

Enjeux de la constitution et de l'exploitation de bases de données en sociologie de l'art et de la culture

Issues Surrounding the Creation and Use of Databases in the Sociology of Art and Culture

Björn-Olav Dozo and Olivier Lapointe

Volume 15, Number 1-2, 2012

URI: <https://id.erudit.org/iderudit/1014625ar>
DOI: <https://doi.org/10.7202/1014625ar>

[See table of contents](#)

Publisher(s)

Globe, Revue internationale d'études québécoises

ISSN

1481-5869 (print)
1923-8231 (digital)

[Explore this journal](#)

Cite this article

Dozo, B.-O. & Lapointe, O. (2012). Enjeux de la constitution et de l'exploitation de bases de données en sociologie de l'art et de la culture. *Globe*, 15(1-2), 41-64. <https://doi.org/10.7202/1014625ar>

Article abstract

This article discusses some of the issues related to the creation and use of databases in research projects in the field of the sociology of art and culture. Using three concrete examples from the history of books ("*Base de données des métiers du livre au Québec*"), the sociology of art ("*Base des prix artistiques au Québec dans l'entre-deux-guerres*"), and literary studies ("*Base de données sur les figurations romanesques de la vie littéraire aux XIX^e et XX^e siècles*"), it explores questions arising from the use of databases in these specific cases, as well as other potential applications. The article also offers general reflections on the development and standardization of databases in these fields.

Mais finalement, à quoi sert une base de données? Qu'espérer de ces quelques tableaux informatiques reliés? Cet article entend aborder quelques usages génériques de cet outil qui permettront aux porteurs de projets de savoir si, tout bien pesé, la base de données est la meilleure carte qu'ils ont à jouer. Nous ne proposerons pas ici une réflexion détaillée sur l'élaboration informatique des bases de données. Cette étape devra être réalisée au sein du projet de recherche, en se fondant sur un dialogue entre les chercheurs et les informaticiens. Elle constitue une phase cruciale, qui conditionne l'exploitation future de la base de données. Il existe, pour le lecteur intéressé, différents ouvrages qui proposent de guider le chercheur novice dans l'élaboration et l'exploitation d'une base de données⁷.

Avant toutefois de penser à conduire la phase de constitution de la base de données proprement dite, il nous semble important de présenter les possibilités d'analyse inhérentes à cet outil. C'est l'objet même de cet article. Pour cela, nous mobiliserons trois approches du social, illustrées par trois projets de recherche qui mettent plus ou moins l'accent sur l'une de ces approches. Nous espérons qu'à partir de la présentation de ces trois études de cas, le lecteur intéressé pourra extrapoler l'un ou l'autre de ces cas à son objet de recherche.

Examinons rapidement trois des entités sociales explorées par les sciences humaines et sociales : les attributs, les relations et les textes. Chaque entité correspond à une certaine perspective pour approcher l'objet étudié. Les trois bases de données présentées dans la suite de l'article émanent de projets de recherche mobilisant une de ces trois entités⁸. Il nous a semblé pertinent de partir de ce degré de généralité élevé pour montrer combien l'usage de bases de données pouvait être divers et cependant fondé sur les mêmes principes de description systématique et de mise en série des objets étudiés.

Les attributs sont ce qui permet de qualifier quelque chose. Tel agent a vingt ans, est de sexe féminin, exerce la profession d'avocat, etc. Ces attributs correspondent à une approche catégorielle de l'objet, classique en sociologie.



7. Mentionnons, entre autres, l'ouvrage de Jacques CELLIER et Martine COCAUD, *Traiter des données historiques. Méthodes statistiques/Techniques informatiques*, Rennes, Presses de l'Université de Rennes, 2001, ainsi que, pour les littéraires, le livre de Benoît HABERT, *Construire des bases de données pour le français*, vol. 1, Paris, Ophrys, 2009.

8. Certains projets ont évolué au fil de leur développement pour rendre compte d'autres dimensions de leur objet : ainsi de la base des Métiers du livre au Québec, qui se concentrait au départ sur une description attributive des agents, pour recenser en plus par la suite les relations que ceux-ci entretiennent entre eux.

Les relations repèrent les liens entre éléments. La nature de ces liens et de ces éléments peut varier grandement : elle dépend en fait d'un choix du chercheur, qui doit la définir. Ces relations correspondent à une approche réticulaire (ou relationnelle) d'un objet, généralement couplée à une analyse structurale. En sociologie, elle constitue un paradigme concurrent de l'approche catégorielle : l'analyse structurale des relations⁹, paradigme dont les tenants, « à rebours des sociologies qui font du social un ordre de réalité *sui generis*, [...] partent de l'hypothèse plus "nominaliste", si l'on veut, que la société est constituée de relations entre individus *et de rien d'autre*¹⁰ ».

