

‘Distant Reading’ in French Music Criticism

Estelle Joubert
Dalhousie University
Email: Estelle.Joubert@dal.ca

*This article offers a series of experiments exploring the potential for ‘distant reading’ in French music criticism. ‘Distant reading’, a term first coined by literary theorist Franco Moretti, refers to quantitative approaches that allow for new insights into a large corpus of texts by aggregating data. While the main corpus employed here is the *Revue et gazette musicale de Paris* (1831–1877), I also use secondary corpora of reviews of Félicien David’s *Herculanum* in 1859, Berlioz’s reviews of Gluck and Beethoven in the *Journal des débats* and reviews that mention Gabriel Fauré in the Library of Congress’ *Chronicling America* database. My experiments employ a text analysis tool named *Voyant*, built by Geoffrey Rockwell and Stéfan Sinclair, thereby also offering a basic introduction to the range of visualizations employed in distant reading. My experiments focus on areas in which quantitative methods are particularly well suited to generating new knowledge: corpus-wide visualizations and queries, moving beyond traditional text searching, investigations of music critics’ authorial styles and detecting sentiment in reviews, and finally, to geographies of music criticism.*

The nineteenth century is widely regarded as the golden age of music criticism, a period during which the sheer number of critics and reviews increased sharply, and the geographic coverage of music criticism expanded vastly. While musicologists have focused on close readings of particular musical works, often looking to criticism to inform these readings, quantitative approaches offer alternate methodologies to study nineteenth-century music. One such approach is ‘distant reading’. Coined by literary theorist Franco Moretti, the term refers to research employing quantitative methods to process the content of large datasets.¹ Queries range widely, often featuring patterns which could include the geographic distribution

I gratefully acknowledge Paul G. Doerwald’s technical assistance and support, especially in building various corpora and with aspects of data manipulation and management. At the Austrian National Library, the Open Access policies of ANNO (AustriaN Newspapers Online) greatly facilitated the building of my main corpus. Finally, I would like to thank Robin Butterhof in the Newspaper and Current Periodical Room at the Library of Congress for assisting me in navigating the API (application program interface) of their *Chronicling America* project, which was indispensable for the section on geographies of French criticism in this article.

¹ This idea was first discussed in Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso, 2005) and subsequently in his *Distant Reading* (London: Verso, 2013). For an excellent critical introduction and case studies using this methodology, see *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*, ed. Matt Erlin and Lynne Tatlock (Rochester: Camden House, 2014).

of a particular genre, changes in vocabulary or semantics, or character-networks in dramatic works, among others.² Whereas ‘distant listening’ of scores requires music encoding,³ rendering this scholarship extremely rare thus far, substantial amounts of nineteenth-century music criticism are already digitized and at least partially encoded, making it, at first glance, an obvious target for computational analysis. But what might distant reading of nineteenth-century music criticism yield?

Quantitative methods are most productive when they execute tasks that humans, unaided, are unable to complete. Distant reading is based on the premise that it is impossible for a single scholar to read an entire corpus of literature, whether it be nineteenth-century novels or music criticism. Indeed, it means that scholars employing distant reading do not actually read the texts, as this task is delegated to technology; it is precisely this facet of his method for which Moretti has been most sharply criticized.⁴ While comprehensive readings of entire corpora are certainly attractive, the ideal remains at least somewhat fleeting. As Moretti quickly learned, what became clear

was the enormous difference between the archive of the Great Unread, and the world of the canon. You enter the archive, and the usual coordinates disappear; all you can see are swarms of hybrids and oddities, for which the categories of literary taxonomy offer very little help. It’s fascinating, to feel so lost in a universe one didn’t even know existed; but it’s hard to extract a rational picture from the *Walpurgisnacht* of discordant voices. And then, to make matters worse, there is the opposite problem, too: working with large quantities, the average becomes an inevitable presence—and the average means loss of distinction, slowness, boredom ... Too much polyphony, and too much monotony.⁵

As Moretti aptly sums up, large corpus analysis, in which every text is theoretically treated as equal, is fraught with challenges and rewards. While his nod to ‘entering the archive’ suggests moving beyond the canon, it should be made clear from the outset that technology does not remove bias. To be sure, human judgement is inherent in how queries are constructed, results are visualized, and conclusions are drawn. Yet the allure of delving into larger data sets does present an opportunity to gain novel perspectives on a particular corpus. Standing back allows one to

² Moretti’s methods are cultivated at the Stanford Literary Lab. Recent publications include Franco Moretti, ed., *Canon/Archive: Studies in Quantitative Formalism* (New York: n+1 Foundation, 2017). For one of the first investigations in musicology, see Elizabeth Monzingo and Daniel Shanahan, ‘The Expression of Self and Grief in the Nineteenth Century: An Analysis through Distant Readings’, *Nineteenth-Century Music Review*, online first, 2020, doi:10.1017/S1479409819000697.

³ Michael Scott Cuthbert first used the term ‘distant listening’ in a paper entitled ‘Distant Listening/Digital Musicology: music21 and Compositional Similarity in the Late Middle Ages’, presented at the Heyman Centre, Columbia University on 1 May 2018. The Music Encoding Initiative (MEI) is an international community effort ‘to define a system for encoding musical documents in a machine-readable structure’. See <https://music-encoding.org/>.

⁴ See for instance Jennifer Schuessler, ‘Reading by the Numbers: When Big Data Meets Literature’, in *The New York Times*, 30 October 2017, and Kathryn Schulz, ‘The Mechanic Muse: What Is Distant Reading?’ in *The New York Times*, 24 June 2011.

⁵ Moretti, *Distant Reading*, 181.

see patterns previously unnoticed. Quantitative methods can verify trends or quantify the presence of themes and ideas that scholars observe anecdotally through close reading. Implicitly, in a distant reading, these patterns or themes require technology to come to the fore. Often, it also requires technology to make sense of the results, lending credence to Moretti's frustrations with a simultaneous lack of anchors as well as a myriad of apparently monotonous results.

This essay offers a series of experiments with distant reading in nineteenth-century French music criticism. My main corpus is limited to the *Revue et gazette musicale de Paris* (1831–1877), in part for technical reasons.⁶ Based on this weekly journal covering musical life in Paris, the corpus contains some 22,437,997 words. For my analyses and visualizations I have used a tool named Voyant, built by Geoffrey Rockwell and Stéfan Sinclair and featured in their book, *Hermeneutica*.⁷ Readers may wish to explore the visualizations via the weblinks, as the figures are interactive; that is, more information is revealed by hovering over a particular word or image, or clicking through. The tools associated with distant reading thus also present new implications of what we consider to be knowledge.⁸ Figures have long supplemented scholarly articles, but these have traditionally been still. Playfully titled 'There's a Toy in My Essay', Rockwell and Sinclair devote an entire chapter to problems of hermeneutical widgets embedded in scholarship.⁹ As readers will realize, this essay's interactive visualizations invoke new ways of producing, communicating and experiencing knowledge. The complete 'picture' is not captured in the Figures; instead, they represent a starting point from which readers can explore the data more fully.

My experiments involve areas where quantitative methods are particularly well suited to generating new knowledge: corpus-wide visualizations and queries, moving beyond traditional text searching, investigations of music critics' authorial styles and detecting sentiment in reviews, and finally, to geographies of music criticism. Since Voyant is a tool primarily intended for scholars of literature, it is worth noting from the outset the particularities of music periodicals as objects for quantitative methods. Periodicals are written by multiple authors, which makes computational investigations of a particular reviewer's aesthetic or political views more complicated than, for instance, a novelist; this would involve training the computer to recognize the beginnings and endings of each periodical section along with authorship for each section. For investigations of a single-author's contributions

⁶ For a brief history of the founding and prominence of the journal in nineteenth-century France, see Ralph Locke's article in this issue, 'How Reliable Are Nineteenth-Century Reviews of Concerts and Operas?: Félicien David's *Le Désert* and His Grand Opéra *Herculanum*'. Quantitative methods in text analysis require consistent and reliable text-files. I have assembled my corpus using ANNO (AustriaN Newspapers Online), an open access repository which makes .txt files readily available. See <http://anno.onb.ac.at>.

⁷ <https://voyant-tools.org/>. Geoffrey Rockwell and Stéfan Sinclair, *Hermeneutica: Computer-Assisted Interpretation in the Humanities* (Cambridge: MIT Press, 2016).

