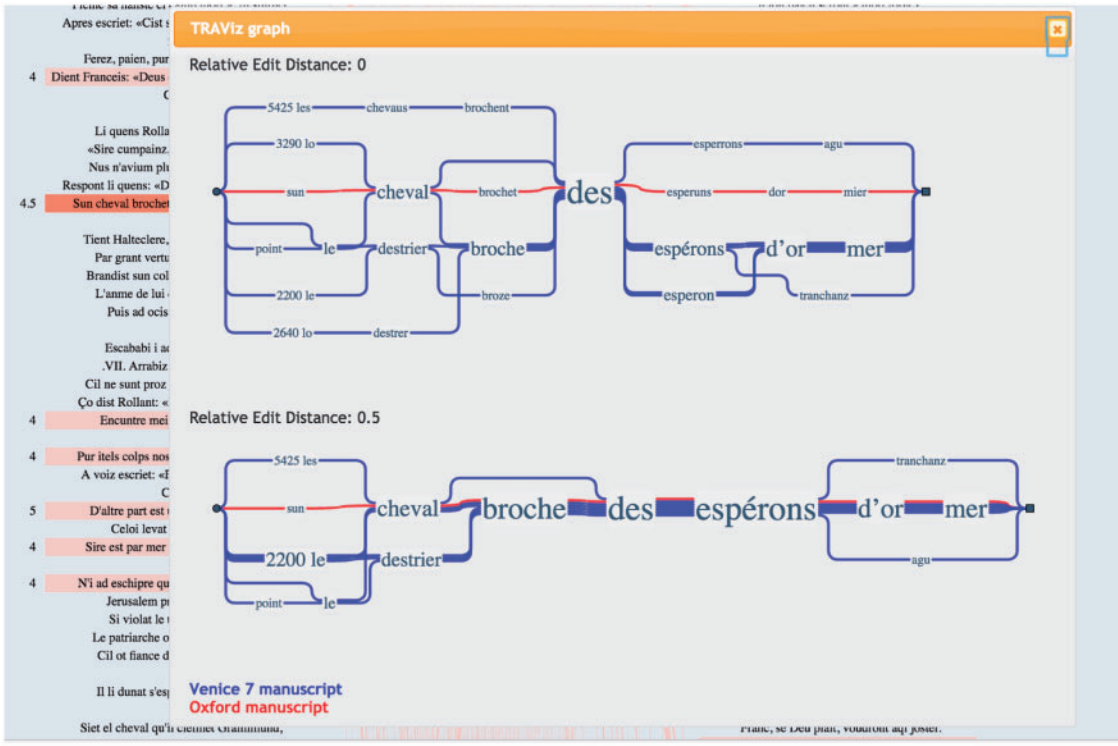

Fig. 12 A TRAViz pop-up example of medial text reuse, line 1,254 of the Oxford Roland, describing the decorative spurs of the horses, visualized with and without the application of RED of string measure

are created in a combinatory fashion around that *n*-gram, with the remaining sections of the line constructed in semantically similar strings. If we factor in the question of temporality and, therefore, of possible influence, it is tempting to conclude that the Venice 4 *Roland* recreates the Oxford *Roland*, or text blocks reminiscent of it, proliferating them by means of the combinatory reuse of text fragments. On the other hand, when we look backward from the perspective of the Venice 4 *Roland* to the Oxford *Roland*, one finds a similar phenomenon, a single line of the later text with which the earlier text matches. One must proceed with caution in overanalyzing the question of influence and intertextuality using visualization. There are nonetheless highly saturated lines on both sides that give provide flavor of what kind of repetitive language is more characteristic of both poems. These high-frequency *n*-grams, no doubt, indicate a free-floating pool of oral formulae repeated, and poetically reactivated, to different extents in individual textual instances.

The last prototype that we created attempts to merge all the previous ones, providing an optimal environment for exploring the full picture of *mouvance* in medieval French epic poetry, implementing the Information Seeking Mantra (Shneiderman, 1996): 'Overview first, zoom and filter, then details-on-demand'.[6] The overview component is a macro-level overview bar (at right in Fig. 13 below), maximizing the height of the screen both to illustrate the larger picture of *mouvance* in a text-less visualization of the two versions of the *Roland*. It is linked to the meso-level view described above, and it can be used to navigate the text and to zoom to occurring alignment patterns. We click on a section of either vertical bar representing the two poems in the overview to select the desired visual frame. Zones of high density in orange point to areas of more intense reuse in the rest of the poem.