Le travail du texte (peu importe la méthode mobilisée ou la nature du texte) constitue le point commun de toutes les disciplines en sciences humaines (de la philologie ou la génétique à la sociocritique ou l'épistémocritique, en passant par la poétique, la rhétorique ou la sémiotique). Le texte constitue l'objet lui-même, voire les objets s'ils sont plusieurs. Il est décortiqué et analysé, contrairement aux attributs et aux relations qui incarnent des unités indivisibles. Le texte contient du social et s'intègre dans le social. Il peut (doit ?) donc être approché selon ces deux axes, notamment lors de l'élaboration d'une base de données le concernant.

Venons-en à la présentation des trois études de cas qui nous permettront d'illustrer tour à tour chacune des trois approches définies précédemment et de démontrer comment celles-ci peuvent s'incarner dans une base de données. Ces études de cas ont été tirées de nos parcours de chercheurs. Nous avons en effet été amenés à plusieurs reprises, et ce, auprès de différents groupes de recherche, à jouer le rôle d'intendants dans le processus de constitution de bases de données. Nous avons de ce fait été régulièrement confrontés aux divers questionnements qui peuvent survenir lors de ce processus.

Après avoir, pour chacune des trois bases de données ciblées par notre démonstration, présenté brièvement la structure de la base de même que les objectifs qui en ont guidé l'élaboration, nous serons amenés à souligner les avantages du recours à cet outil dans le cadre du ou des projets de recherche associés.



9. Voir, comme introduction, l'ouvrage d'Alain DEGENNE et Michel FORSÉ, *Les réseaux sociaux*, 2^e édition, Paris, Armand Colin, 2004, et Pierre MERCKLÉ, *Sociologie des réseaux sociaux*, 2^e édition, Paris, La Découverte, 2011.

10. Frédéric CLAISSE, « De quelques avatars de la notion de réseau en sociologie », Daphné DE MARNEFFE et Benoît DENIS (dir.), *Les Réseaux littéraires*, Bruxelles, Le Cri / CIEL-ULB-ULg, 2006, p. 24.

QUESTIONS DE VOCABULAIRE : LES BASES DE DONNÉES ET LEURS COMPOSANTES

Il nous paraît utile, avant de passer à la présentation de nos trois études de cas, et cela afin d'en faciliter la compréhension, de discuter très brièvement de ce qu'est une base de données de même que des principaux éléments du vocabulaire associé à cet outil informatique. Il suffit en fait, pour les besoins de notre démonstration, de concevoir une base de données, du moins, une base de données relationnelle, comme une série de tableaux interreliés. Ces tableaux sont dénommés « tables » et sont structurés en plusieurs colonnes : les « champs ». Chacune de ces tables contient un certain nombre de lignes : les « enregistrements ». À chaque enregistrement est assignée une valeur d'index, une « clé primaire », qui permet de retrouver facilement l'enregistrement ciblé et d'ainsi avoir accès aux informations qui lui sont liées. On peut, pour diverses raisons, mais généralement pour éviter les redondances et ainsi optimiser la base de données, vouloir relier une table à une autre. Par exemple, on pourrait vouloir lier une table contenant une liste d'œuvres littéraires avec une table contenant les informations biographiques d'un certain nombre d'écrivains. On serait ainsi en mesure d'associer à plusieurs œuvres littéraires recensées dans notre base de données les informations biographiques concernant leur auteur, sans avoir à inscrire à plusieurs reprises ces informations dans la base.

Lorsqu'on désire ajouter, modifier ou supprimer un enregistrement dans une des tables d'une base de données, nous devons effectuer une « requête ». On peut, pour ce faire, soit directement envoyer une requête au serveur, soit passer par une « interface » graphique qui simplifie fortement cette opération. Dans des logiciels commerciaux de grande diffusion, tels FileMaker et, pour certains usages, Microsoft Access, cette interface est créée par l'utilisateur au même moment que la base de données et, dans l'esprit de cet utilisateur, l'interface est bien souvent confondue avec la base de données. Dans d'autres solutions logicielles, plus robustes, dont Oracle Database ou MySQL, l'utilisateur peut créer sa propre interface. Celle-ci permet notamment à des personnes qui ne sont pas familiarisées avec le langage SQL¹¹ d'effectuer des requêtes sur la base de données. Comme ces requêtes sont le principal moyen d'interagir avec la base de données, même pour une opération aussi simple que la recherche dans un champ précis, il est nécessaire, dès lors que l'on désire recourir à ce type d'outil dans le cadre d'un projet de

✦ ✦ ✦

11. Le langage SQL (*Structured Query Language*) est un standard informatique permettant la définition, la manipulation et le contrôle des données inscrites dans les bases de données relationnelles.

recherche, de construire une interface graphique suffisamment ergonomique et paramétrable pour permettre à chaque chercheur de la consulter en toute autonomie, c'est-à-dire sans recourir pour chaque requête au service d'un informaticien.