⁸ For a discussion on 'knowledge bearing instruments and practices' see Rockwell and Sinclair, *Hermeneutica*, 150–52.

⁹ See Chapter 5: 'There's a Toy in My Essay: Problems with the Rhetoric of Text Analysis', in Rockwell and Sinclair, *Hermeneutica*, 83–104.

(e.g. Berlioz's music reviews in the *Journal des débats*), creating a separate corpus is the preferred approach.

One of the underlying currents in this experimental exploration is how tools shape the research questions or even approaches particular to musicology. As one digital humanities team remarked, 'tools are not just tools. They are cognitive interfaces that presuppose forms of mental and physical discipline and organization. By scripting in action, they produce and transmit knowledge, and, in turn, model a world'.¹⁰ Musicologists like me who are accustomed to close reading in reception studies have developed strong text-search skills to navigate an ever-increasing body of reviews dealing with music, among other reception documents. We are trained to then generate nuanced arguments with the evidence gathered. The aim of these experiments is to explore new ways of gathering evidence and, by extension, new ways of interpreting musicological evidence.¹¹ Navigating the proverbial archive of nineteenth-century music criticism with a range of new tools at hand may serve as a critical introduction to the relevant practical and conceptual issues and inspire future work in the field of quantitative musicology.

Entering the Archive

We begin by exploring the *Revue et gazette musicale de Paris* from 1831 to 1877 – a room in the 'Great Unread' of nineteenth-century French music criticism. The aim, ideally, is to simply experience some first impressions; to see what catches one's eye. One of the simplest and widely used visualizations is a *word cloud*.¹² Having selected French as language and implemented some stop words (a list of words that the computer should omit in the visualization, normally words that do not carry significant meaning such as and, the, from and so forth), we are left with a glimpse of the most frequently recurring words in the journal. [Figure 1](#) is a word cloud of the *Revue musicale de Paris* corpus.

Theatre, opera and the piano are clearly at the centre of musical life in nineteenth-century Paris. The 'drastic' qualities of musical sound, to borrow Carolyn Abbate's phrase, come to the fore in terms such as *sons*, *moment*, *brillante*, *effet* and *grâce*, while verbs such as *voir* and *entendre* signal the predominant sensory

¹⁰ Ann Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner and Jeffrey Schapp, eds., *Digital Humanities* (Cambridge: MIT Press, 2012): 105.

¹¹ I am indebted to Matthew L. Jockers's discussion of new ways of gathering evidence and generating knowledge, as described in his chapter on 'Evidence' in his book, *Macroanalysis: Digital Methods & Literary History* (Urbana: University of Illinois Press, 2013): 5–10.

¹² A word cloud is an image consisting of words placed in a pseudo-random position whose relative size belies the underlying frequency of that word in the corpus. For a description of 'Cirrus' (the word cloud function in Voyant), see <https://voyant-tools.org/docs/#!/guide/cirrus>.



Fig. 1 Word cloud of the *Revue musicale de Paris* corpus, <http://bit.ly/ncmr-fig1>.¹³

modes of engaging with musical sound.¹⁴ That terms such as *société* and *public* are markedly larger than *roi* points to the fading social and political influence of the *Ancien Régime*. Results such as *produit*, *francs*, *obtenu* and *prix* betray nineteenth-century consumer culture concomitant with laudatory terms such as *succès*, *nouveau/nouvelle*, *talent*, *célèbre* and *artiste*. The cult of the individual also emerges in words

¹³ Hovering over the word in the interactive visualization reveals the number of times the term occurs in the corpus.

¹⁴ Carolyn Abbate, 'Music—Drastic or Gnostic?', *Critical Inquiry* 30/3 (2004): 505–36.

such as *compositeur*, *directeur*, *seule*, *chef*, *cantatrice*, *pianiste* and *l'auteur*. Only four names enter the word cloud and are all are composers: Beethoven dominates the corpus, appearing no fewer than 10,359 times, a result that may deserve closer examination given the 250th anniversary of the composer's birth; he is followed by Mozart (6,973 results), Meyerbeer (6,552 results) and Rossini (6,360 results). Given the centrality of theatre and opera in Paris as confirmed by this word cloud, the absence of famous performers along with the prominence of 'great composers', who have reigned supreme in histories of music, is, at the very least, provocative.

Broadly speaking, the strength of the word cloud is that it provides a rich array of people, places, objects and ideas that feature prominently in a large corpus and are thus believed to characterize musical life in a particular time and place. A major shortcoming of this visualization lies in that it provides a seemingly static snapshot of music in Paris. The dangers of 'static slices' and 'unconnected dots' in recreating music history are manifold, and recent attempts by scholars such as Ben Piekut to reorient music history in terms of actors and varying kinds of relationships are one way to move beyond the static nature of the word cloud, which is not dissimilar from musicological approaches that reconstruct history by works and composers most often performed.¹⁵ To this end,

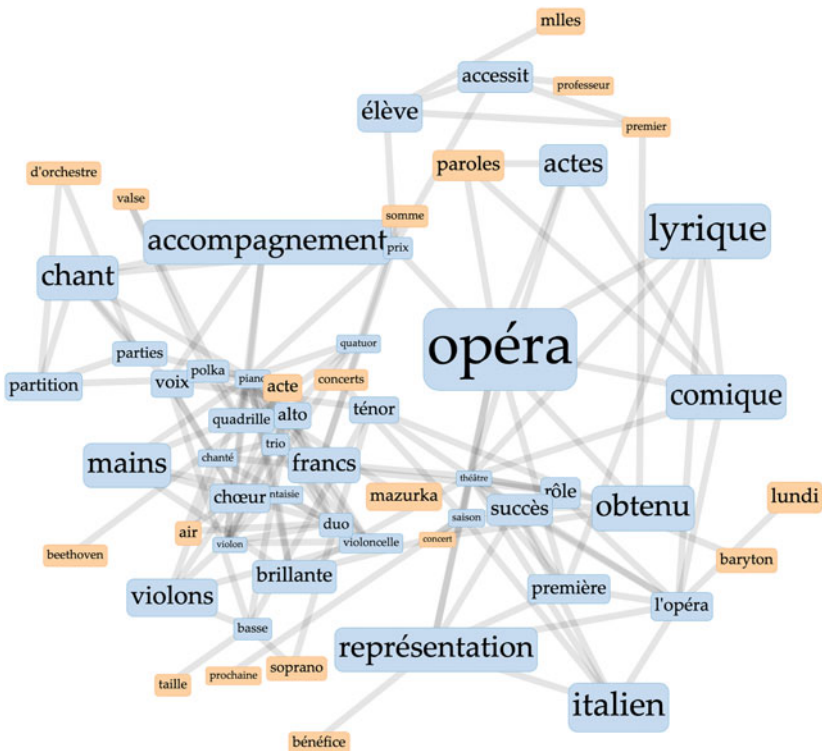


Fig. 2 Network graph of higher frequency terms in the *Revue musicale* corpus, <http://bit.ly/ncmr-fig2>.¹⁶

¹⁵ Benjamin Piekut, 'Actor-Networks in Music History: Clarifications and Critiques', *Twentieth-Century Music* 11/2 (2014): 191–215.

¹⁶ In the interactive visualization, hovering over the word 'lights up' the edges connected to each node, thereby revealing various pathways of connection.

Voyant provides a relatively primitive but nevertheless intriguing 'links' option, allowing the user to convert their word cloud (or a portion thereof) to a network diagram. Figure 2 represents a network graph of higher frequency terms in the *Revue musicale* corpus.

In the context of the diagram, it becomes clear then that *brillante* is connected specifically to *mains*, *piano*, *fantaisie*, *concert* and *théâtre*. *Conservatoire* is connected only to *concerts*, and *société* is also connected only to *concerts*. Thus, what seemed at first glance to be merely a collection of words could be rendered as a host of social, political, artistic and economic threads, intimately connected and driving musical culture in nineteenth-century Paris.¹⁷

Voyant also offers the possibility of graphing the frequency of terms, and here, overlaying terms is most productive so as to produce comparative studies. As seen in Figure 3, the 'Loom' function produces some of the busier frequency visualizations and is best explored interactively as hovering over the line will reveal the term, or the terms are listed alphabetically in the left margin.

