

BIENS
SYMBOLIQUES
Revue de sciences sociales
sur les arts, la culture et les idées



A Social Science Journal
on Arts, Culture and Ideas
SYMBOLIC
GOODS



BIENS
SYMBOLIQUES
SYMBOLIC
GOODS



PRESSES
UNIVERSITAIRES
DE VINCENNES



n°2 / 2018

Arpenter la vie littéraire *Surveying Literary Life*

Arpenter la vie littéraire

Les bases de données dans les études sociohistoriques sur l'imprimé et la littérature

Surveying Literary Life

The Use of Databases in Sociohistorical Studies of Print and Literature*

Claire Ducournau | Anthony Glinoyer

traduction | translation
Katharine Throssell



Figure 1. Meuble de classement dans une bibliothèque
Homestead1997.
Licence CC 0 Creative Commons (source : Pixabay)

Figure 1. Filing cabinet in a library
Homestead1997.
Licence CC 0 Creative Commons (source : Pixabay).

* Where no official English translation is available, translations have been carried out by BS/SG.

Introduction

À l'heure où les *big data* (ou données massives), les bases qui les accueillent et les programmes qui en permettent l'extraction prolifèrent, ce deuxième dossier de la revue *Biens symboliques / Symbolic Goods* propose de s'arrêter sur la conception, la conservation et les usages des bases de données conçues par ou pour les chercheur·se·s sur l'imprimé, la littérature et leur histoire. Qu'elles portent sur des mots, des références, des individus, des lieux, des événements, ou encore des objets matériels, les bases de données ont toujours existé dans la recherche en sciences humaines et sociales. Toutefois, le développement de logiciels spécialisés a permis de simplifier leur création et leur utilisation ainsi que leur croisement, tandis que, dans un nombre croissant de cas, l'Internet en a ouvert la consultation à qui veut bien s'y plonger.

Appelons ici bases de données des présentations structurées, interrogeables et significatives d'informations préalablement sélectionnées, catégorisées et harmonisées en vue d'un usage scientifique. Encore faut-il distinguer entre deux types de bases de données dans les sciences humaines et sociales : d'une part, la constitution d'un répertoire ou d'un catalogue d'informations, disponible à la consultation et à l'interrogation ; d'autre part, la structuration de ces informations en préparation à leur traitement statistique (du plus simple au plus complexe). La première possibilité a pu offrir aux spécialistes de littérature des matériaux, par exemple pour des expositions, des colloques, des livres, à l'instar de bases qui mettent en série des informations sur des parutions ou des auteurs, comme [l'Index Translationum de l'Unesco](#) qui a inventorié des titres traduits, ou [Weblitaf](#) qui recense les titres de littérature africaine francophone parus. Mais seul le second type de base implique des raisonnements quantitatifs et permet des représentations graphiques originales

Introduction

At a time when we are seeing the proliferation of big data, of the databases that store it, and of the programs that allow its extraction, this second dossier of *Biens symboliques / Symbolic Goods* suggests we pause to reflect on the conception, maintenance, and use of databases designed for or by researchers who study print, literature, and their history. Whether they cover words, references, individuals, places, events, or material objects, databases have always existed in social science research and in the humanities. However, for the last decades, the development of specialized computer programs has made it easier to create, use, and combine them. Moreover, in an increasing number of cases, the internet has meant that databases can now be consulted by whomsoever might wish to delve into them.

Databases are defined here as structured, consultable, and meaningful presentations of information that has been previously selected, categorized, and harmonized in preparation for scientific use. It is important to distinguish between two types of database in the humanities and social science: on one hand, the construction of a repertoire or catalogue of information available for consultation and investigation; and on the other hand, the structuring of this information so that it might be used for statistical processing (from the simplest to the most complex). The first provides literary specialists with materials, for example for exhibitions, conferences, or books, like, for instance, databases that provide information on publications or authors, such as [Unesco's Index Translationum](#), which provides an index of translated works, or [Weblitaf](#), which has established an inventory of Francophone African literature. Nevertheless, only the second type of database incorporates quantitative reasoning and allows for original graphic



des données recueillies, moyennant les méthodes et les outils adaptés. Inversement, les approches quantitatives, qui ont permis de mieux comprendre les processus de production et de réception de la littérature (pour un bilan, voir Sapiro 2010) ne nécessitent pas toujours la production de bases de données, qu'il s'agisse par exemple de compter des items dispersés ou de manier des chiffres déjà existants (produits, par exemple, par des professionnel·le·s). Force est de constater cependant que ces approches quantitatives, y compris bibliométriques, mobilisant souvent des bases, ont été plus souvent le fait d'historien·ne·s et de sociologues, ou de linguistes, que de littéraires (Viala 1985, Vaillant 1992, Genet & Lafon 2003).

Dans les études littéraires proprement dites, la place des bases de données et des réflexions méthodologiques qui les accompagnent est au contraire longtemps restée très limitée. La raison principale de ce désintérêt est à chercher dans l'espace des études littéraires lui-même. Là, comme longtemps en histoire de l'art (Joyeux-Prunel 2010), l'étude (voire la simple conception) de bases de données a subi le discrédit symbolique dont sont souvent victimes les outils de recherche et les travaux quantitatifs. C'est aussi que le texte, placé au cœur des approches littéraires traditionnelles puis de la nouvelle critique des années 1960-1980, est une réalité analogique, qui impose une lecture continue, elle-même variable en fonction de la subjectivité, des outils herméneutiques et des caractéristiques sociales de celui ou de celle qui s'y adonne : son interprétation et sa compréhension ne vont pas traditionnellement de pair avec une retraduction numérique, qui l'encoderait sous forme de chiffres pour le rendre comparable à d'autres corpus. Les analyses textuelles et le textualisme, dominants dans les études littéraires à l'échelle mondiale depuis l'après-guerre jusqu'à la fin du xx^e siècle, se sont généralement tenus en retrait de l'usage d'outils statistiques et même du recours à de vastes corpus, préférant concentrer leurs analyses sur des textes jugés majeurs ou représentatifs.

representations of the data collected, using specifically adapted methods and tools. Conversely, the quantitative approaches that have enabled us to better understand the process of production and reception of literature (for an overview, see Sapiro 2010) do not always require the production of databases, whether they involve counting diffuse items or manipulating already-existing figures (produced by professionals, for example). We note, however, that these quantitative approaches – including bibliometric ones, often drawing on databases – have more frequently been the work of historians and sociologists, or linguists, than literature specialists (Viala 1985; Vaillant 1992; Genet & Lafon 2003).