Cette interface peut être différente de celle destinée, par exemple, au grand public, lors de la mise en place d'un accès ouvert à une communauté d'utilisateurs plus large que celle des chercheurs associés au projet. Une base de données bien pensée doit offrir la possibilité de mobiliser différentes interfaces, répondant aux différents usages de la base : une interface de consultation, une de saisie des données, une présentant seulement certaines informations, etc. Il est important, pour le porteur d'un projet de recherche, de bien faire la différence entre la base de données, qui peut être comparée à un grand stock d'informations, et les différentes interfaces, qui sont autant de vitrines mettant en évidence et en valeur ces informations. L'interface publique d'une base de données peut ainsi être développée en fin de projet, mais il ne faut pas perdre de vue que c'est une opération qui a un coût, tant financier qu'en termes de temps de développement. On connaît trop de bases de données qui existent sur les ordinateurs personnels des membres d'une équipe, mais qui n'ont jamais été rendues publiques parce que le développement d'une interface spécifique à cet usage n'était pas prévu. On ne trouve d'ailleurs sur Internet que très peu de bases de données en sociologie ou en histoire de la culture mises en ligne par des chercheurs québécois¹². Or, à notre sens, ce processus de mise en ligne garantit aux citoyens la possibilité d'accéder à une partie fondamentale du travail des chercheurs financés par des fonds publics.

Les trois bases présentées dans la suite de cet article sont encore en développement, mais envisagent (voire proposent déjà) une interface publique. Nous nous attarderons néanmoins plus spécifiquement sur les interfaces de saisie des données et de consultation par l'équipe de recherche, car elles correspondent à l'usage premier d'une base de données dans un projet de recherche et parce que, sans elles, l'interface publique n'aurait pas grand sens.

✦ ✦ ✦

12. Soulignons, à ce sujet, l'intérêt du Projet DÉCALCQ du Laboratoire Ex Situ visant à mettre en ligne les fonds documentaires et les bases de données liés aux nombreux projets antérieurs et actuels du Centre de recherche interuniversitaire sur la littérature et la culture québécoise (CRILCQ). Le lecteur intéressé pourra se référer à la description du projet accessible, en ligne, à l'adresse <http://carnets.contemporain.info/ex-situ/projets> (10 avril 2012).

LA « BASE DE DONNÉES SUR LES MÉTIERS DU LIVRE AU QUÉBEC »

Nous nous pencherons tout d'abord sur la « Base de données sur les métiers du livre au Québec » qui constitue, en fait, le support documentaire principal du Groupe de recherche et d'études sur le livre au Québec (GRÉLQ) pour le projet « Les métiers du livre au Québec » (MLQ).

Le projet MLQ s'inscrit dans la foulée d'une histoire du livre qui, prise dans son acceptation la plus large, s'intéresse au contexte de production et de diffusion du livre, de même qu'aux pratiques et aux usages de la lecture. Les individus qui interviennent dans le milieu du livre ont jusqu'ici encore peu retenu l'attention des chercheurs. Pourtant l'étude de leurs trajectoires permet de développer notre connaissance de chacun des métiers du livre tout en améliorant notre compréhension générale du système-livre. Qui sont, par exemple, les traducteurs, ouvriers de l'ombre sur qui repose le succès de plusieurs auteurs ? Quand le métier de distributeur est-il né et pourquoi a-t-il un poids économique si important à l'heure de la convergence dans le monde des médias ? Autant de questions auxquelles le projet tentera de répondre¹³.

Le développement d'une base de données sur Internet est le prolongement naturel des activités du GRÉLQ. Le groupe collecte en effet depuis ses débuts de très nombreuses informations inédites sur des individus, des entreprises, des organismes et des lois concernant les métiers du livre, de la Nouvelle-France à nos jours. Ces informations étaient conservées jusqu'en 2008 au sein de fichiers FileMaker partagés. Les travaux du CIEL¹⁴ ont persuadé les directrices du GRÉLQ, Marie-Pier Luneau et Josée Vincent, de suivre cette voie : François Melançon et nous avons donc travaillé sur la constitution des structures de la base et sur son développement informatique. Depuis lors, de nombreux étudiants y inscrivent des données, en vue de faire de cette base un guide et un soutien pour « les travaux actuels des rédacteurs des notices du *Dictionnaire historique des gens du livre au Québec*. [La base] pourra, à terme, servir à développer de nouveaux axes de recherche sur l'histoire du livre au Québec¹⁵ ».

✦ ✦ ✦

13. Cette description correspond au texte de la page de présentation du projet de recherche : « Groupe de recherche et d'études sur le livre au Québec », <http://www.usherbrooke.ca/grelq/recherche/projets-de-recherche/les-metiers-du-livre-au-quebec> (21 juin 2011).

14. Le CIEL est le Collectif interuniversitaire d'étude du littéraire, un groupe de recherche belge de l'Université Libre de Bruxelles et de l'Université de Liège, qui a développé une base de données sur les auteurs, les œuvres et les revues belges francophones concernant la période 1920-1960. Cette base est accessible à l'adresse <http://www.ciel-litterature.be> sur inscription gratuite.