In essence, 'Loom' shows frequency of word over time; it allows the scholar to see words whose frequency changes over time (these words can be distinguished from those that whose frequency remains fairly static). For instance, *valse* appears to be part of the 'noise' below, but then spikes and dips dramatically toward the end. *Voix* is among the more frequently mentioned terms, and its frequency varies a great deal, though the overall trajectory across the nineteenth century seems to decline. This is curious, in view of the overall increase of *théâtre*, clearly one of the most frequently cited terms in this journal.

While the 'loom' function might initially seem overwhelming, it does provide a means of exploring a particular corpus, particularly when one knows little about it and when a sense of the major overarching themes coupled with the comparative frequency of their appearance is desired. However, scholars might well have more specific research questions to explore across a global corpus, in which case, a less busy frequency visualization such as the 'trends' function may be preferred. For instance, given recent interest in the physicality of the body in musical performance,¹⁸ one might wish to pit *voix* against *mains*. Figure 4 is a line graph comparing the appearance of these two terms across time.

The human voice is clearly the dominant instrument for nineteenth-century Parisian audiences, yet interest in the voice declines over the decades while interest in the hands increases. Or, returning to the question of canonicity, visualizing trends of when particular composers are mentioned, it useful to see both relative fame, measured by how often music critics mention their names, as well as peaks and troughs of celebrity. Figure 5 is a line graph comparing the relative frequencies of Beethoven, Meyerbeer, Rossini, Berlioz, Bizet and Fauré.

Notably, Beethoven might be the dominant composer in the *Revue et gazette musicale de Paris*, but his fame spikes dramatically towards the end of the nineteenth century. At times, Meyerbeer, Rossini and Berlioz receive more discussion

¹⁷ Voyant's current 'links' function is governed by proximity searches. True network visualization software is supported by a graph database structure, allowing for much more nuanced networks, including directionality in edges, varying types of nodes and so forth. One of the leading graph database platforms and communities is neo4j. For a musicological study on fantasy in this period, see Francesca Brittan, *Music and Fantasy and the Age of Berlioz* (Cambridge: Cambridge University Press, 2017).

¹⁸ See for example, James Q. Davies, *Romantic Anatomies of Performance* (Berkeley: University of California Press, 2014).

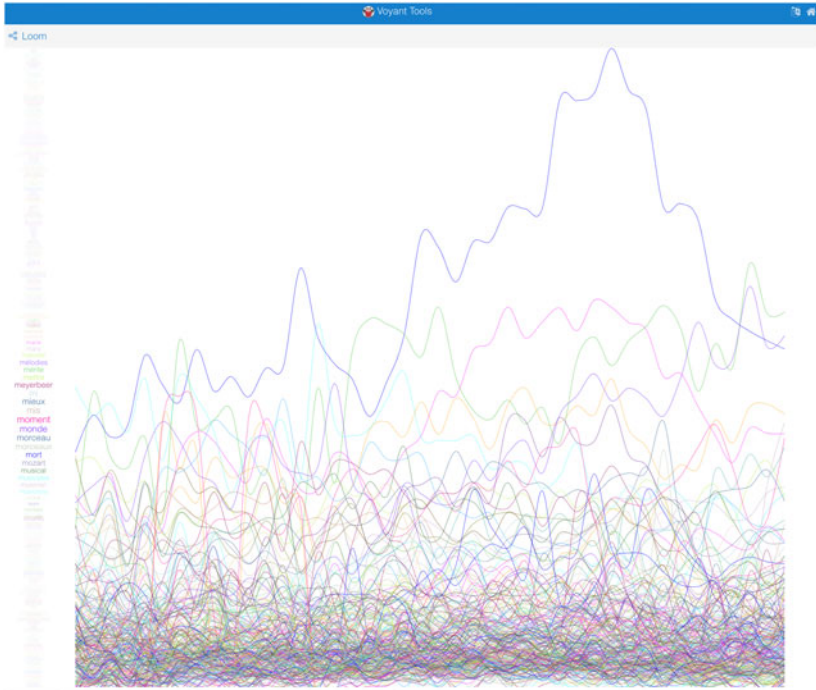


Fig. 3 Frequency of word over time using the 'Loom' function for the *Revue musicale* corpus, <http://bit.ly/ncmr-fig3>.

that Beethoven. Fame is thus dynamic, ever changing in response to an infinite range of factors. Yet even these trends have limitations, most notably, the question of connectedness: what is their fame connected to? It is a published work, a performance, a performer, a particular event, a political moment, or a specific aesthetic of sound? For this, it is necessary to search much purposefully within the archive, which in turn, demands a different set of tools and methods.

Search 2.0

Musicologists have long been interested in studying music in various contexts, and the obvious scholarly technique when it comes to database research is to begin by searching for key terms in music criticism. Computational approaches need not replace this time-tested technique. In fact, they can enhance and extend it. Consider, for example, the term *classique*, a concept central to Christopher Moore's article in this issue, albeit in a period slightly later than this present corpus. The search term *classique* occurs 2,437 times in this corpus. This simple query can of course easily be undertaken in databases such as Gallica, ANNO or RIPM.¹⁹ Text analysis is designed to move beyond search results to display findings in ways that

¹⁹ Gallica is the Bibliothèque Nationale de Paris' digital library. See <https://gallica.bnf.fr>; ANNO is the Austrian National Library's historic journals digital database. See <http://anno>.

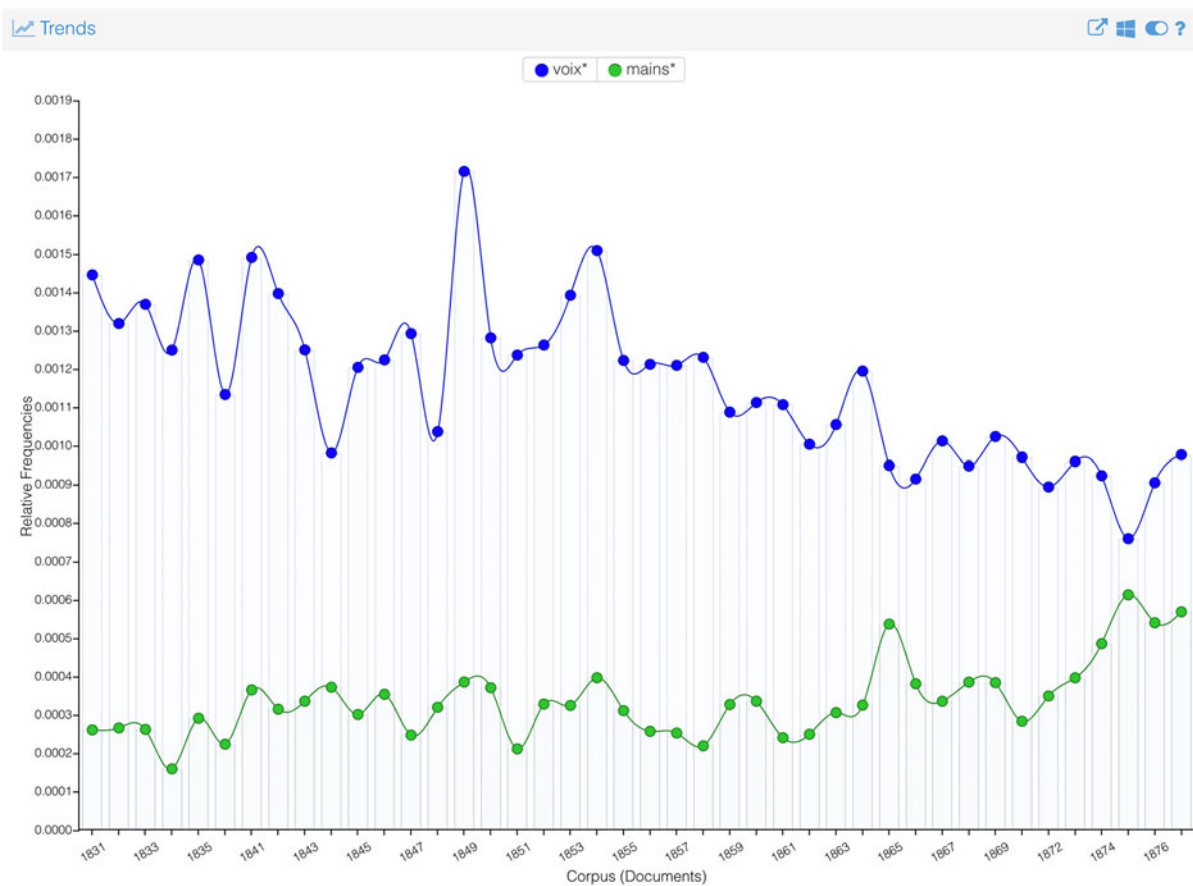


Fig. 4 Line graph comparing *voix* and *mains* in the *Revue musicale* corpus, <http://bit.ly/ncmr-fig4>.