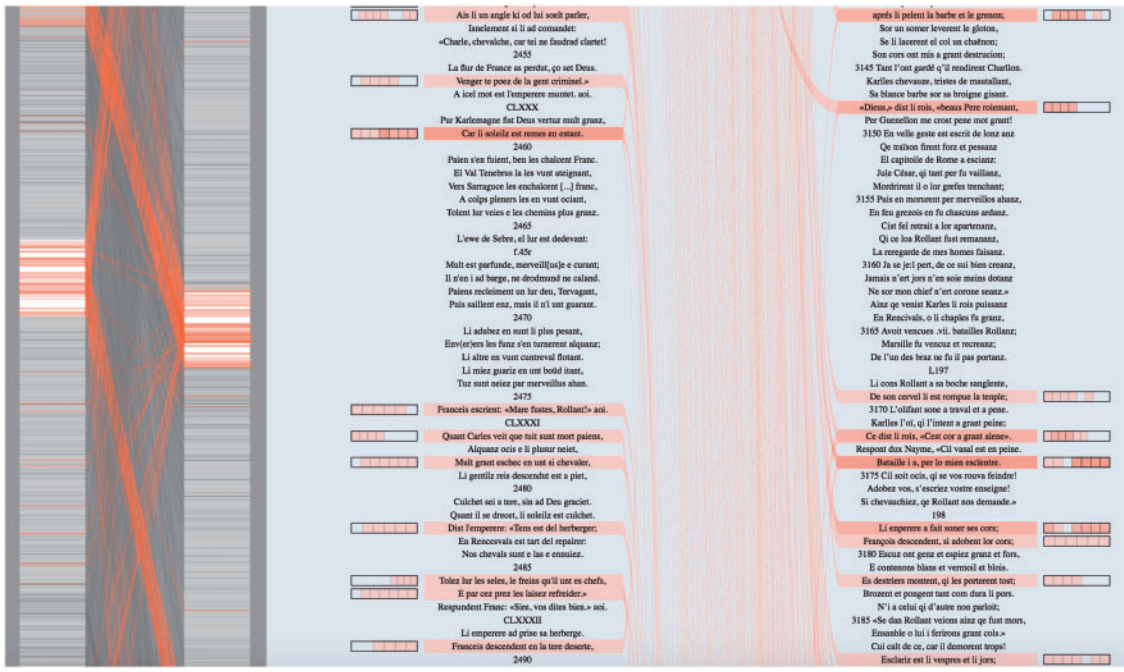

Fig. 13 Depicts an integrated visualization of meso- and macro-views of zones of high reuse in the poetic comparison around laisse 280. The white segments at left indicate the sections of two versions of the Roland visualized at right in the text columns

One high reuse zone indicated by a dense set of orange lines falls around *laisse* 280 in the Oxford *Roland*, the part of the poem in which Charlemagne is hurt in the confrontation with the emir Baligant's armies, at the point where Thierry and Pinabel fight a duel. Such a pattern might lead the interested researcher to investigate how the commonly matched fragments of texts in the Oxford version potentially impact the diction in the Venice 4 version. This is but one of many examples of the potential value of applying the comprehensive view (micro, meso, and macro) to the exploration and interpretation, of repetition in the *Roland*. One small addition to this interaction that serves as a meso-level fingerprint of micro-level alignments shown in TRAViz is the small rectangular bars found to the left and right of the saturated stream graphs. Bertin called these 'frame rectangle symbols' (Monmonier, 1993). These 'rectangular heat bars' are divided into buckets matching word segmentation on the verse line. Color indicates the position in the line where the *n*-gram match takes place and saturation of the same color of the stream graph points to the frequency of reuse across the text. This is designed as a convenient overview, not only as a visual cue to drill down to micro-level zones of high reuse but also to give an indication of what kind of positional reuse it is: frontal, medial, or terminal.

9 Conclusion

Textual reuse, or repetition, what Duggan called 'formulaic style', across the different versions of the *Roland* points to a genre-specific oral-like language that scholars have suggested is integral to the process of poetic creation. There are a number of problems in text traditions with such oral inflection. There are not only different kinds of variance (single word/string variance, half-line or hemistich variance, transposition and reorganization of rhyming verse lines, or interpolation of entirely new lines) but patterns of variance are also not uniform

across a text, making the desired comparative visualization of texts difficult, even pushing screen-based representation to the limit. Duggan's 1969 concordance was only a small first step to understand this complexity. Alignments can also mix and confuse kinds of variance. Our article has examined possible ways to handle such heterogeneous forms of variance, allowing them to be assembled carefully and empirically in a panoptic view. Setting the *n*-gram at a minimum of four limits the kinds of variance we have detected, assuring that computationally derived alignment is not gratuitous, but further research may experiment with other parameters and methods of alignment.