15. Selon les termes de la présentation de la base de données à l'adresse <http://www.usherbrooke.ca/grelq/recherche/projets-de-recherche/les-metiers-du-livre-au-quebec/base-de-donnees-sur-les-metiers-du-livre-au-quebec>.

L'ambition du GRÉLQ est donc d'inscrire dans la pérennité le développement de cet outil. L'idée générale est de fonctionner autour de différentes entités, que nous allons détailler, et qui se trouvent interconnectées entre elles. Chaque entité est décrite par une série d'attributs qui permettent de la catégoriser : par exemple, les individus peuvent être regroupés par sexe, métiers, lieux de résidence, etc. Cette logique attributive s'inspire largement de celle qui a prévalu lors de la constitution de la base du CIEL. Organisée elle aussi à partir de différentes entités (auteurs, œuvres et revues), elle présente des similitudes de découpage du social. La base MLQ s'étend néanmoins sur une période bien plus vaste et concerne un tout autre cadre socio-historique. Il a donc fallu repenser certaines catégorisations. La parenté entre les deux bases reste malgré tout patente.

Avant de passer au détail de la structure actuelle de la base, il est utile d'insister sur la souplesse qu'induit une structuration autour de quelques grandes entités décrites chacune par une table principale et des tables secondaires qui lui sont reliées. Une des principales adaptations, qui illustre également la fécondité qu'il y a à travailler sur plusieurs projets, a été la création d'une table de relations entre individus. La structure de la base de données du CIEL ne prévoyait pas de moyen de notifier des liens directs entre agents autre qu'un champ commentaire non structuré. Les seuls liens repérés et structurés dans cette base étaient ceux d'affiliation (d'un individu à un groupe) et, par transitivité, entre les membres affiliés à un même groupe¹⁶. Le projet « Jurys », dont il sera question plus loin et sur lequel nous travaillons à l'époque du développement de la base MLQ, a attiré notre attention sur les relations interpersonnelles directes et la nécessité de les saisir¹⁷ d'une manière ou d'une autre dans la base MLQ. Cela fut fait dans une table des relations qui permet à l'heure actuelle de générer directement, d'un clic dans l'interface produisant une requête programmée, une liste des liens d'un individu avec tous les autres qui ont un rapport avec l'histoire des métiers du livre, présents ou non dans la base de données.



16. Cette structure a cependant déjà permis d'étudier le réseau des lieux de sociabilité au sein du sous-champ francophone belge de l'entre-deux-guerres et de mettre en évidence le rôle des animateurs de la vie littéraire. À ce sujet, voir Björn-Olav DOZO, « Sociabilités et réseaux littéraires au sein du sous-champ belge francophone de l'entre-deux-guerres », *Histoire & mesure*, vol. 24, n° 1, 2009, p. 43-72.

17. La saisie, dans cet article, correspond à l'opération de structuration de l'information dans une base de données. Elle diffère du codage qui correspond pour sa part à une catégorisation au service d'une première analyse. Par exemple, pour les professions, l'enregistrement de la profession telle qu'elle existe dans la source est l'opération de saisie ; la normalisation des noms de professions à partir d'une nomenclature permettant de traiter ces données de manière systématique coïncide à l'opération de codage.

Structure et contenu

Venons-en à la description détaillée de la base de données MLQ. Elle tient compte de trois dimensions de l'histoire du livre et de l'édition au Québec et est structurée en fonction de celles-ci : les acteurs, les organismes et les lois. On y trouve ainsi trois tables principales interreliées (« Acteurs », « Organismes » et « Lois »), ainsi qu'une multitude de tables secondaires contenant des informations associées aux enregistrements de ces tables principales. Ces informations sont de natures diverses et ont, de fait, été regroupées en différentes catégories. Ainsi retrouve-t-on pour chacun des acteurs identifiés par le projet des informations générales sur sa vie (dates de naissance et de décès, nationalité, etc.), sa carrière, sa formation, ses lieux de résidence, ses prix et récompenses, son appartenance à des associations ou des partis politiques, etc. L'interface, développée pour faciliter la consultation et la modification des données de la base MLQ, permet à son utilisateur de visualiser les informations répertoriées catégorie par catégorie, chaque table de la base étant liée à une page du site. On peut ainsi, à titre d'exemple, obtenir très facilement la liste des emplois occupés par les acteurs du milieu du livre recensés par la base du GRÉLQ. C'est ce qu'on peut observer sur l'illustration suivante, qui reproduit l'interface de visualisation des informations sur les individus rassemblées dans la base MLQ.

ILLUSTRATION 1.

UNE PAGE DE L'INTERFACE DE CONSULTATION DE LA BASE MLQ.
ICI, LA LISTE DES POSTES OCCUPÉS
PAR L'IMPRIMEUR PAUL-ÉMILE ARBOUR.