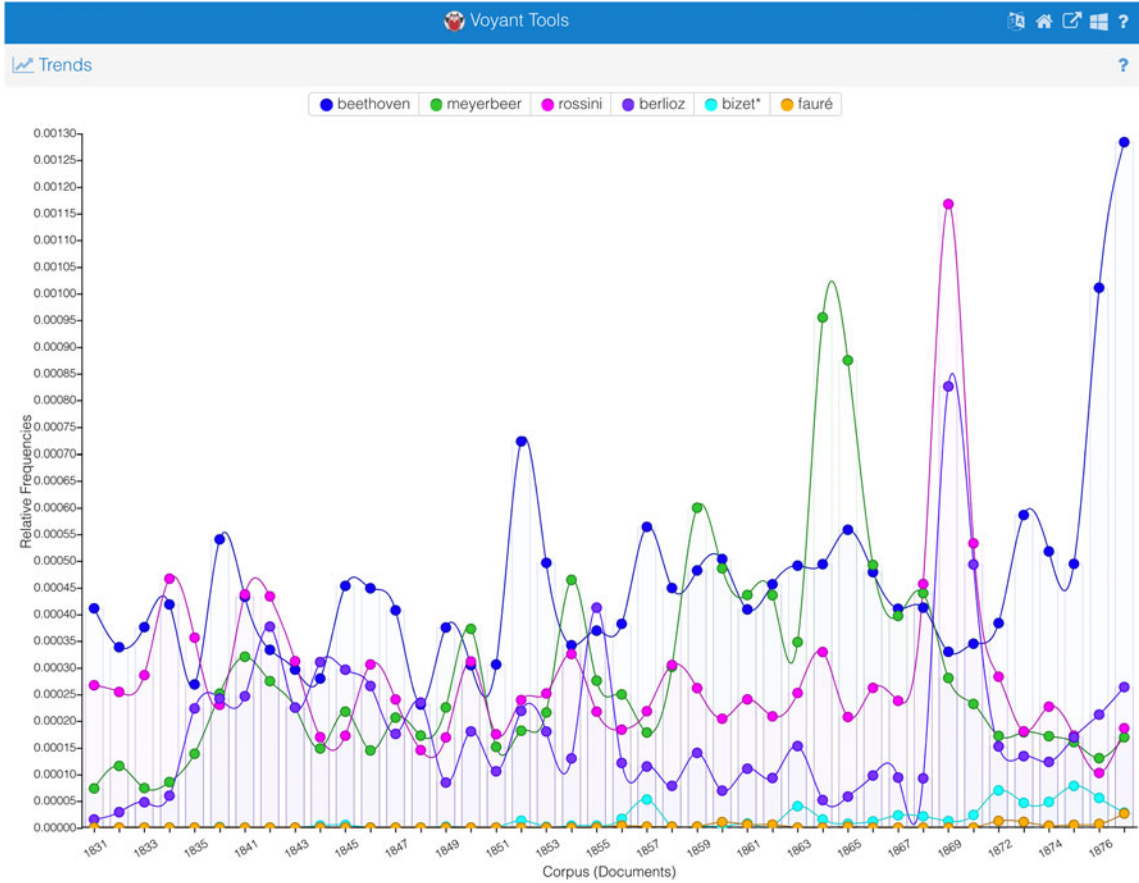


Fig. 5 Line graph comparing the relative frequencies of six composers' names in the *Revue musicale* corpus, <http://bit.ly/ncmr-fig5>.

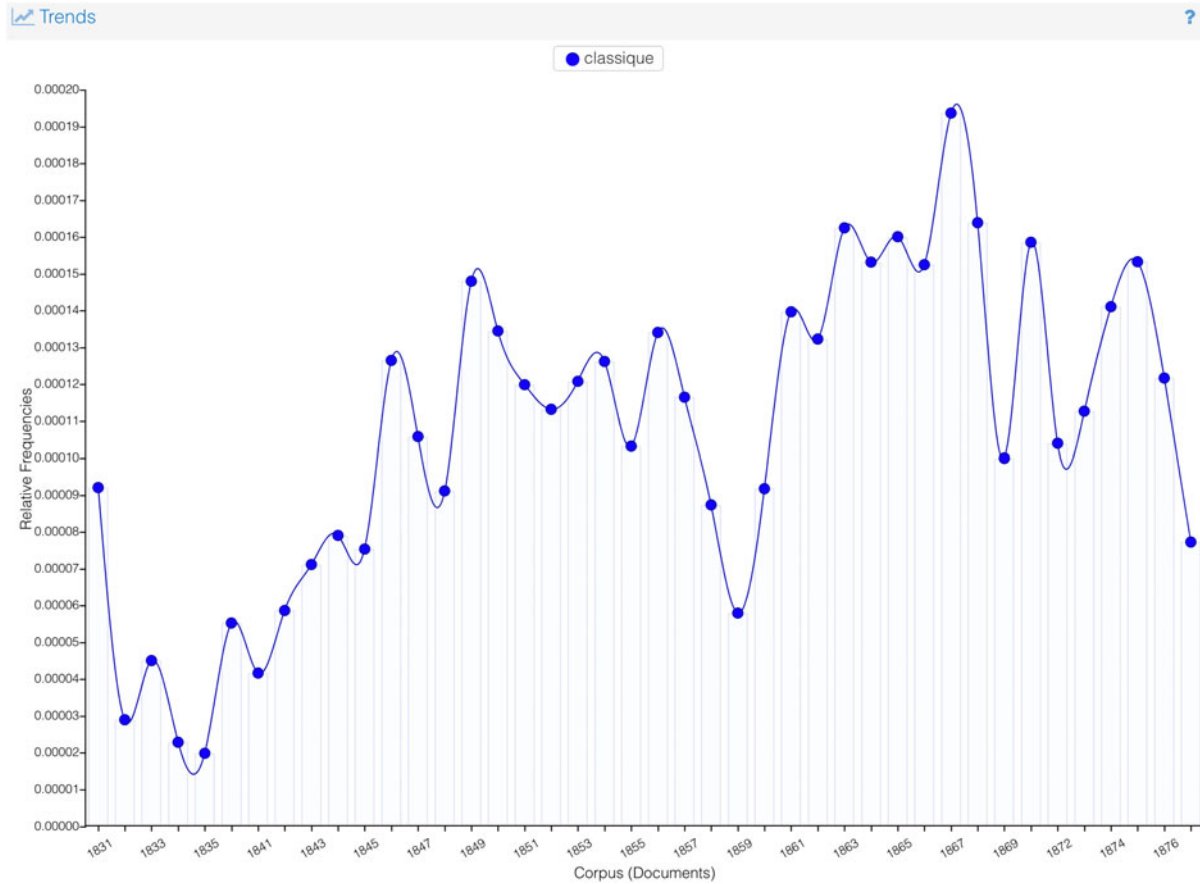


Fig. 6 Line graph illustrating the frequency of the term *classique* in the *Revue musicale* corpus, <http://bit.ly/ncmr-fig6>.

The image shows a screenshot of the Voyant Tools web application. The title bar is blue and contains the text 'Voyant Tools' along with several icons. Below the title bar, there is a header for the current analysis, 'Collocates', with a grid icon on the left and a question mark on the right. The main content is a table with three columns: 'Term', 'Collocate', and 'Count (context)'. The table lists 25 rows of data, each starting with a checkbox in the 'Term' column. The terms are all 'classique', and the collocates are various words. The counts range from 241 to 40.