In literary studies more specifically, though, the role of databases and the methodological reflections associated with them were for a long time limited. The main reason for this lack of interest can be found in the realm of literary studies itself. As was long the case in art history (Joyeux-Prunel 2010), the study (or even simply the design) of databases was subject to the symbolic discredit to which quantitative approaches and their research tools often fall victim. This is also because the text, which was at the heart of traditional approaches to literature, and of French "New Criticism" in the 1960s-1980s, is an analogous reality, requiring a continuous reading that varies according to the subjectivity, hermeneutic tools, and social characteristics of the reader. Its interpretation and its comprehension are not traditionally associated with digitalization – with being encoded and rendered comparable with other corpora. Textual analysis and textualism, which were dominant in literary studies worldwide between the post-war period and the late twentieth century, generally held back from using statistical tools or even large corpora, and focused instead on texts that were considered significant or representative.

Si les travaux sociohistoriques sont venus relativement tard à l'informatisation des bases de données et ont longtemps accordé la préférence aux fiches manuscrites, les linguistes et les lexicologues ont au contraire, dès les années 1950, fait appel aux ordinateurs pour procéder au traitement statistique de leurs corpus de phonèmes, mots, syntagmes ou membres de phrases, dont ils ont pu ainsi analyser les concordances (classement par ordre alphabétique des mots, des expressions, des thèmes, etc., présents dans un texte). Coïncée entre linguistique et littérature, la statistique textuelle est cependant restée, jusqu'à récemment, une branche assez isolée des études littéraires (Delcourt 1995 ; Morin, Bosc, Hebrail, Lebart 2002).

De nombreuses recherches en histoire du livre et en sociologie du littéraire se sont aussi parallèlement converties à la collecte et à l'organisation de données (biographiques, bibliographiques, contextuelles, textuelles). La première, telle qu'elle a été renouvelée par Roger Chartier et Robert Darnton¹, a notamment pris en compte les supports matériels pour mieux étudier la production et la réception des textes. La construction de biographies collectives, depuis les travaux pionniers de Rémy Ponton (1977), Christophe Charle (1979) ou Alain Viala (1985), a permis de porter au jour des espaces de positions et d'interrelations, y compris en temps de crise (Sapiro 1999), de comparer des ressources sociales individuelles et collectives, de révéler l'existence de stratégies, d'analyser la hiérarchie sociale des genres littéraires, en déplaçant bien des questionnements usuels de l'histoire littéraire. D'autres travaux plus récents s'appuyant sur de telles bases de données prosopographiques se sont davantage intéressés au

1 Qui a mis certains de ses matériaux de travail sur le monde du livre dans la France pré-révolutionnaire en libre accès sur [un site](#).

Sociohistorical studies came relatively late to the digitalization of databases, and have long preferred handwritten forms. Linguists and lexicologists, however, turned to computers as early as the 1950s to conduct statistical analyses of their corpora of phonemes, words, phrases, or portions of phrases, and thus to study overall coherence (by putting the texts' words, expressions, themes and so forth, into alphabetical order for example). Caught between linguistics and literature, textual statistics remained a relatively isolated branch of literary studies until quite recently (Delcourt 1995; Morin, Bosc, Hebrail, Lebart 2002).

In parallel, many studies in the history of the book and the sociology of literature have also turned to the collection and organization of data (biographical, bibliographical, contextual, or textual). After Roger Chartier and Robert Darnton's¹ innovative researches, the renewed history of the book has, for example, taken the material aspects of printed texts into account to have a better understanding of their production and reception. The construction of collective biographies – as in the pioneering works by Rémy Ponton (1977), Christophe Charle (1979), and Alain Viala (1985) – shed new light on the spaces of positions and interrelations, even in times of crisis (Sapiro 1999). These studies also made possible the comparison of individual and collective social resources; revealed the existence of strategies; and analysed the social hierarchy of literary genres, all while shaking up the typical investigations of literary history. Other more recent studies based on this kind of prosopographical database have shown interest in the role of cultural intermediaries

1 Robert Darnton has made some of his research material on the world of books in pre-revolutionary France freely accessible [online](#).



rôle des intermédiaires culturels et aux mécanismes d'accès à la consécration littéraire (Dubois 2008, Dozo 2011, Ducournau 2017).

Comme l'histoire de l'édition², celle de la presse³ a grandement profité des travaux qui se sont adossés sur des bases de données pour décrire la vie littéraire à partir du traitement de larges corpus médiatiques ou de *realia* comme les catalogues des éditeurs et des bibliothécaires, ce qui a permis de renouveler des traditions historiographiques qui étaient avant tout monographiques. Des initiatives récentes se proposent d'étudier les pratiques de lecture à partir de bases répertoriant les traces textuelles de scènes de réception⁴ ou encore [les réseaux de correspondances entre les hommes \(et femmes\) de lettres des Lumières](#).

L'affirmation de ce type de travaux, marqués par des collaborations accrues avec des littéraires, souvent situés dans des franges favorables aux sciences sociales, le recul du *théorisme* littéraire et des débats sur la théorie littéraire sur les campus américains et européens, l'éclatement des objets de recherche, l'accès plus aisé aux ordinateurs, aux outils numériques et à l'Internet, la tendance

-
- 2 La somme de *L'Histoire de l'édition littéraire au Québec au xx^e siècle* (Michon 1999, 2004, 2010) s'est ainsi largement appuyée sur des bases de données mises en place à l'université de Sherbrooke, comme le restitue, dans les présentations de bases de données jointes à ce dossier, le témoignage de Josée Vincent.
 - 3 Voir les bases de données proposées sur le site du réseau international [Transfopress](#), consacré à l'étude de la presse en langue étrangère, et sur la plateforme [Médias 19](#) créée en 2011 sous la direction de Guillaume Pinson et de Marie-Ève Thérénty, consacrée à l'étude de la culture médiatique au xix^e siècle, parallèlement à la publication d'un ouvrage scientifique décisif sur cette question (Kalifa, Régnier, Thérénty, Vaillant 2011) .
 - 4 Voir le site [Reading Experience Database, 1450-1945](#), développé en accès ouvert au Royaume-Uni, à destination des enseignant•e•s et des chercheur•se•s.

and means of obtaining literary acclaim (Dubois 2008; Dozo 2011; Ducournau 2017).

Like the history of publishing,² the history of newspapers³ has broadly benefited from studies relying on databases to describe literary life through the processing of large-scale press corpora, or *realia* such as publishers' and librarians' catalogues: this contributed to the renewal of historiographic traditions which were for the most part monographic. Similarly, recent works have studied reading practices using databases of either textual records of the experience of reading⁴ or [correspondence networks between men \(and women\) from the Enlightenment period](#).