GRÉLQ-Individus

Identification	Pseudos	Métiers	Autres métiers	Formations	Résidences	Prix	Appartenance	Sociabilités	Relations	Sources	Accueil	Déconnexion
Arbour, Paul-Émile												
1913	1922	Imprimerie	[Indéterminé]	?			Arbour & Dupont ltée					Employé
		Commentaire: On mentionne dans "Le Maître Imprimeur" qu'il était connu à cette date comme un chef très compétent.										
1922	1946	Imprimerie	Directeurs	Gérant de l'imprimerie			Arbour & Dupont ltée					Dirigeant
		Commentaire: Paul-Émile est à l'emploi de l'entreprise depuis 1913 et c'est son père qui le nomme gérant à la suite du décès de Charles Dupont.										
1930	1930	Imprimerie	Opérateurs de machines à imprimer	Imprimeur			[Indéterminé]					[Indéterminé]
		Commentaire: Imprimeur, sans lieu d'affectation dans l'Annuaire de Montréal, 1930, mais habite déjà la rue Saint-Denis.										
1933	1946	Imprimerie	Directeurs	Gérant d'imprimerie			Association des maîtres imprimeurs de Montréal					Dirigeant
		Commentaire: Présent à l'assemblée de fondation.										

On aperçoit, par ailleurs, dans cette illustration, une série d'onglets de navigation (« Identification », « Pseudos », « Métiers », « Autres métiers », etc.) qui constituent, en fait, la transposition de la structure de la base de données, chaque onglet donnant accès aux informations inscrites dans une des tables de cette base. Cette interface, accessible à partir du site du GRÉLQ¹⁸, permet ainsi de naviguer de façon très efficace au sein de la quantité extrêmement grande d'informations répertoriées. Si la base se veut notamment l'équivalent informatisé d'une encyclopédie papier, elle offre tout de même à ceux qui la consultent un grand nombre d'avantages non négligeables, parmi lesquels des croisements plus nombreux, des classifications à la volée, une navigation plus rapide et des mises à jour aisées et régulières.

Les avantages

Ce rassemblement dans une base de données des informations mises au jour par le projet MLQ facilite, en fait, grandement leur exploitation. Celles-ci peuvent, en effet, être réellement appréhendées dans leur globalité. C'est qu'à l'interface de saisie et de modification des données de la base s'ajoute une autre interface : un outil de recherche qui permet la production de diverses statistiques sur lesquelles peut s'appuyer la production de prosopographies¹⁹ des milieux visés par le projet. On trouve dans l'illustration à la page suivante, par exemple, la liste de l'ensemble des acteurs répertoriés dans la base du GRÉLQ ayant suivi une formation universitaire et travaillé dans le secteur de l'imprimerie.

La possibilité de croiser les critères de recherche et de créer ainsi, en quelques clics de souris, des sous-corpus d'analyse permet aussi, soulignons-le, de mettre en évidence des réalités difficilement repérables avec les moyens d'analyse traditionnels. Le travail d'analyse statistique reste à faire sur cette base, mais l'expérience de la base du CIEL montre combien cette démarche peut être fructueuse²⁰. Autre avantage, l'inscription de ces données dans un support informatisé permet leur exportation dans un format utilisable par différents logiciels de statistiques, notamment pour des analyses

✦ ✦ ✦

18. On peut accéder à ce site à l'adresse suivante : <http://pages.usherbrooke.ca/mlq>.

19. La prosopographie consiste en la définition d'une population à partir d'un ou de plusieurs critères, puis à l'établissement à son propos d'un questionnaire biographique dont les diverses variables serviront à la décrire dans ses dynamiques sociale, privée, publique, voire culturelle, idéologique ou politique, selon la population et le questionnaire retenus.

20. Voir Björn-Olav DOZO, *La vie littéraire à la toise. Études quantitatives des professions et des sociabilités des écrivains belges francophones (1918-1940)*, préface de Jean-Marie KLINKENBERG, Bruxelles, Le Cri, 2010, et *Mesures de l'écrivain. Profil socio-littéraire et capital relationnel dans l'entre-deux-guerres en Belgique francophone*, Liège, Presses Universitaires de Liège – Sciences humaines, 2011.

factorielles des correspondances, dont l'usage heuristique – voire interprétatif en sociologie des champs – s'est répandu en sciences humaines et sociales depuis plus de 30 ans. Cette exportation de données en vue de leur utilisation dans des outils statistiques s'accompagne généralement d'un « toilettage » de ces données, qui peut aller de diverses retouches formelles à des recatégorisations plus profondes. Il ne faut donc pas s'imaginer qu'une fois l'information structurée dans la base de données, il suffit d'appuyer sur un bouton pour voir des graphiques parfaits apparaître sur l'écran. L'analyse statistique repose sur une démarche longue et souvent laborieuse, faite d'ajustements multiples, mais les résultats qu'elle peut produire en valent le prix²¹.

ILLUSTRATION 2.