Term	Collocate	Count (context)
<input type="checkbox"/> classique	concert	241
<input type="checkbox"/> classique	populaire	191
<input type="checkbox"/> classique	concerts	159
<input type="checkbox"/> classique	direction	120
<input type="checkbox"/> classique	populaires	95
<input type="checkbox"/> classique	société	88
<input type="checkbox"/> classique	piano	83
<input type="checkbox"/> classique	séance	80
<input type="checkbox"/> classique	moderne	70
<input type="checkbox"/> classique	répertoire	68
<input type="checkbox"/> classique	aujourd'hui	66
<input type="checkbox"/> classique	pur	65
<input type="checkbox"/> classique	style	62
<input type="checkbox"/> classique	napoléon	55
<input type="checkbox"/> classique	cirque	52
<input type="checkbox"/> classique	séances	49
<input type="checkbox"/> classique	soirée	48
<input type="checkbox"/> classique	mu	47
<input type="checkbox"/> classique	dernier	47
<input type="checkbox"/> classique	religieuse	46
<input type="checkbox"/> classique	volume	45
<input type="checkbox"/> classique	série	45
<input type="checkbox"/> classique	programme	42
<input type="checkbox"/> classique	pianiste	40
<input type="checkbox"/> classique	donnée	40
<input type="checkbox"/> classique	c0ncerts	40

Fig. 7 Table displaying terms most frequently found in proximity to the term *classique*, <http://bit.ly/ncmr-fig7>.

Document	Left	Term	Right
1) 1831	saisir. INSTITUTION ROYALE DE MUSIQUE	classique	, DInsaéz PAR M. cnonon. L'institution
1) 1831	M. cnonon. L'institution de musique	classique	donnera cette an- née six
1) 1831	chargés. ÉCOLE ROYALE DE MUSIQUE	classique	. Exercice extraordinaire pour l'exécution d'une
1) 1831	de l'École royale de musique	classique	ont exécuté cette messe avec
1) 1831	et l'École royale de musique	classique	seraient fondus en un seul
1) 1831	de l'École royale de Musique	classique	, dirigée par M. Choron, aura
1) 1831	français. L'institu- tion de musique	classique	dirigée par M. Choron a
1) 1831	de l'Institution royale de musique	classique	, de ce certain je ne
1) 1831	des Rameaux, l'Institution de musique	classique	dirigée par M. Choron, exécutera
1) 1831	FÉTIs. INSTITUTION ROYALE DE MUSIQUE	classique	, DIRIGÉE PAR M. cHoRoN. Concert
1) 1831	Ne soyez ni romantique ni	classique	d'une manière - exclusive; soyez l'un
1) 1831	l'un et l'autre à propos;	classique	= par vos études, romantique par
1) 1831	inspirations. Si - vous n'étiez que	classique	, vous ne parviendriez pas - à
1) 1831	de l'Institution royale de Musique	classique	dirigée par M. Choron ; mais
1) 1831	n'est l'ombre décharnée de l'école	classique	, qui s'en va toute ridée
1) 1831	peau de l'ombre de l'école	classique	. Et cet imbécille de Mo
1) 1831	a pris le nom de	classique	ou d'académique; en musique, elle
1) 1831	a pris le nom de	classique	ou d'acadé- - mique; en musique
1) 1831	article sous le nom de	classique	, qui s'est avisée de substituer
1) 1831	nature ! En littérature, c'est le	classique	; en politique, c'est le despotisme
1) 1831	se trouvait sur une terre	classique	, il dirigea son attention sur
1) 1831	a pris le nom de	classique	ou d'académique; en musi- - que
1) 1831	du talent si pur, si	classique	de M. Kalkbrenner, c'est-à
1) 1831	élèves du Conservatoire de musique	classique	, par M. Choron, fondateur et
1) 1831	Voy. le Manuel de Chant	classique	et le Manuel de Musique
1) 1831	et savantes qu'a tracées le	classique	, il ne craint point de

Fig. 8 Table showing the word *classique* in context through the *Revue musicale* corpus,²⁰ <http://bit.ly/ncmr-fig8>.

offer greater insight into when and how the term was used across decades of music criticism. One might begin with the trends function for the term in order to get a sense of when the term was used in the journal. Figure 6 is a line graph showing the frequency of the term *classique* in the *Revue musicale* corpus.

As Figure 6 reveals, the term *classique* was most often used in 1867, followed by 1849 and 1836. At first glance, the peaks and troughs do not correspond to any particular performance or political event, except perhaps the Exposition Universelle of 1867.

The 'Collocate' tool reveals words most often found in close proximity to *classique*. Figure 7 is a table showing terms most frequently found with number of occurrences in proximity to the term *classique*.

Glancing at the top ten terms – *concert/s*, *populaire/s*, *direction*, *société*, *piano*, *séance*, *moderne*, *répertoire*, *aujourd'hui*, *pur* – it is evident that the term *classique* was not associated with a timeless canonic work, though it does seem to be closely associated with instrumental music, at least in this corpus. The person's name in closest proximity to the term *classique* is Napoleon; this finding deserves closer

onb.ac.at; RIPM (Répertoire international de la presse musicale) is a database of music periodicals. See <https://ripm.org/index.php>.

²⁰ In the interactive visualization, readers can expand the preceding and following text for each row.

The screenshot shows the Voyant Tools web interface. At the top, there are several browser tabs and the URL 'voyant-tools.org'. Below that, a search bar contains the word 'Napoleon'. The main area displays a list of search results, each with a row of text containing the word 'Napoleon' in context. The results are organized into columns: 'Left', 'Term', and 'Right'. The first result shows the word 'Napoleon' in a sentence about a 'morceau de Rossini'. Other results mention 'Napoleon' in various contexts, such as 'Napoleon et C.', 'Napoleon Bonaparte', and 'Napoleon III'. The interface also includes a search bar at the bottom with the text 'Napoleon' and a search button.

Fig. 9 Results for 'Napoleon' using the 'Contexts' function in Voyant. The results are expanded to reveal a larger portion of the preceding and following text for the term.

examination. Diving deeper into how the term was used, the 'Contexts' function in Voyant may be employed to reveal the text to the left and right of each occurrence of the word. Figure 8 is a table showing the most *classique* in context throughout the corpus.

It is also possible to search for Napoleon in the corpus, and expanding the results allow for a clearer understanding of this connection. Figure 9 shows the results for 'Napoleon' in using the 'Contexts' function with most of the results expanded to reveal a significant portion of the preceding and following text.

It quickly becomes clear that very few results are mentions of the political figure. Instead, our text analysis is catching the place of print: Imprimerie centrale de Napoléon. This illustrates one of the potential pitfalls of quantitative methods: the computer will catch details that humans would likely (correctly) ignore. Nevertheless, there is a valuable lesson here: results are best verified through close contextual analysis, even if it is just glancing to see whether they make sense. Put another way, the veracity of distant reading, in some instances, still benefits from close contextual study.

Yet this is not a reason to downplay the strengths of quantitative methods. As Matthew L. Dockers has remarked, 'massive corpora offer us unprecedented access to the literary record and invite, even demand, a new type of evidence gathering

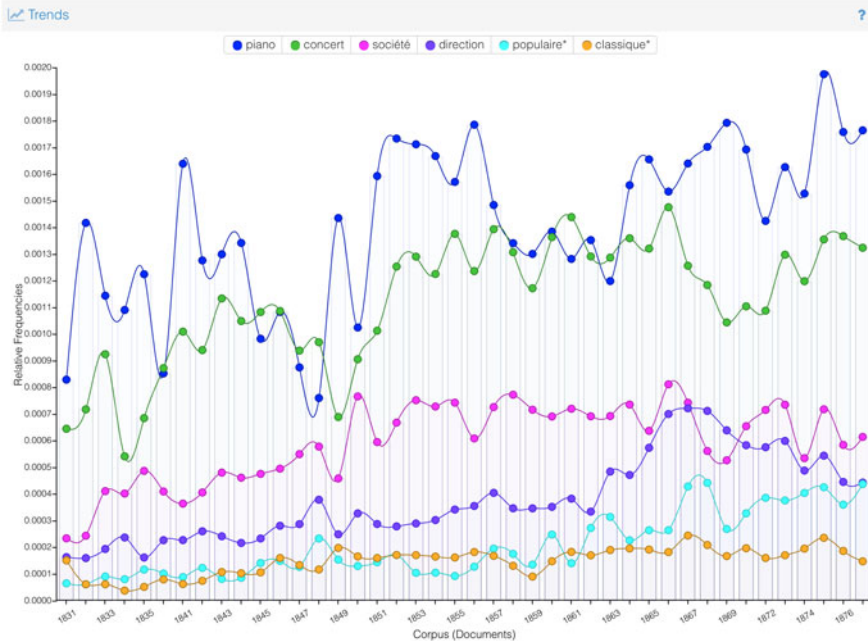


Fig. 10 Line graph showing the relative frequency of the terms *société*, *direction*, *populaire*, *concert* and *piano* in the *Revue musicale* corpus, <http://bit.ly/ncmr-fig10>.

and meaning making'.²¹ He advocates for moving beyond 'anecdotal evidence', often used to support close readings. To be sure, the dangers of selective evidence gathering both by music critics and scholars have already been identified by Ralph Locke in this volume.²² One way to address some of Locke's concerns surrounding bias and innuendo is to employ quantitative methods unlocking new kinds of evidence and by extension, new kinds of interpretive work.