Things changed in this area from the early twenty-first century. The shift can be partly explained by factors such as the emergence of studies involving increased collaboration between sociologists, historians, and literary specialists often situated on the fringes of the field and open to social science; the decline debate on literary theory on American and European campuses; the diversification of research objects; easier access to computers, digital tools, and the

-
- 2 The significant *Histoire de l'édition littéraire au Québec au xx^e siècle* (Michon 1999, 2004, 2010) is, for instance, essentially founded on databases put into place at the University of Sherbrooke, as we can see in Josée Vincent's contribution to this dossier.
 - 3 See the databases available on the website of the international network [Transfopress](#), dedicated to the study of newspapers in various languages, or on the [Media 19 platform](#) created in 2011 under the supervision of Guillaume Pinson and Marie-Ève Thérénty, and dedicated to the study of media culture in the nineteenth century, beside the publication of a key scientific book on this issue (Kalifa, Régnier, Thérénty, Vaillant 2011) .
 - 4 See the open access website [Reading Experience Database, 1450-1945](#), developed in the United Kingdom for teachers and researchers.

au sein des organismes subventionnaires à privilégier les projets pluridisciplinaires, comportant un volet quantitatif et un volet numérique, sont autant de facteurs qui peuvent contribuer à expliquer que depuis le début du *xxi*^e siècle, les lignes aient bougé. Dans les sciences humaines et sociales plus généralement, la multiplication des données numériques, rendues plus accessibles, a conduit à des réorganisations disciplinaires, favorisant les partenariats avec les sciences de l'informatique (Ollion & Boelaert 2015). Depuis une décennie a ainsi émergé et s'est institutionnalisé le champ dit des « humanités numériques », entendues comme une nouvelle épistémologie pour l'étude des cultures patrimoniales écrites, du fait des transformations que les outils et les méthodes informatiques imposent aux savoirs en sciences humaines. Leur rapide progrès a conduit certains à les célébrer comme une véritable révolution scientifique, d'autres à les dénoncer, dans le contexte anglo-américain tout du moins, comme une étiquette à la mode susceptible de nourrir le conservatisme des institutions universitaires, en allant par exemple à l'encontre des interprétations critiques des textes littéraires, qu'elles soient marxistes, féministes ou postcoloniales (Allington, Brouillette, Golumbia 2016). Dans le domaine des études littéraires francophones, pour mentionner quelques initiatives qui se revendiquent des humanités numériques, sont ainsi apparus le Labex [Observatoire de la vie littéraire \(OBVIL\)](#) en France, dont le site offre des textes numérisés et transcrits, à côté de catalogues de fonds d'archives, le laboratoire [Ex situ, littérature et technologie](#) au Québec, ou la [Chaire de recherche du Canada sur les arts et les littératures numériques](#). L'une des plus abouties et la mieux connue de ces entreprises collectives est celle du [Literary Lab](#) fondé par Franco Moretti à l'université de Stanford (Moretti 2016). Les travaux qui en sont issus, publiés sous forme de plaquettes disponibles en accès libre, reposent sur une méthode d'analyse computationnelle qui implique d'importants corpus (en particulier sur le roman du *xix*^e siècle européen), des analyses statistiques et des représentations graphiques (arbres, cartes, graphes, etc.). Ce dynamisme a fait des émules, comme le [Bodmer Lab](#) à l'université de Genève.

internet; as well as the tendency of funding bodies to prioritize pluridisciplinary projects with quantitative and digital aspects. In social science and humanities more broadly, increases in (accessibility to) digital data have led to disciplinary reorganizations promoting partnerships with information science (Ollion & Boelaert 2015). Over the last decade, we have therefore seen the emergence and institutionalization of the field known as “digital humanities,” which is perceived as a new epistemology for the study of written cultural heritage, resulting from the transformations that digital tools and methods have imposed on knowledge in human science. With its rapid progress, this field was praised by some commentators as a genuine scientific revolution, while others have condemned it (in the Anglo-American context at least) as merely a fashionable label likely to encourage conservatism in university institutions by going, for instance, against the critical – whether Marxist, feminist, or postcolonial – analysis of literary texts (Allington, Brouillette, Golumbia 2016). To mention a few initiatives that claim the digital humanities label in Francophone literary studies, we have seen the emergence of the [Observatoire de la vie littéraire \(OBVIL\)](#) in France, whose website provides digital texts and transcripts alongside catalogues and archival sources. Quebec has the [Ex situ, Littérature et Technologie au Québec](#) laboratory, or the [Canada Research Chair in Digital Art and Literature \(ALN\)](#). One of the most well-known and comprehensive of these collective undertakings is that of the [Literary Lab](#) founded by Franco Moretti at the University of Stanford (Moretti 2016). The studies that have resulted from this, published as brochures available online through open access, are based on computational analysis that involves significant corpora (on nineteenth-century European novel particularly), statistical analysis, and graphics (maps, graphs, mind maps). This dynamism has sparked similar projects, such as the [Bodmer Lab](#) at the University of Geneva.



Au sein de ce courant, les études formalistes restent cependant les plus présentes, et s'étendent parfois du texte à ses avant- et après-textes. Les résultats qui en découlent, du point de vue de l'interprétation textuelle, se situent tantôt dans une continuité avec les approches traditionnelles, tantôt dans une rupture revendiquée (Alexandre 2016). Le monopole de l'autorité toute personnelle des critiques littéraires s'y efface en effet en partie devant la possibilité de faire intervenir d'autres médiateurs ou médiatrices, comme les informaticien·ne·s, au prix d'une transformation des représentations classiques du texte, devenu terrain d'expérimentation pour mettre à l'épreuve des hypothèses. La mise à disposition d'une quantité très importante de données permet par exemple de traiter également le corpus littéraire canonique et des textes méconnus, sans avaliser directement les différences de légitimité existant entre eux. Le *distant reading*, défendu par Franco Moretti, tente précisément de comparer les caractéristiques formelles de différents types de textes (le style, les régularités phonétiques et sémantiques), ce qui peut aplanir des hiérarchies littéraires pour mettre en évidence des proximités parfois étonnantes quant au lexique (Moretti 2016), quant aux motifs et quant aux structures narratives. Faire ainsi de la littérature un « discours parmi d'autres » (Gefen 2015 : 7) soumis à des protocoles de mesure quantitative à partir des seules données textuelles ne devrait cependant pas faire oublier les rapports de pouvoir et les hiérarchies sociales qui ont historiquement prévalu dans la vie littéraire (comme ailleurs), et qui déterminent en partie ces réalités textuelles, comme le montrent les travaux d'histoire sociale du littéraire. Ce tournant dit computationnel n'annule évidemment pas non plus la nécessité de l'interprétation patiente des résultats mathématiques ou graphiques obtenus, et, partant, le recours aux méthodes traditionnelles d'interprétation des textes, comprenant des approches de détail, le dépouillement de sources primaires et les efforts de contextualisation.