UN EXEMPLE DE RÉSULTATS D'UNE RECHERCHE SUR LA BASE MLQ

GRELQ-Recherche

Individus Organismes Lois Accueil Déconnexion

Critères de recherche:

- **Niveau de formation:** Enseignement universitaire
- **Secteur (Métier du livre):** Imprimerie

Résultats de la recherche:

Nom de l'individu	Date de naissance	Date de décès
Boulangier, E. Gérard	1911 ?	avril 1950 ap.
Fontaine, Lucien	1950 av.	septembre 1964 ap.
Kaufmann, Otto	1922 av.	juin 1949 ap.
Lefebvre, Gaston	1947 av.	janvier 1952 ap.
Lefebvre, Paul-Émile	juillet 1949 av.	juillet 1949 ap.
Lévesque, Paul-É.	1949 av.	1954 ap.
Péladeau, Pierre	1925 v.	juin 1963 ap.
Turgeon, François	1936 av.	septembre 1951 ap.

Nouvelle recherche

On mesure avec la base MLQ combien l'utilisation d'une base de données, dans le cadre d'un projet sociohistorique, peut amener d'importants profits tant heuristiques qu'herméneutiques. Ces profits ne sont d'ailleurs pas

✦ ✦ ✦

21. Voir la note 17 sur la différence entre saisie et codage en vue d'une exploitation statistique.

tous visibles, ni programmés au moment de la conception du projet : les usages de bases de données en sciences humaines et sociales ne peuvent tous être prévus dès le départ²². Le développement d'une base de données doit donc être pensé sur le long terme, en choisissant des solutions souples et pérennes, qui répondent à des standards technologiques si possible ouverts (pour que les données ne se retrouvent pas prisonnières d'un format de programme informatique non documenté qui n'est plus soutenu par la firme qui l'a produit, par exemple). Le GRÉLQ nous semble avoir choisi cette voie.

LA BASE DE DONNÉES SUR LES PRIX ARTISTIQUES QUÉBÉCOIS DE L'ENTRE-DEUX-GUERRES

Penchons-nous maintenant sur une base de données beaucoup plus modeste, destinée à un usage très spécifique au départ, mais que nous avons voulue néanmoins potentiellement évolutive : la base de données sur les prix artistiques québécois de l'entre-deux-guerres (communément appelée base « Jurys²³ »). Celle-ci a été conçue en collaboration avec Michel Lacroix et est liée à un projet de recherche ayant pour objet l'histoire des prix artistiques québécois de l'entre-deux-guerres, projet dont les fondements méthodologiques et théoriques ont été fortement nourris par la sociologie des réseaux. Ce projet est une ramification des travaux de l'équipe PHVC²⁴ (« Penser l'histoire de la vie culturelle »), qui souhaite notamment penser les pratiques artistiques et culturelles de manière transdisciplinaire. Cette base de données résulte de la mise en commun de données disciplinaires (les différents prix « disciplinaires » organisés par des instances culturelles transdisciplinaires). L'étude de cette base de données propose une approche qui mobilise un type de données spécifique dont on constate la présence partout (donc dans toutes les disciplines) et qui appelle une définition minimale, transversale à toutes les disciplines : la relation effective entre deux agents, permettant de reconstituer des structures de réseaux relationnels²⁵.

Dans un premier temps, la base de données devait donc mettre l'accent non pas sur les attributs des individus et des prix (même si cette possibilité de développement n'est pas impossible dans le futur), mais bien plutôt sur les relations pouvant être établies entre eux. Nous avons recensé,



22. La réorientation (de l'histoire littéraire vers l'historiographie et la didactique de la littérature) de la Banque de données d'histoire littéraire du Centre Hubert de Phalèse (disponible à l'adresse <http://www.phalese.fr/bdhl/bdhl.php>) est un bon exemple de reconversion d'usage d'une base de données.