Moreover, as is the case in experimental science, positive and negative evidence are equally valid. For instance, in our quest to more fully understand the use of the term *classique* in this corpus, one might visualize the frequency of the term alongside the top five words identified to be in close proximity: *société*, *direction*, *populaire*, *concert* and *piano*. Figure 10 is a line graph showing the relative frequency of the aforementioned terms.

Immediately evident is the relatively low frequency of change in the appearance of the *classique*. Also, the dramatic outlier in terms of contour is *piano* – this term does not seem to 'track' *classique* in any way. What this means is that even though *piano* occurs in close proximity to *classique*, discussions of the instrument are not concomitant or dependent on the notion of the classic; its appearance in the *Revue musicale de Paris* is unrelated to the piano. While *classique* and *populaire* have more similar trend trajectories, they still crisscross as much as they move in parallel. Thus, although the term 'classic' might have a more similar overall trend trajectory to the idea of the popular (suggesting that these two ideas might

²¹ Dockers, *Macroanalysis*, 8.

²² See Locke, 'How Reliable Are Nineteenth-Century Reviews'.

indeed be co-dependent), they also finish in the inversion of the position in which they began: in 1831 *classique* was used more frequently, but by 1866 the idea of the popular clearly breaks away and becomes more widely used in the *Revue musicale de Paris*. Moretti's aforementioned lament about large datasets comes to mind here: 'the average becomes an inevitable presence—and the average means loss of distinction, slowness, boredom ... Too much polyphony, and too much monotony'.²³ In this instance, one kind of bias (that of a scholar selecting various bits of evidence in support of a particular argument) has been removed, in so far as all the appearances of the term *classique* are weighted equally in the visualizations. In this particular instance we haven't yet found terms that track very closely with *classique* (and readers may try out their own ideas in the search box). But quantitative approaches engage in questions of bias in other ways, most notably, as we shall see, not by eliminating it but rather by bringing it to the fore.

Content, Sentiment, Critical Style

Thus far this investigation has focused on decades of a particular music journal with multiple authorship. One strategy in producing meaningful data visualization is to create a corpus with a constant: a collection of reviews of a particular composer or work, for example. I shall use six 1859 reviews of Félicien David's *Herculanum* as case study.²⁴ Six authors – Berlioz, D'Ortigue, Escudier, Heugel, Smith and Vernes – react to performances of the same work in a very short time span: 6–13 March 1859. As illustrated in Figure 11, the 'Bubblelines' function in Voyant shows the distribution of eight of the most frequently cited words across these six reviews: *scène*, *lyrique*, *choeur*, *Roger*, *Borghi*, *voix*, *compositeur* and *style*. The name of the critic and journal is stipulated on the left-hand side and each term is distinguished by a unique colour. The size of the bubble represents the frequency of appearance for each term. Hovering over each bubble reveals the term and the number of times that it appears at that point in the review.

The term *lyrique* shows the complications of a distant reading as it can refer both to an *oeuvre lyrique*, a musical work for the stage, and it forms part of *théâtre lyrique*, a term for an opera house.²⁵ Voyant will not distinguish between the two, though a sophisticated machine-learning training model would have this capacity. For now,

²³ Moretti, *Distant Reading*, 181. See above at n.5.

²⁴ The *Herculanum* corpus is comprised of E. Vernes, [Herculanum de David], *La France Musicale*, 6 March 1859; Léon Escudier, [Herculanum de David], *La France Musicale*, 13 March 1859; J.L. Heugel, [Herculanum de David], *Le Ménestrel*, 6 March 1859; Paul Smith, [Herculanum de David] *Revue et Gazette Musicale*, 13 March 1859; Hector Berlioz, [Herculanum de David], *Journal des Débats*, 12 March 1859; J. d'Ortigue, [Herculanum de David], *Le Ménestrel*, 13 March 1859. These reviews may be found on [http://bruzanemedia-base.com/fre/Documents/Articles-de-presse/\(searchText\)/herculanum](http://bruzanemedia-base.com/fre/Documents/Articles-de-presse/(searchText)/herculanum) (last accessed 7 March 2020). This website accompanies Gunther Braam, 'La réception d'*Herculanum* dans la presse contemporaine' / 'The Reception of *Herculanum* in the Contemporary Press', in the booklet to Félicien David, *Herculanum*, with Véronique Gens, Edgaras Montvidas and other vocal soloists, Flemish Radio Choir, Brussels Philharmonic, cond. Hervé Niquet, Ediciones Singulares/Palazzetto Bru Zane (Centre de musique romantique française), no. 10 in their series 'Opéra français'. I am indebted to Ralph Locke's article in this issue for drawing my attention to this source. For this corpus I added character names to the stopwords list, as well as the composer and work.

²⁵ I am indebted to Ralph Locke for pointing this out to me.

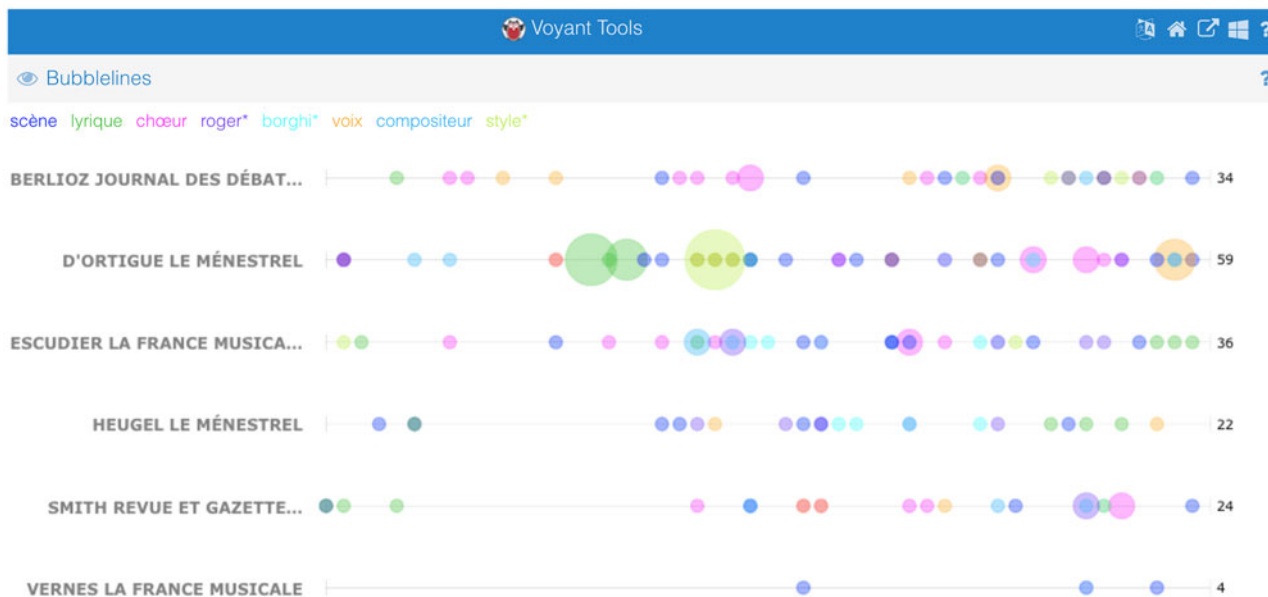


Fig. 11 The 'Bubblelines' function in Voyant displays seven terms found in each of the six reviews, specifying the point in the review at which they appear,²⁶ <http://bit.ly/ncmr-fig11>.