Within this approach, formalist studies are the most prevalent, however, and sometimes spread from the text itself to its *avant-textes*, and subsequent editions. The results of this, from the perspective of textual interpretation, are sometimes situated within the continuity of traditional approaches, and sometimes in a clear and intentional break with them (Alexandre 2016). The monopoly of the very personal authority of literary critics has indeed waned, partly due to the possibility of calling on other mediators such as computer specialists, with the result of transforming classical representations of text, which has become a field for experimentation and hypothesis-testing. The availability of very substantial amounts of data allows to work with less well-known texts as well as texts of the literary canon without directly endorsing the differences in legitimacy that exist between them. *Distant Reading*, as defended by Franco Moretti, specifically tries to compare the formal characteristics (style, phonetic, and semantic regularities) of different types of text, which can level literary hierarchies, in order to reveal sometimes surprising proximities in terms of vocabulary (Moretti 2016), motifs, and narrative structures. Making thus literature “one discourse among others” (Gefen 2015: 7) and subject to quantitative protocols using only textual data, should not obscure the power relations and social hierarchies that have historically prevailed in literary life (as elsewhere), and which at least partly determine these textual realities, as we can see in sociohistorical studies of literature. Of course, this so-called computational turn should not either expunge the necessity for patient interpretation of the mathematical and graphic results obtained, through the use of traditional interpretive methods, including attention to detail, assessment of primary sources, and attempts at contextualization.

À l'heure des premiers bilans, force est ainsi de constater que l'accessibilité et la masse croissante des données numériques rendent d'autant plus nécessaire l'exercice de la vigilance méthodologique et épistémologique (Ollion & Boelaert 2015). La production de ces données, ni neutre ni transparente, repose souvent en effet sur « l'invisibilisation et [le] masquage des opérations dont elles dépendent et des petites mains qui les réalisent » (Jaton & Vinck 2016 : 496), sur des contraintes et sur des « biais considérables » (Gefen 2015 : 3-4), qui méritent d'être étudiés et restitués de manière réflexive. Malgré quelques travaux récents (Bernier & Couturier 2007 ; Vesna 2007 ; Flichy & Parasie 2013 ; Jaton & Vinck 2016), les bases de données restent en effet peu interrogées en tant que pratiques de recherche et dans les démarches méthodologiques concrètes qu'elles imposent, en particulier dans les études littéraires où elles viennent nourrir davantage de travaux et de problématiques.

La restitution des résultats dans les « humanités numériques » y voisine pourtant volontiers avec celle des cheminements suivis, sur le mode plus narratif du « journal de bord » comme le formule Franco Moretti (2016 : 11-12). La publication des tomes de *La Vie littéraire au Québec*, fondée sur des bases de données présentées dans ce dossier par Marie-Frédérique Desbiens et Chantal Savoie, est ainsi abordée dans ce qui l'a rendue possible, dans ses évolutions et son élaboration collective. Comme dans les recherches de Franco Moretti prônant « l'équipe de cinq à six chercheurs » (2016 : 8), la nécessité de la discussion et de l'argumentation à plusieurs s'y accompagne d'une division du travail permettant d'expérimenter plusieurs directions, ce que ne pourrait faire un·e chercheur·se isolé·e.

Taking stock of the first results, we are forced to observe that increased accessibility and quantity of digital data have made the need for methodological and epistemological vigilance all the more pressing (Ollion & Boelaert 2015). The production of this data – a process which is neither neutral nor transparent – is indeed often based on the “invisibility and obscuring of the operations upon which they depend, and of those who conduct them” (Jaton & Vinck 2016: 496), and on constraints and “considerable biases” (Gefen 2015: 3-4) which deserve to be studied and rendered reflexively. But, with the exception of some recent studies (Bernier & Couturier 2007; Vesna 2007; Flichy & Parasie 2013; Jaton & Vinck 2016), databases remain relatively free from investigation as a research practice, and in the concrete methodological approach that they impose, particularly in literary studies where they are mobilised in an increasing number of studies.

However, the presentation of results in the “digital humanities” often also includes the exposition of the different stages that have led to these findings in a more narrative mode, as in the “log-book” proposed by Moretti (2016: 11-12). The publication of the volumes of *La Vie littéraire au Québec*, based on data presented in this dossier by Marie-Frédérique Desbiens and Chantal Savoie, is thus presented from the perspective of what made it possible, in its collective development and evolutions. As in Moretti's research, which advocates “a team of five to six researchers” (2016: 8), the need for collective discussion and argumentation is accompanied by a division of labour that allows several directions to be pursued at once, which would be impossible for a researcher working alone.



Ce dossier veut donc contribuer à rompre avec l'idée de rupture qui nourrit certains débats actuels, en insistant sur la manière dont les études sociohistoriques sur l'imprimé et le littéraire ont de plus longue date construit et utilisé des bases de données. Ce geste va en effet souvent de pair avec le recours à des concepts sociologiques tels que le champ et le réseau, mobilisant (et, parfois, revisitant) certaines approches théoriques. On le voit ici à l'œuvre dans la contribution de Marie-Frédérique Desbiens et Chantal Savoie, sur les méthodes de recherche permettant d'évaluer l'autonomisation d'un champ littéraire au Québec. C'est aussi le cas dans l'article de Florence Bonifay, qui questionne les conditions d'application de ce même concept dans l'histoire littéraire. À travers une déconstruction rigoureuse d'un certain nombre de représentations communes attachées au groupe abusivement dit de la « Pléiade », celui-ci fait ressortir les tensions qui traversent la vie poétique française dans la seconde moitié du xv^e siècle, ainsi que l'importance des sociabilités et des échanges dans la construction sociale des carrières individuelles, mesurés ici de manière originale à partir d'un corpus de textes publiés.

Pour les études sociohistoriques sur l'imprimé et la littérature, de tels recours à des bases de données répondent, en conséquence, à des enjeux à la fois techniques, et, surtout, scientifiques aussi décisifs que l'administration de la preuve ou la définition des objets et des problématiques de recherche (que l'utilisation de bases de données contraint justement, dans bien des cas, à expliciter). En ouvrant sur les coulisses de leur utilisation, ce dossier explore, dès lors, comment ces données numériques, devenues essentielles dans bien des travaux, permettent d'arpenter la vie littéraire à différentes périodes (au double sens de la mesurer et de la revisiter, à plus ou moins grands pas). Cette mise au jour se fait selon trois axes de recherche : la conception des bases de données, leurs usages, et leur stockage.

This dossier therefore seeks to contribute to breaking away from the idea of radical change that still nourishes certain contemporary debates, emphasizing instead the way in which sociohistorical studies of print and literature have traditionally constructed and used databases. This movement indeed often goes together with the use of sociological concepts such as field and network, mobilizing (and sometimes revisiting) certain theoretical approaches. We can see this at work in the contribution by Marie-Frédérique Desbiens and Chantal Savoie, on the research methods used to assess the autonomisation of the literary field in Quebec. This is also the case for Florence Bonifay's article, which questions the conditions for the application of this same concept to literary history. Through a rigorous deconstruction of a certain number of shared representations associated with the group – incorrectly – known as “the Pléiade,” Bonifay reveals the tensions that ran through the French poetry scene in the second half of the sixteenth century, as well as the importance of sociability and exchanges in the social construction of individual careers, measured in an innovative way here using a corpus of published texts.