23. Celle-ci est accessible à l'adresse <http://jurs.phvc.ca>.

24. Pour une description détaillée de ce projet, voir son site, accessible à l'adresse <http://www.phvc.ca>.

25. Une première mouture de cette étude est disponible à l'adresse <http://hdl.handle.net/2268/131425>.

grâce au dépouillement de nombreuses sources archivistiques et médiatiques, les participants des jurys associés à différents concours artistiques québécois de 1918 à 1943 (prix David, Prix de l'Association catholique de la jeunesse canadienne-française, Prix d'Europe, etc.) et inséré ces informations dans la base de données. L'analyse de cette base a pu produire différents résultats²⁶. Nous avons ainsi fait apparaître les figures importantes de la vie culturelle québécoise de cette époque (ce que nous avons appelé ses principaux « instituants »), en mettant au jour les liens entre les jurys (et, par transitivité, entre les jurés) associés aux différentes pratiques culturelles et artistiques de l'époque. Nous avons aussi pu mettre à l'épreuve, grâce à un cas empirique original distinct, la proposition de Paul Aron et Benoît Denis concernant l'importance des réseaux dans les « institutions faibles²⁷ », en nous concentrant dans ce cas-ci sur la phase liminaire de la constitution du champ artistique et littéraire québécois, de ses structures institutionnelles et des réseaux « au sein » de ses appareils (d'un côté, celui du gouvernement du Québec ; de l'autre, celui de l'Université de Montréal ou de la Société royale du Canada). Enfin, l'exploitation de cette base nous a permis d'appréhender un usage différent des bases de données : nous ne l'avons pas construite en pensant à ses usages futurs, ni seulement comme un stock documentaire, mais bien plutôt en vue d'une utilisation particulière : l'approche réticulaire des jurys artistiques. Sa structure, nous allons le voir, est donc minimaliste. Elle n'est cependant pas fermée sur elle-même et se veut évolutive : l'objectif est de la connecter, à terme, avec d'autres « petites » bases formant autant de facettes du grand projet PHVC. Pour la comparer à la base du GRÉLQ, on pourrait dire qu'elle constitue une entité spécifique, isolée pour le moment, mais dont certaines tables pourraient être connectées à d'autres qui mobilisent les mêmes objets. Examinons sa structure plus en détail.

Structure et contenu

Les diverses informations inscrites dans les quelques tables qui composent la base « Jurys » le sont en fonction de son objectif principal : la recension des membres des jurys ciblés. On trouve ainsi, regroupées dans la table principale « Jurys », diverses informations concernant chacune des éditions annuelles des prix recensés par le projet. À cette table principale se trouvent rattachées quelques autres tables dont la table « Acteurs », à l'intérieur de laquelle ont été

✦ ✦ ✦

26. *Ibidem.*

27. Paul ARON et Benoît DENIS, « Introduction. Réseaux et institution faible », Daphné MARNEFFE et Benoît DENIS (dir.), *Les réseaux littéraires*, Bruxelles, Le Cri / CIEL-ULB-ULg, 2006, p. 7-18.

inscrites diverses informations sur les membres de ces jurys, de même que la table « Concours », qui recèle en son sein diverses informations sur les concours auxquels sont rattachés les jurys étudiés. Ce sont ces tables secondaires (en particulier la table « Acteurs ») qui pourraient aisément constituer des passerelles vers d'autres bases.

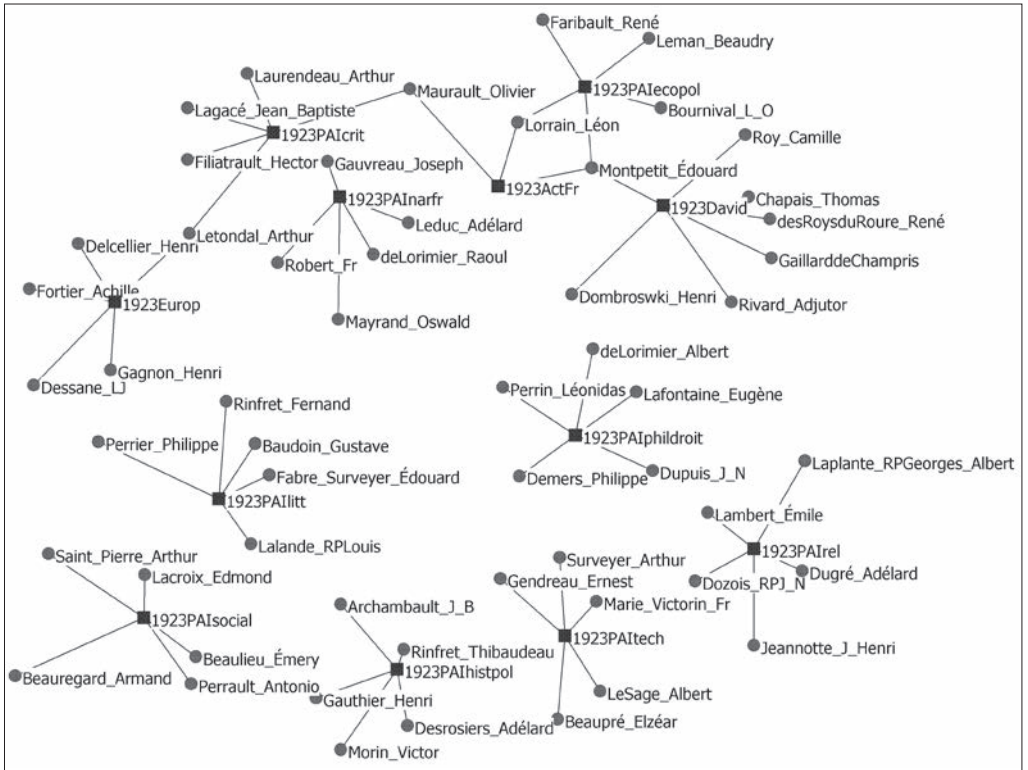
Il semble légitime de s'interroger, dès lors que l'on prend en compte la simplicité de la structure de cette base et des informations qui y ont été inscrites, sur la pertinence d'une telle entreprise. De simples fiches papier consignait, pour chacune des éditions des prix, les informations nécessaires au projet n'auraient-elles pas, en effet, suffi ?