²⁶ Each term is distinguished by a unique colour. The size of the bubble corresponds to the number of times that the term appears at that point in the review.

we have to work with the Voyant results. The term is briefly mentioned either at the beginning of the review (Berlioz, Escudier, Smith) and/or at the end (Escudier, Heugel, Smith). Yet d'Ortigue devotes considerable attention to the term around a third of the way through his review (he mentions it seven times altogether). It is both the scope and placement of the discussion on *lyrique* that sets him apart from the others. Moreover, near the end of his exploration of *lyrique*, he segues into a discussion of the scene, much like Heugel does toward the end of his review. A close reading of all of the reviews suggests that d'Ortigue's focus on the term does indeed concern the lyrical oeuvre of the work, setting him apart from the others, whose use of the term predominantly refers to the opera theatre. In this example, an idiosyncrasy was detected with a distant reading, though a close reading was required to interpret this result.

The presence and function of the chorus is the most frequently mentioned musical feature covered in the six reviews. Yet their distribution and placement are worth exploring further. Berlioz discusses the chorus throughout his review and the most extensive discussion is in the middle. D'Ortigue leaves his discussion of the chorus to the end of the review, though he devotes considerable space to it. The chorus is also discussed alongside the performer, Borghi. Escudier's coverage of the chorus occurs sporadically throughout the first two-thirds of his review; one discussion located around the middle of his review sandwiches comments on the chorus between a discussion about the composer and the singers. Heugel and Vernes do not mention the chorus at all. The relative brevity of Vernes' review means that his does not contain much content, which is reflected in this visualization. Heugel's lack of interest in the chorus, however, is unusual. He seems much more focused on the scene and the performers (Roger and Borghi). His, it would seem, veers toward an analysis of drama as opposed to sound. Overall, visualizations such as this allows one to compare authorial preferences in terms of what critics cover, the relative size of such coverage, and whether or not it stands isolated or is connected to other key musical features that emerged across the entire corpus.

It is worth acknowledging the silences, or white space. These are areas in which the discussion features either unique terms (a comparison to another composer, perhaps), or there may be terms that may not seem to carry much meaning, but in fact could convey irony or sentiment. While this current tool is not capable of advanced natural language processing (NLP), a custom machine-learning algorithm would be able to analyse irony, innuendo, allusion and sentiment (elements that arose in Lesley Wright's article in this issue) in a large corpus (ideally tens of thousands) of music reviews. This would involve training an algorithm to process music-specific terminology in a particular (historic) language.

In the meantime, other approaches lend further insight into the authorial styles and 'signatures' of music critics. Consider, for example, Berlioz's music reviews for the *Journal des Débats*, a treasure trove for scholars in search of a clean corpus to examine a single music critic's aesthetic preferences and style.²⁷ Does Berlioz cover composers differently? Do his reviews follow stylistic patterns depending on genre (i.e. opera versus symphony)? To answer the first question, I built a corpus of Berlioz's Gluck and Beethoven reviews published in the *Journal des Débats*. The 'TermsRadio' function in Voyant exhibits Berlioz's treatment of the

²⁷ These are available on www.hberlioz.com/feuilletons/debatsindex.htm (last accessed 8 March 2020). I'm indebted to Ralph Locke's 'How Reliable Are Nineteenth-Century Reviews' for making me aware of this corpus.

two respective composers over time. Figure 12 illustrates Berlioz' treatment of Gluck's operas over the course of the *Journal des Débats*, while Figure 13 illustrates the composer-critic's treatment of Beethoven's works in the same corpus. In each of the illustrations, the collective 'mini-corpus' of composer-specific criticism is divided into distinct segments, thereby charting the ebbs and flows of terms over time.

Beginning with Figure 12, it becomes evident that, as perhaps is to be expected, the most important elements in Berlioz's assessment of Gluck's operas are the voice, chorus and orchestra. It is also clear that some reviews of *Alceste* are briefly interrupted by a review or two of *Orphée*, returning again to *Alceste*. One of the interesting changes over time is Berlioz's interest in branding Gluck a French composer – the *français* line begins high but declines over time. Moreover, opera is a musico-dramatic genre, and critics tend to lean one way or the other in their criticism. In Berlioz's writing on Gluck's operas, *dramatique* reaches high point around the middle of the 16 reviews, creating an inverse relationship with *l'orchestre* and *instruments*, which begin and end high but slump in the middle. Berlioz's Beethoven reviews, when viewed through the 'TermsRadio' function speak to the prevalence of the critic's attention to style, oeuvre, authorship and the orchestra in covering the German composer (Fig. 13). The most dramatic change in his series of reviews on Beethoven emerges in some *Fidelio* reviews about two-thirds from the end. His focus on 'effects' remained constant in his Beethoven coverage, which is curious, because his interest in the orchestra dips down before it rises around the same time as *Fidelio* gets a lot of coverage. This substantially muddies our understanding of timbral effects during the nineteenth century²⁸ and also suggests that Berlioz's interest in Beethoven's orchestral writing featured prominently in his writings on the composer's only opera. Quantitative studies are well-suited to studies of authorial style, as this is a dimension of the field that is especially well developed by literary scholars. As such, further inquiries into the style of a particular music critic such as Berlioz would yield great dividends, even if it is beyond the scope of this present study.

Geographies of French Music Criticism

Transnational studies of literature lie at the core of Moretti's vision for quantitative methods. In fact, one of the goals of his 2013 *Distant Reading* was the ambitious pursuit of studying 'World Literature'.²⁹ Naturally, once the barrier of human limitations in reading novels has been removed (a lifetime is an insufficient amount of time to read all the literature of the world, not to mention foreign language competence), it then becomes theoretically possible to study world literature. The questions one might imagine include: are there stylistic and syntactical particularities to national literary traditions? How do genres differ, and can computers classify literature into various genres? For music criticism, similar questions arise: can one detect national and transnational trends in musical taste? Do musical genres receive similar critical treatment across national borders? Did criticism surrounding canonic works differ from non-canonic works? Is it possible to map the critical reach of music critics? These (and many more) questions require large

²⁸ See for instance, Emily I. Dolan, *The Orchestral Revolution* (Cambridge: Cambridge University Press, 2013).

²⁹ See chapter 2, 'Conjectures on World Literature' in Moretti's *Distant Reading*, 43–62.

transnational corpora of Open Access content in a consistent and easy-to-process format (text files) and with a consistently high-quality of OCR (Optical Character Recognition). While global corpora of music criticism are not yet ready for this kind of inquiry,³⁰ it is possible to explore some smaller-scale questions related to geographies of nineteenth-century French music criticism.

In her article on Fauré in Boston, Heather de Savage offers a comparative statement on the musical scene in New York as compared to Boston in the early twentieth century: 'while New York was staunchly devoted to Austro-German instrumental works, Wagner's music dramas, and Italian operas (as was the case in other American cities), Boston moved to an aesthetic that would also embrace French music as a complement to the other repertoire'.³¹ There is no reason to dispute this assertion, as it is likely based on years of close reading and an intimate study of musical life in those two major American centres. At the same time, it might be interesting, at the very least, to quantify evidence in support of such statements. The main corpus for reviews in America is the Library of Congress' *Chronicling America*, an Open Access resource which currently contains over 140,000 bibliographic title entries from 1789–1963.³² As is the case with most digital resources, data is continuously added, and the Boston titles are not yet available. Thus, de Savage offered the most accurate and up-to-date information, given the resources available at present.

The kinds of issues that de Savage raises – to what extent Boston embraced French music more than any other American capital, for example – are well suited to quantitative methods, if the data is available. And while we are not able to give much insight into Boston at this time, it is possible to look at the reception of Fauré's music in other American cities and towns. Creating a new corpus of Fauré reviews in America using the *Chronicling America* site reveals some 23 centres that published local newspapers mentioning Gabriel Fauré. Since Voyant's geographic capabilities are still under development, I mapped the findings using Tableau. Figure 14 maps Fauré results by journal and number of results from the *Chronicling America* corpus. Hovering over each hub reveals the name of the journal and number of articles with results for Gabriel Fauré.

Indeed, the French composer's music was performed or at least known from Omaha to South Dakota, Nevada to Salt Lake City, as well as major centres such as Washington DC, San Francisco and of course New York. This shows a 'bird's eye view' perspective of the geographies of Fauré reception in America. But it is also possible to uncover the 'Great Unheard', metaphorically entering into some of the more unusual milieus in which Fauré's music was heard in early twentieth-century America.³³ Having created a sub-corpus of more remote locations with coverage that mentions Fauré (all mention him at most twice), the 'TermsRadio'

³⁰ See my review of RIPM in this issue. RIPM currently does not offer an API (application program interface) through which to access a corpus of texts. Like other commercially driven databases, RIPM has made it difficult to extract content in a programmatic way, thereby rendering bulk analyses such as distant reading impossible.