We propose a prototype, not a final implementation of a visualization system, that allows for variants to be tracked and visualized in a way that no print-bound critical apparatus or concordance could possibly visualize. Our visualizations attempt to provide a framework for a medievalist's 'comparative, not archeological' reading in the words of Cerquiglini, and this, in as interactive and intuitive a fashion as possible. Moving from individual word variants to more complex forms of variance and repetition should provide insight into textual behavior at a larger scale. It suggests that the concept of an urtext, a lost original document that could be reconstructed from which all other witnesses derive—a concept against which different traditions of textual criticism reacted quite vigorously—for certain generic forms might be replaced altogether by a shared pool of computationally collected multi-word strings, lines, or line segments of lesser or greater orthographic similarity. Much more research needs to be done to consider the evolving genre of the more than 200 examples of *chanson de geste* currently documented (Suard, 2011). There is also room, however, for comparative approaches between and across genres. The example of the single *fabliau* visualized at the beginning of the article provided an example of a more 'written' and literary recombination of text blocks.

Historical text reuse must think not only about the algorithms used to discover that reuse but also the frameworks in which such reuse can be visualized. In contrast to the manual, narrative-driven alignment illustrated in the visualizations above (Figs 5 and 6), an automated alignment yields a very different picture with respect to the degree of *mouvance*. Computational approaches to *mouvance* produce much more than just semantically significant variants. It remains to be seen how literary history will receive such an expanded definition of *mouvance*. Will they find the repetition of the *chanson de geste* 'monotones—comme sont monotones les sculptures romanes—et même gothiques de nos cathédrales, avec leurs thèmes identiques' [*monotonous, as the Romanesque, even Gothic sculptures of our cathedrals, with their repetitive themes*] (Lejeune, 1954; p. 331) or will it be the fodder of new forms of distant pattern searching?

We argue here that the complexity of the intersection of literary traditions and visualization is a fertile field for future research, not only for the insights they offer of larger-scale problems in the textual humanities but also inasmuch as they challenge the visualization community to integrate non-contradictory, visual languages for multi-scalar problems particularly at the meso-level, where the very results of computation on text are incorporated into the visualization, rather than being occulted.

Future work focuses on the improvement of the computational alignment by turning this visual analysis method into a visual analytics approach that keeps the 'human in the loop'. This could be accomplished through the development of scoring sessions, allowing the textual scholar to rate individual alignments according to their relevance. For example, false positives such as shown in Fig. 11 could be marked accordingly. Also, undiscovered alignments could be manually added. With this human-generated input, we can design quality measures that reconfigure the proposed alignment algorithm. Examples of such measures include taking *n*-gram length, character length of words in *n*-grams, or percentage of matching words across the whole line into account. Following these scoring sessions, the alignment can be recomputed and the visualizations updated. Additional design features such as order the editions and filtering may be implemented, but taking a user-centered approach,

alignment is not to be understood as a final product, but rather a process for understanding variant text traditions, supporting the generation of new hypotheses about textual behavior, and even integrating the intricacies of medieval *mouvance* into new visualization strategies.

References

Bowers, K. (2016). Tweet during Ted Underwood's lecture at #emcmillionwords. https://twitter.com/kab3d/status/794618896741318656 (accessed 4 November 2016).

Brun, L. (2016). La Chanson de Roland. In *ARLIMA Archives de littérature du moyen âge*. http://www.arlima.net/no/11 (accessed 31 October 2016).

Byron, L., and Wattenberg, M. (2008). Stacked graphs–geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, **14.6**: 1245–52.

Cerquiglini, B. (1999). *In Praise of the Variant: A Critical History of Philology*. Baltimore/London: The Johns Hopkins Press.

Dembowski, P. (1977). *La vie de sainte Marie l'Egyptienne, versions en ancien et en moyen français*. Geneva: Droz.

Duggan, J. (1969). *A Concordance of the Chanson de Roland*. Columbus: Ohio State University Press.

Duggan, J. (1973). *The Song of Roland. Formulaic Style and Poetic Craft*. Berkeley/Los Angeles/London: University of California Press.

Duggan, J. (ed.) (2005). *La Chanson de Roland/The Song of Roland: The French Corpus*. Geneva: Droz.

Ford, A. E. (1973). *L'Evangile de Nicodème: les versions courtes en ancien français et en prose*. Geneva: Droz.

Franzini, G., Franzini, E., and Büchler, M. (2015). *Historical Text Reuse: What Is It?*. http://etrap.gcdh.de/?page_id=332 (accessed 3 March 2016).

Jänicke, S., Geßner, A., Büchler, M. and Scheuermann, G. (2014). Visualizations for text re-use. In *2014 International Conference on Information Visualization Theory and Applications (IVAPP)*, pp. 59–70.

Jänicke, S. and Geßner, A. (2015). A distant reading visualization for variant graphs. In *Proceedings of the Digital Humanities 2015*. University of Western Sydney, 1 July 2015. http://dh2015.org/abstracts/xml/JANICKE_Stefan_A_Distant_Reading_Visualization_fo/J_NICKE_Stefan_A_Distant_Reading_Visualization_for_Vari.html (accessed 1 August 2017).

Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S. and Scheuermann, G. (2015). TRAViz: A Visualization for Variant Graphs. In *Digital Scholarship in the Humanities*, **30(suppl 1)**: i83–99. http://www.informatik.uni-leipzig.de/~stjaenicke/TRAViz.pdf (accessed 27 October 2015)

Lejeune, R. (1954). Technique formulaire et la chanson de geste. *Le Moyen âge*, **60**: 331–4.

Martin, J. -P. (1987). Les motifs dans la chanson de geste. Définition et utilisation. *Cahiers de civilisation médiévale*, **30**(120): 315–29.

Méla, C. (ed.) (1990). *Le Roman de Perceval* (Chrétien de Troyes). *Édition du ms. 354 de Berne, traduction critique, présentation et notes*. Paris: Librairie générale française.

Moffat, M. (2014). *The Châteauroux Version of the "Chanson de Roland": A Fully Annotated Critical Text*. Berlin/Boston: De Gruyter.

Monmonier, M. (1993). The visual variables of Jacques Bertin. In *Mapping It Out: Expository Cartography for the Humanities and Social Sciences*, Chicago: University of Chicago Press, 59–70.

Mortier, R. (ed.) (1940-44). *Les textes de la "Chanson de Roland."* Paris: Éditions de la Geste Francor.

Nichols, S. (1961). *Formulaic Diction and Thematic Composition in the Chanson de Roland*. Chapel Hill: University of North Carolina Press.

Nichols, S. (2016). *From Parchment to Cyberspace: Medieval Literature in the Digital Age*. New York: Peter Lang.

Parry, M. (1971). *The Making of Homeric Verse*. Ed. Parry, A. Oxford: Clarendon Press.

Pierazzo, E. (2011). A rationale of digital documentary editions. *Literary and Linguistic Computing*, **26**(4): 463–77.

Richards, E. J. (1987). Finding the "authentic" text: editing and translating medieval and modern works as comparable interpretive exercises (Chrétien's Charrette, Christine de Pizan's Cité des Dames, and Diderot's Neveu de Rameau). *L'Esprit créateur*, **27**(1): 111–21.

Robinson, P. (1989). The collation and textual criticism of Icelandic manuscripts (1): collation. *Literary and Linguistic Computing*, **4**(2): 99–105.

Robinson, P. (1989). The collation and textual criticism of Icelandic manuscripts (2): textual criticism. *Literary and Linguistic Computing*, **4**(3): 174–81.

Roman de la Rose Digital Library (2017). http://romandelarose.org/ (accessed 1 August 2017).

Rychner, J. (1955). *La Chanson de geste. Essai sur l'art épique des jongleurs*. Geneva/Lille: Droz/Giard.

Rychner, J. (1960). *Contribution à l'étude des fabliaux: variantes, remaniement, dégradations*. Geneva: Droz.

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*. Washington: IEEE Computer Society Press, pp. 336–43.

Suard, F. (2011). *Guide de la chanson de geste et de sa postérité littéraire (XIe–XIVe siècles)*. Paris: Champion.

Sugiyama, K., Tagawa, S., and Toda, M. (1981). Methods for visual understanding of hierarchical system structures. *IEEE Transactions on Systems, Man and Cybernetics*, **11**(2), 109–25.

TransVis (2016). Version, variation, visualization. http://www.delightedbeauty.org/vvv (accessed 5 November 2016).

Uitti, K. *et al.* (2006). Project Materials, The Princeton Charrette Project. http://www.princeton.edu/~lancelot/ss/materials.shtml (accessed 31 October 2016).

Versioning Machine (2015). http://v-machine.org/samples/prophecy_of_merlin.html (accessed 27 October 2015).

Wheeles, D. and Jensen, K. (2013). Juxta Commons. In *Proceedings of the Digital Humanities 2013.* University of Nebraska-Lincoln, 17 July 2013. http://dh2013.unl.edu/abstracts/ab-142.html (accessed 1 August 2017).

Zumthor, P. (1992). *Toward a Medieval Poetics*. Minneapolis: University of Minnesota Press.

Notes

1 Other platforms supporting the micro-analysis of textual variance such as CollateX (http://collatex.net/) or Stemmaweb (https://stemmaweb.net/) are discussed in (Jänicke *et al.*, 2015).

2 On the Gothenburg model, see http://wiki.tei-c.org/index.php/Textual_Variance.

3 The visualization is available at http://informatik.uni-leipzig.de:8080/Fabliaux/.

4 The visualization is available at http://informatik.uni-leipzig.de:8080/roland/index2.html?ftsize=11.

5 The visualization is available at http://informatik.uni-leipzig.de:8080/Duggan/.

6 The visualization is available at http://informatik.uni-leipzig.de:8080/EditionAlignment2/.