As a result, for sociohistorical studies of print and literature, the use of such databases responds to challenges that are not only technical but also scientific, as fundamental as the production of evidence or the definition of objects and research questions (which the use of databases necessarily makes explicit in most cases). By shedding light on what happens behind the scenes when databases are mobilised, this dossier therefore explores how digital data – which have become essential in many studies – allow us to survey literary life (in terms of both measuring and re-evaluating it) in different periods. In this perspective, three lines of questioning are proposed: how databases are designed; how they are used; and how they are stored.

La conception

Les choix méthodologiques en apparence anodins qu'engage la construction d'une base de données peuvent paradoxalement conduire à d'importantes questions d'ordre théorique, voire éthique et épistémologique. La machine, maniant des informations qui, par leur quantité même, peuvent sembler plus objectives et plus complètes que celles que peut traiter un esprit humain, effectue cependant des opérations qui dépendent toujours de choix et de questions de départ formulés par des chercheur·se·s. Comment, pour quelles raisons et par qui les bases de données sont-elles conçues ? Quelles relations, voire quelles négociations, s'établissent entre ceux ou celles (informaticien·ne·s, professionnel·le·s) qui en dessinent l'infrastructure (parfois par défaut) et les chercheur·se·s qui les utilisent, les mettent en rapport les unes avec les autres, les croisent avec d'autres données (textuelles, iconographiques, bibliographiques, archivistiques) afin de produire de nouveaux éléments de connaissance et de nouvelles problématiques ? La mise en tables et en catégories, l'ordonnement même des données ne sont en effet jamais neutres, et dépendent des choix toujours *situés* de ceux ou celles qui les mettent en œuvre (Hayles 2012). L'abondance et la diversité des données disponibles sur certains objets, du fait notamment de l'Internet, peuvent conduire à une volonté d'intégrer un ensemble quasi illimité d'informations. Mais la qualité des sources et des matériaux mobilisés pour construire une base de données peut influencer, comme des ingrédients en cuisine, sur celle des résultats. Il apparaît ainsi important de respecter l'hétérogénéité des sources et de permettre de s'y référer (Lemercier & Zalc 2008 : 50). Certaines données semblent plus fiables que d'autres, du fait de leurs récurrences, de leur cohérence ou de leurs régularités formelles sur la longue durée par exemple. Inversement, des données en apparence et en soi non significatives peuvent se révéler par la suite utiles. Où s'arrêter

Design

The apparently trivial methodological decisions that are involved in the construction of a database can lead to significant theoretical and even ethical and epistemological questions. A machine processes information that may, by its sheer quantity, seem more objective and more comprehensive than information that could be processed by the human mind; yet such a machine must conduct operations that still depend on choices and questions initially formulated by researchers. How, why, and by whom are such databases constructed? What relations, or even negotiations, are established between those (professional IT specialists) who design the infrastructure (sometimes by default) and the researchers who use it, and connect data sources or compare them with other textual, iconographic, bibliographic, or archival data to produce new elements of knowledge and new problem questions? Constructing tables, categories, and data sequences is never neutral and always depends on the *situated* choices of those who implement them (Hayles 2012). The abundance and diversity of available data on certain objects, notably due to the internet, can promote a desire to integrate an almost limitless collection of information. However, the quality of sources and materials used in the construction of a database can have an influence on the result – just like ingredients in a recipe. It therefore appears important to respect the heterogeneity of sources and allow them to be consulted (Lemercier & Zalc 2008: 50). Some data appears more reliable than others, because of recurrences, coherence, or formal regularity over the long term for example. Conversely, data that may be or may seem insignificant, may prove useful later. Where does one stop in the delimitation of a corpus, how does one avoid it becoming arbitrary? How can we select relevant data respecting an adequate legislative framework concerning intellectual work (such as legal submission and copyright)? Certain databases, such as



dans la délimitation d'un corpus en évitant l'écueil de l'arbitraire ? Comment sélectionner les données pertinentes en respectant un cadre législatif adéquat concernant les œuvres de l'esprit (dépôt légal, *copyright*) ? Certaines bases, comme l'[Index Translationum de l'Unesco](#) ou le catalogue Electre, nécessitent d'être replacées dans leurs conditions de production de manière à identifier les biais qui expliquent tel ou tel résultat chiffré (prise en compte, ou pas, des rééditions, variations de la déclaration des publications d'un pays de publication à l'autre, en fonction des politiques culturelles, etc.). C'est en particulier le cas lorsque ces bases ont été conçues pour répondre aux besoins de professionnel·le·s et non à des questions de recherche. La possibilité de croiser différentes sources et de mener des comparaisons internes semble en ce sens fructueuse, moyennant une approche critique et réflexive, comme le montrent les récents travaux dirigés par Gisèle Sapiro sur les flux de traductions (Bokobza & Sapiro 2008). Lorsqu'elle est une démarche individuelle, la constitution d'une base de données engage des difficultés de départ et des errances potentiellement décourageantes. Souvent chronophage, la conception de vastes bases de données mobilise la plupart du temps des équipes, ce qui impose une répartition des tâches, des plus simples en apparence (la saisie) aux plus chargées de sens (les codages, qui constituent déjà des interprétations, la construction de variables, puis les analyses). Quels sont les enjeux spécifiques de ce travail collectif, qui nécessite une réflexion préalable et, souvent, continue, sur chaque étape (Genet 2002) ? Comment réduire la marge de jeu et limiter les variations subjectives parfois importantes lorsque ces étapes sont réalisées par différents protagonistes aux interprétations divergentes (Merllié 1985) ? Idéalement, la division du travail doit favoriser l'érudition sur le cas abordé quant à l'organisation de la base, mais aussi le recul critique permettant de confronter différentes sources, d'identifier des doublons, de relativiser l'importance de matériaux. Enfin, certains projets

Unesco's [Index Translationum](#) or the Electre catalogue, need to be considered in the context of the conditions of their construction in order to identify the biases inherent in any given numerical result (for example whether or not reprints are taken into account, variations in the publication details from one country of publication to another depending on cultural policy, and so forth). This is particularly relevant in cases where databases were conceived to respond to professional requirements rather than to research questions. The possibility of combining different sources and conducting internal comparisons seems fruitful in this sense, assuming a critical and reflective approach, as we can see in the recent studies led by Gisèle Sapiro on the flow of translations (Bokobza & Sapiro 2008). When it results from an individual project, the construction of a database involves initial difficulties and potentially discouraging diversions. Often extremely time-consuming, the design of vast databases generally relies on teams, which implies the distribution of tasks, from the apparently simplest (data entry) to the most significant (coding, which already constitutes interpretation, the construction of variables, and analysis). What are the specific challenges of this collective operation, which requires prior and often ongoing reflection at each stage (Genet 2002)? How can we reduce the leeway and subjective variations that are sometimes significant when these stages are conducted by different players with diverging interpretations (Merllié 1985)? Ideally, the division of labour should facilitate erudition on the case under study regarding the organization of the database, but also the critical distance necessary to confront different sources, identify duplicate entries, and relativize the importance of certain materials. Finally, certain projects include a collaborative approach and encourage everyone to participate in enriching the database with their own contributions: what are the advantages and potential problems of this kind of approach inscribed in the open access movement?

incluent une démarche collaborative et invitent tout un chacun à venir enrichir de sa contribution la base de données constituée – quels sont les atouts et les écueils d'une telle démarche inscrite dans le mouvement de l'*open access* (ou accès ouvert) ?