³¹ See Heather de Savage, 'Under the Gallic Spell': Boston's Embrace of Gabriel Fauré, 1892–1924' in this issue.

³² See <https://chroniclingamerica.loc.gov/about/> (last accessed 8 March 2020).

³³ The term 'Great Unheard' is obviously a play on the 'Great Unread', frequently cited by Moretti but borrowed, as Amir Khadem has noted, from Margaret Cohen. See Amir Khadem, 'Annexing the Unread: a Close Reading of "Distant Reading"', *Neohelicon* 39 (2012): 409–21, here 410. With respect to music, the term 'Great Unheard' could be used to

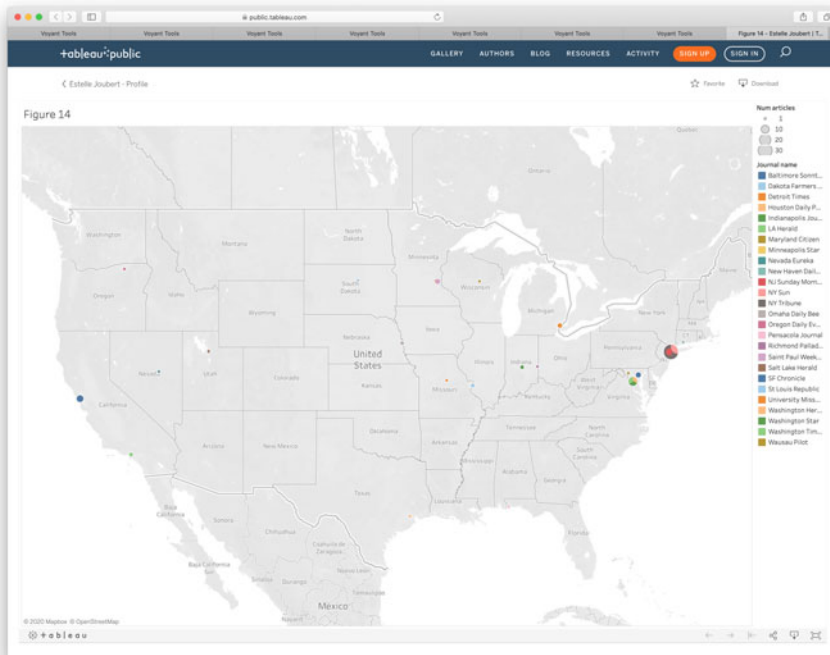


Fig. 14 Map of journal articles with results for Gabriel Fauré in the *Chronicleing America* corpus between 1890 and 1910, <http://bit.ly/ncmr-fig14>.

function in Voyant displays the relative frequency of common words in these issues, this time divided into different segments representing each journal. Figure 15 illustrates the varying local contexts in which the music of Gabriel Fauré was known in America. Journal names, revealing their locations, are on the x-axis.

In effect, one gets a glimpse of the key political, social and artistic issues particular to each location, thereby, in a sense, painting the varied contexts in which Fauré's music was known. Since all the terms visualized appear in all five newspapers, one might argue that these are issues common to America, though varying in importance depending on location. In Houston in 1902, leisure activities such as books, reading, art, song and music are near the top of the curve, underpinned by business, government, county and state. By contrast, crops are near top of mind for the *Dakota Farmers Leader* of 1905 and the *Maryland Citizen* of 1906, a seemingly unlikely context for Fauré's music. Whereas the *Nevada Eureka* of 1906 is focused on stock, companies and time, the *Omaha Daily Bee* of 1907 and 1908 zeros sharply in on the railroad company, though still carving out time for art, song, music and reading.

refer to historic performances whose fleeting sounds are long gone, and/or musical scores, some of which may not have been performed.

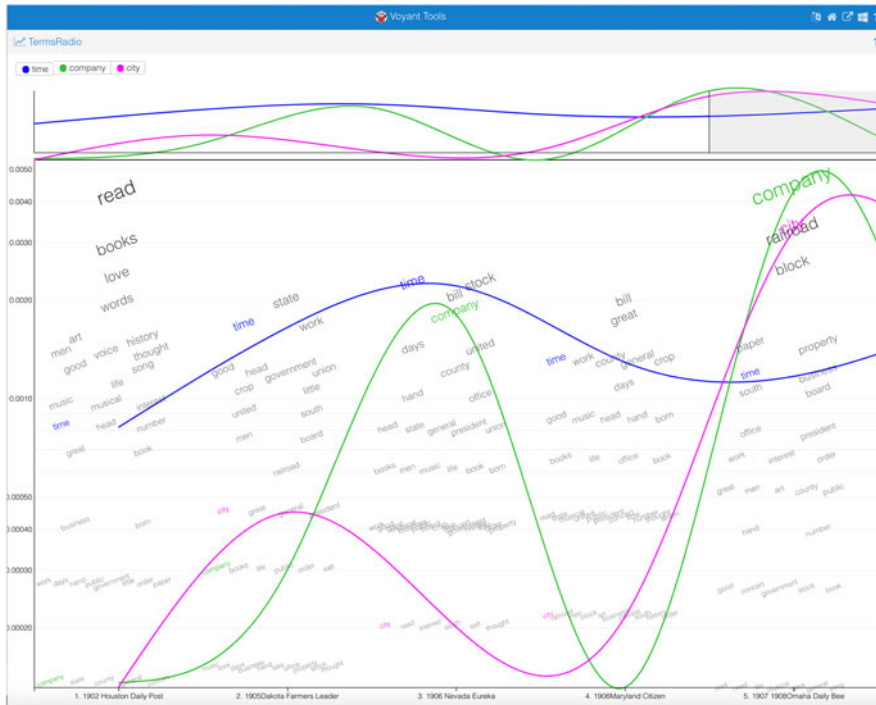


Fig. 15 A comparative visualization of other topics of local interest in journals with reviews on Fauré in America from 1890–1910, <http://bit.ly/ncmr-fig15>.

Concluding Remarks

These experiments with distant reading in nineteenth-century French music criticism represent starting points for the kinds of evidence-gathering and interpretation possible with quantitative methods. One of the undercurrents in this investigation has been one of Moretti's long-standing interests: the canon/archive dichotomy.³⁴ Whereas distant reading has the allure of being able to remove bias when needed, weighing all authors equally (even outside the canon, and notably composers and performers from equity-seeking groups), it is striking that global (that is, corpus-wide) queries did not 'undo' the canon. Rather, it seemed to reinforce it. Discussions of Beethoven, as we might recall, dominate the *Revue musicale de Paris*. This result is sobering, and some reflection seems prudent. It might be that our 'Great Unread' – a corpus of nineteenth-century French music criticism has already filtered the 'Great Unheard' – works composed and performed in the nineteenth century. In other words, while many musical works and performances were left out of the canon for complex reasons, it is perhaps difficult to underestimate the weight of bias in music criticism. The urgent opportunity for future quantitative work in music criticism, I suspect, lies in developing sufficiently nuanced machine-

³⁴ The idea is present already in his *Distant Reading* (pp. 67–71) but forms the title of his more recent collection of pamphlets generated by the Stanford Literary Lab. See Moretti, ed., *Canon/Archive*.

learning algorithms capable of uncovering bias, sentiment, irony, and allusion, amongst other stylistic traits. On the flip side, distant reading did uncover some 'outliers' in Fauré reception in America. This reinforces the fortunate truth that tools can be used precisely for the tasks that scholars require, if they are properly designed (and in the case of machine-learning and AI), trained. Another strong undercurrent pertains to expanding research techniques beyond complex text-searching in digitized documents. The extent to which our discipline has been shaped by this simple and powerful, yet still limited technique, has yet to be fully assessed. Finally, readers might have been struck by the number of times at which this study seemed to push against the limits of large corpus databases and visualization tools. Offering musicologists an Open Access periodical database of music-related reviews from around the world with an API (Application Program Interface) designed from the outset for quantitative studies is a worthwhile collective pursuit. Now that we've entered into the archive with a new set of tools in hand, it might soon be bustling with new discoveries.