Les usages

Si l'on peut céder à la fascination pour les gisements d'informations potentiellement illimitées que représentent les bases de données, ils ne constituent pas en eux-mêmes des garanties de rigueur scientifique. Les données assemblées dans une base peuvent, dans certains cas, donner lieu à des exploitations statistiques articulant différentes méthodes, des tris croisés à la régression logistique, en passant par diverses représentations graphiques. Ces démarches quantitatives, du comptage le plus basique aux analyses statistiques plus sophistiquées, s'articulent cependant le plus souvent à des présupposés, des définitions et des décisions d'ordre qualitatif, qui peuvent engager toute une vision du monde (délimitations géographiques ou historiques plus ou moins étendues, de l'événement ponctuel à la longue durée, représentation pacifiée ou conflictuelle des relations sociales, etc.). Qu'implique pour la recherche en études littéraires l'utilisation d'algorithmes et de programmes informatiques spécialisés dans le traitement des données ? Quel rôle accorder aux requêtes (simples ou croisées) qui sont souvent la voie d'accès principale aux bases de données ? Au-delà de ces paramètres techniques, l'exploitation de telles bases de données informatisées peut conforter ou déplacer, de manière heuristique, les questions usuellement posées en histoire littéraire. Quelles (re)découvertes ces démarches permettent-elles ? Quelles lectures des événements, des textes, des trajectoires individuelles et de groupe suscitent-elles, voire forcent-elles ? À quelles conditions permettent-elles de soutenir une argumentation, de vérifier une intuition ou de suggérer une nouvelle

Use

Although stockpiles of potentially limitless information are fascinating, they do not in themselves constitute a guarantee of scientific rigour. The data assembled in a database can, under certain circumstances, be treated statistically with different methods, from cross tabulation to logistic regression, as well as various forms of visual representation. These quantitative processes, from the most basic counting to the most sophisticated statistical analysis, are generally coupled with assumptions, definitions and decisions that are qualitative, and which may involve a whole worldview (broad or narrow geographical or historical delimitations, from one-off events to the *longue durée*, harmonious or conflictual conceptions of social relations, and so forth). What does the use of algorithms and specialized data analysis programs mean for research in literary studies? What role is there for queries (whether simple or combined) which are often the main form of access to databases? Beyond these technical parameters, the use of this kind of digital database can heuristically consolidate or shift the questions that are usually asked in literary history. What (re-)discoveries do these approaches enable? What readings of events, texts, or individual and group trajectories do they provoke or require? Under what conditions do they enable us to defend an argument, confirm an intuition, or suggest a new avenue for research? Although the construction of databases may seem laborious and thankless, can't it be an opportunity for playful and original explorations, allowing to consider in a new light some of the major notions in literary studies, such as intertextuality or literary events? In this perspective, it



piste ? Si la construction de bases de données peut sembler laborieuse et ingrate, leurs exploitations ne permettent-elles pas au contraire des explorations ludiques et originales, prenant à nouveaux frais certaines des notions majeures dans les études littéraires, telles que l'intertextualité ou l'événement littéraire ? Il peut s'agir de redessiner les chronologies habituelles en faisant ressortir certaines démarcations inaperçues à partir du corpus le plus légitime, de construire des générations d'auteur•e•s en partant d'indicateurs biographiques comme la date de la première publication, de faire apparaître des populations absentes du canon, féminines par exemple, comme le fait ici Florence Bonifay, mais aussi de s'intéresser aux possibles non advenus, d'examiner les distinctions admises entre le légitime et le populaire, le national et le transnational, etc.

Si la réalisation de ces bases impose bien souvent la prise en compte et l'interprétation de données empiriques (sur les producteur•rice•s et les biens littéraires, les localisations, les termes récurrents, etc.) en affinité avec les approches sociales et historiques de l'imprimé et du littéraire, elles permettent aussi de dialoguer avec la linguistique, l'économie ou la géographie par exemple. En quoi les bases de données nécessitent-elles ou favorisent-elles des perspectives interdiscursives ou interdisciplinaires (sciences sociales, histoire, sciences du langage, histoire de l'art) ?

Le stockage

De la même façon que les fiches réalisées à la main risquent de ne plus faire qu'un tas de papier dépourvu de signification après la retraite ou la mort de leur(s) concepteur(s) ou conceptrice(s), les bases de données informatisées peuvent devenir inutilisables faute de machine en état de les lire ou, pire, disparaître une fois le projet de recherche terminé. Quelle pérennité pour les bases

becomes indeed possible to redesign habitual chronologies by emphasizing certain – once imperceptible – demarcations in the most legitimate corpus; to reconstruct generations of authors based on biographical indicators such as the date of first publication, or to draw attention to populations that are absent from the canon, women for example, as Florence Bonifay does here. But it may also be a matter of focusing on unfulfilled possibilities, or re-examining the established distinctions between legitimate and popular, national and transnational, and so forth.

Although their construction often goes together with the necessity to take into account and interpret empirical data (on literary producers and literary goods, localizations, recurring terms and so forth) in connection with social and historical approaches to print and literature, databases also allow us to engage in dialogue with linguistics, economics or geography, for example. To what extent do they require or facilitate inter-discursive or interdisciplinary perspectives (between social science, history, language sciences, and art history)?

Storage

In the same way that old handwritten forms may end up as a mere pile of paper with no meaning after the retirement or death of their creator(s), digital databases may become unusable when there are no machines to read them, or worse they may disappear entirely once a research project is finished. What longevity is there for databases after their initial purpose? How can they be maintained



de données après une première vie ? Comment les maintenir en tant qu'outils pertinents ? Peut-on envisager de vérifier, voire de nettoyer les données par exemple ? Comment les rendre interopérables et où les conserver pour permettre de nouvelles explorations qui croiseraient plusieurs bases de données ? Certain·e·s chercheur·se·s isolé·e·s ne rendent au contraire jamais accessibles les bases de données construites et exploitées pour leurs recherches, par souci de contrôle ou bien faute de temps et d'informations – d'importants progrès ayant été réalisés en ce domaine, du moins en France, avec la très grande infrastructure de recherche [Huma-Num](#). Le coût d'accès aux données rassemblées, dont l'exploitation, le stockage et la sauvegarde peuvent dépendre de la détention d'outils informatiques et de la compétence pour les faire fonctionner, est en effet susceptible de renforcer les inégalités structurelles propres au monde de la recherche.

Claire Ducournau
Université Paul-Valéry-Montpellier 3 – [RIRRA21](#)
Anthony Glinoyer
[Université de Sherbrooke](#) – Chaire de recherche du Canada sur l'histoire de
l'édition et la sociologie du littéraire

Références bibliographiques

ALEXANDRE Didier (dir.) (2016). « Études littéraires et calcul numérique ». *Revue d'histoire littéraire de la France*, 116(3).

ALLINGTON Daniel, BROUILLETTE Sarah, GOLUMBIA David (2016). « Neoliberal Tools (and Archives) : A Political History of Digital Humanities ». *Los Angeles Review of Books*. [En ligne] <https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/> [consulté le 5 juin 2017].

as relevant tools? Can we consider testing them, or cleaning the data, for example? How can we make them interoperable or preserve them in order to enable new explorations to compare several databases? Some individual researchers never make the databases they constructed and exploited for their research accessible to others, out of a desire for control, or due to lack of time and information. However, significant progress has been made in this area, at least in France, with the development of the very large research infrastructure known as [Huma-Num](#). The cost of accessing assembled data (the exploitation, storage, and preservation of which may depend on having the digital tools and computer skills required to operate them) is indeed likely to reinforce the structural inequalities specific to the world of research.

Claire Ducournau
Université Paul-Valéry-Montpellier 3 – [RIRRA21](#)
Anthony Glinoyer
[Université de Sherbrooke](#) – Chaire de recherche du Canada sur l'histoire de
l'édition et la sociologie du littéraire

References

ALEXANDRE Didier (ed.) (2016). “Études littéraires et calcul numérique.” *Revue d'histoire littéraire de la France*, 116(3).

ALLINGTON Daniel, BROUILLETTE Sarah, GOLUMBIA David (2016). “Neoliberal Tools (and Archives): A Political History of Digital Humanities.” *Los Angeles Review of Books*. [On line] <https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/> [accessed on 5 June 2017].



BERNIER Sylvie & COUTURIER Yves (2007). *Les Banques de données numérisées dans la recherche en sciences humaines. Enjeux éthiques, juridiques et patrimoniaux*. Sherbrooke, Université de Sherbrooke.

BOKOBZA Anaïs & SAPIRO Gisèle (2008). « L'analyse des flux de traductions et la construction des bases de données ». In Sapiro Gisèle (dir.). *Translatio. Le Marché de la traduction en France à l'heure de la mondialisation*. Paris, CNRS Éditions : 45-64.

CHARLE Christophe (1979). *La Crise littéraire à l'époque du naturalisme : roman, théâtre et politique. Essai d'histoire sociale des groupes et des genres littéraires*. Paris, Presses de l'École normale supérieure.

DOZO Björn-Olav (2011). *Mesures de l'écrivain. Profil socio-littéraire et capital relationnel dans l'entre-deux-guerres en Belgique francophone*. Liège, Presses Universitaires de Liège, 2011.

DUBOIS Sébastien (2008). « Mesurer la réputation. Reconnaissance et renommée des poètes contemporains ». *Histoire & Mesure*, 23(2) : 103-143.

DUCOURNAU Claire (2017). *La Fabrique des classiques africains. Écrivains d'Afrique subsaharienne francophone (1960-2012)*. Paris, CNRS Éditions.

FLICHY Patrice & PARASIE Sylvain (dir.) (2013). « Sociologie des bases de données ». *Réseaux*, 178-179(2-3).

GEFEN Alexandre (2015). « Les enjeux épistémologiques des humanités numériques ». *Socio. La nouvelle revue des sciences sociales*, 4 : 61-74.

GENET Jean-Philippe (2002). « Analyse factorielle et construction des variables. L'origine géographique des auteurs anglais (1300-1600) ». *Histoire & Mesure*, 17(1-2) : 87-108.

GENET Jean-Philippe & LAFON Pierre (2003). « Des chiffres et des lettres. Quelques pistes pour l'historien ». *Histoire & Mesure*, 18(3-4) : 215-223.

BERNIER Sylvie & COUTURIER Yves (2007). *Les Banques de données numérisées dans la recherche en sciences humaines. Enjeux éthiques, juridiques et patrimoniaux*. Sherbrooke, Université de Sherbrooke.

BOKOBZA Anaïs & SAPIRO Gisèle (2008). « L'analyse des flux de traductions et la construction des bases de données. » In SAPIRO Gisèle (ed.). *Translatio. Le Marché de la traduction en France à l'heure de la mondialisation*. Paris, CNRS Éditions: 45-64.

CHARLE Christophe (1979). *La Crise littéraire à l'époque du naturalisme: roman, théâtre et politique. Essai d'histoire sociale des groupes et des genres littéraires*. Paris, Presses de l'École normale supérieure.

DOZO Björn-Olav (2011). *Mesures de l'écrivain. Profil socio-littéraire et capital relationnel dans l'entre-deux-guerres en Belgique francophone*. Liège, Presses Universitaires de Liège.

DUBOIS Sébastien (2008). « Mesurer la réputation. Reconnaissance et renommée des poètes contemporains. » *Histoire & Mesure*, 23(2): 103-143.

DUCOURNAU Claire (2017). *La Fabrique des classiques africains. Écrivains d'Afrique subsaharienne francophone (1960-2012)*. Paris, CNRS Éditions.

FLICHY Patrice & PARASIE Sylvain (eds.) (2013). « Sociologie des bases de données. » *Réseaux*, 178-179(2-3).

GEFEN Alexandre (2015). « Les enjeux épistémologiques des humanités numériques. » *Socio. La nouvelle revue des sciences sociales*, 4: 61-74.

GENET Jean-Philippe (2002). « Analyse factorielle et construction des variables. L'origine géographique des auteurs anglais (1300-1600). » *Histoire & Mesure*, 17(1-2): 87-108.

GENET Jean-Philippe & LAFON Pierre (2003). « Des chiffres et des lettres. Quelques pistes pour l'historien. » *Histoire & Mesure*, 18(3-4): 215-223.



HAYLES N. Katherine (2012). *How We Think. Digital Media and Contemporary Technogenesis*. Chicago/London, University of Chicago Press.

JATON Florian & VINCK Dominique (2016). « Processus frictionnels de mises en bases de données ». *Revue d'anthropologie des connaissances*, 10(4) : 489-504.

JOYEUX-PRUNEL Béatrice (dir.) (2008). « Art et mesure ». *Histoire & Mesure*, 23(2).

JOYEUX-PRUNEL Béatrice (2010). « L'art et les chiffres : une mésentente historique ? Généalogie critique et tentatives de conciliation ». In Joyeux-Prunel Béatrice (dir.), avec la collaboration de Sigalo-Santos Luc. *L'Art et la Mesure. Histoire de l'art et méthodes quantitatives : sources, outils, bonnes pratiques*. Paris, Éditions Rue d'Ulm : 19-58.

KALIFA Dominique, RÉGNIER Philippe, THÉRENTY Marie-Ève, VAILLANT Alain (dir.) (2011). *La Civilisation du journal. Histoire culturelle et littéraire de la presse française au XIX^e siècle*. Paris, Nouveau Monde.

Lemercier Claire & Zalc Claire (2008). *Méthodes quantitatives pour l'historien*. Paris, La Découverte.

MERLLIÉ Dominique (1985). « "Comment vous êtes-vous connus ?" Une expérience de codification multiple ». *Actes de la recherche en sciences sociales*, 57 : 89-92.

MICHON Jacques (dir.) (1999). *Histoire de l'édition littéraire au Québec au XX^e siècle. La naissance de l'éditeur. 1900-1939*, volume 1. Montréal, Fides.

MICHON Jacques (dir.) (2004). *Histoire de l'édition littéraire au Québec au XX^e siècle. Le temps des éditeurs. 1940-1959*, volume 2. Montréal, Fides.

HAYLES N. Katherine (2012). *How We Think. Digital Media and Contemporary Technogenesis*. Chicago/London, University of Chicago Press.

JATON Florian & VINCK Dominique (2016). "Processus frictionnels de mises en bases de données." *Revue d'anthropologie des connaissances*, 10(4): 489-504.

JOYEUX-PRUNEL Béatrice (ed.) (2008). "Art et mesure." *Histoire & Mesure*, 23(2).

JOYEUX-PRUNEL Béatrice (2010). "L'art et les chiffres: une mésentente historique? Généalogie critique et tentatives de conciliation." In JOYEUX-PRUNEL Béatrice (ed.), avec la collaboration de SIGALO-SANTOS LUC. *L'Art et la Mesure. Histoire de l'art et méthodes quantitatives: sources, outils, bonnes pratiques*. Paris, Éditions Rue d'Ulm: 19-58.

KALIFA Dominique, RÉGNIER Philippe, THÉRENTY Marie-Ève, VAILLANT Alain (eds.) (2011). *La Civilisation du journal. Histoire culturelle et littéraire de la presse française au XIX^e siècle*. Paris, Nouveau Monde.

LEMERCIER Claire & ZALC Claire (2008). *Méthodes quantitatives pour l'historien*. Paris, La Découverte.

MERLLIÉ Dominique (1985). "'Comment vous êtes-vous connus?' Une expérience de codification multiple." *Actes de la recherche en sciences sociales*, 57: 89-92.

MICHON Jacques (ed.) (1999). *Histoire de l'édition littéraire au Québec au XX^e siècle. La naissance de l'éditeur. 1900-1939*, volume 1. Montréal, Fides.

MICHON Jacques (ed.) (2004). *Histoire de l'édition littéraire au Québec au XX^e siècle. Le temps des éditeurs. 1940-1959*, volume 2. Montréal, Fides.



MICHON Jacques (dir.) (2010). *Histoire de l'édition littéraire au Québec au xx^e siècle. La bataille du livre. 1960-2000*, volume 3. Montréal, Fides.

MORETTI Franco (dir.) (2016). *La Littérature au laboratoire*. Paris, Ithaque.

MORIN Annie, BOSC Patrick, HEBRAIL Georges, LEBART Ludovic (2002). *Bases de données et statistiques*. Paris, Dunod.

OLLION Étienne & BOELAERT Julien (2015). « Au-delà des *big data*. Les sciences sociales et la multiplication des données numériques ». *Sociologie*, 6(3). [En ligne], <http://sociologie.revues.org/2613> [consulté le 6 février 2017].

PONTON Rémy (1977). *Le Champ littéraire de 1865 à 1906. Recrutement des écrivains, structures des carrières et production des œuvres* (thèse de doctorat en sociologie). Paris, Université de Paris V.

SAPIRO Gisèle (1999). *La Guerre des écrivains, 1940-1953*. Paris, Fayard.

SAPIRO Gisèle (2010). « Mesure du littéraire : approches sociologiques et historiques ». In Joyeux-Prunel Béatrice (dir.), avec la collaboration de Luc Sigalo-Santos. *L'Art et la Mesure. Histoire de l'art et méthodes quantitatives : sources, outils, bonnes pratiques*. Paris, Éditions Rue d'Ulm : 59-93.

VAILLANT Alain (dir.) (1992). *Mesure(s) du livre*. Paris, Bibliothèque nationale.

VESNA Victoria (2007). *Database Aesthetics : Art in the Age of Information Overflow*. Minneapolis, University of Minnesota Press.

VIALA Alain (1985). *Naissance de l'écrivain. Sociologie de la littérature à l'âge classique*. Paris, Minuit.

MICHON Jacques (ed.) (2010). *Histoire de l'édition littéraire au Québec au xx^e siècle. La bataille du livre. 1960-2000*, volume 3. Montréal, Fides.

MORETTI Franco (ed.) (2016). *La Littérature au laboratoire*. Paris, Ithaque.

MORIN Annie, BOSC Patrick, HEBRAIL Georges, LEBART Ludovic (2002). *Bases de données et statistiques*. Paris, Dunod.

OLLION Étienne & BOELAERT Julien (2015). "Au-delà des *big data*. Les sciences sociales et la multiplication des données numériques." *Sociologie*, 6(3). [On line], <http://sociologie.revues.org/2613> [accessed on 6 february 2017].

PONTON Rémy (1977). *Le Champ littéraire de 1865 à 1906. Recrutement des écrivains, structures des carrières et production des œuvres* (PhD dissertation, sociology). Paris, Université Paris V.

SAPIRO Gisèle (1999). *La Guerre des écrivains, 1940-1953*. Paris, Fayard.

SAPIRO Gisèle (2010). "Mesure du littéraire: approches sociologiques et historiques." In JOYEUX-PRUNEL Béatrice (ed.), with the collaboration of Luc Sigalo-Santos. *L'Art et la Mesure. Histoire de l'art et méthodes quantitatives: sources, outils, bonnes pratiques*. Paris, Éditions Rue d'Ulm: 59-93.

VAILLANT Alain (ed.) (1992). *Mesure(s) du livre*. Paris, Bibliothèque nationale.

VESNA Victoria (2007). *Database Aesthetics: Art in the Age of Information Overflow*. Minneapolis, University of Minnesota Press.

VIALA Alain (1985). *Naissance de l'écrivain. Sociologie de la littérature à l'âge classique*. Paris, Minuit.