Les avantages

En fait, l'inscription de ces informations dans notre base de données ne joue pas seulement un rôle documentaire. Celle-ci a en effet été structurée de façon à faciliter l'exportation des données dans un format qui permet d'effectuer toute une série d'analyses, à l'instar de la base du GRÉLQ qui fut construite, du moins en partie, en fonction d'une exploitation statistique des données consignées. Ainsi, à titre d'exemple, mentionnons le fait que l'encodage dans cette base de données des informations recueillies a permis leur traitement par des logiciels d'analyse structurale des réseaux sociaux. Cette exploitation spécifique a révélé l'importance, dans le milieu culturel canadien-français, d'individus (les « instituants »), tel Édouard Montpetit, qui, fortement impliqués dans l'univers des prix culturels et artistiques, se trouvaient bien souvent à faire le pont entre disciplines, entre des réseaux relationnels généralement clos sur eux-mêmes.

L'encodage, dans un format spécifique et selon une structure adaptée, des données dans la base « Jurys » a aussi facilité grandement la production de différents graphes relationnels permettant d'illustrer les conclusions obtenues par les logiciels d'analyse structurale des réseaux sociaux. On peut ainsi observer, à titre d'exemple, l'ensemble des relations entre les jurés des différents prix culturels canadiens-français de 1923 dans l'illustration suivante :

GRAPHIQUE 1.
LES RÉSEAUX FORMÉS PAR LES MEMBRES DES JURYS DES PRIX
CULTURELS
ET ARTISTIQUES CANADIENS-FRANÇAIS DE 1923²⁸



Ce graphe permet d'appuyer l'analyse structurale qui nous a permis de constater l'importance d'Édouard Montpetit au sein du réseau formé par les « instituants ». L'économiste de l'Université de Montréal fait ainsi le lien entre trois jurys : celui du Prix du Concours dramatique de l'*Action française*, celui du prix David et celui du Prix d'économie politique de l'Association catholique de la jeunesse canadienne-française. En dehors de

✦ ✦ ✦

28. PAI est l'acronyme de Prix d'action intellectuelle (de l'Association catholique de la jeunesse canadienne-française) ; PAIhistpol, de Prix d'histoire et de politique ; PAIitt, de Prix de littérature ; PAIrel, de Prix de littérature et sciences religieuses ; PAItech, de Prix de travaux scientifiques et techniques ; PAIecopol, de Prix d'économie politique ; PAIcrit, de Prix de critique littéraire et de critique d'art ; PAInarfr, de Prix de narration française ; PAIsocial, de Prix de sciences sociales ; ActFr, de Prix du Concours dramatique de l'*Action française* ; PAIphildroit, de Prix de philosophie et de droit ; David, de prix David ; et Europ, de Prix d'Europe.

ce potentiel d'analyse semi-automatisée des données et de production de graphes relationnels, précisons tout de même que cette base constitue en soi un important support documentaire touchant à la fois l'histoire des prix culturels canadiens-français de la première moitié du vingtième siècle et l'histoire des acteurs impliqués dans la distribution de ces prix. L'encodage de ces données accélère fortement, par ailleurs, certaines opérations à teneur documentaire, fondées sur des croisements et des classements de données. On peut ainsi, à titre d'exemple, à l'aide de l'interface de visualisation des membres des jurys recensés, obtenir très rapidement la liste complète des jurys auxquels ces derniers ont été rattachés. Ainsi peut-on observer, dans l'illustration suivante, la liste des jurys dont fut membre Olivier Maurault, historien de l'art et recteur de l'Université de Montréal de 1933 à 1955. Ce dernier, on le voit bien ici, joua un rôle central dans le processus de légitimation des acteurs de la vie littéraire du Québec de l'entre-deux-guerres.

ILLUSTRATION 3.

L'INTERFACE DE VISUALISATION DES INFORMATIONS ASSOCIÉES AUX ACTEURS RECENSÉS DANS LA BASE « JURYS ». ICI, LA LISTE DES CONCOURS DANS LESQUELS FUT IMPLIQUÉ OLIVIER MAURULT

PHVC-Jurys

Visualisation Recherche Gestion B.D. Wiki

Visualisation des acteurs répertoriés

Maurault, Olivier

Prénom:	Olivier
Nom:	Maurault
Date de naissance:	1886
Date de décès:	1968
Occupation(s):	Prêtre ; Professeur universitaire ; Recteur de l'Université de Montréal
Jurys liés:	0000-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1918-Concours littéraire - Société Saint-Jean-Baptiste de Montréal - nouvelle 1920-Concours de l'Action française 1921-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1922-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1923-Concours de l'Action française 1923-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1924-Concours littéraire - Société Saint-Jean-Baptiste de Montréal - conte et légende 1925-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1926-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1927-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1928-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1929-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1930-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1930-Prix David littérature 1932-Prix David littérature 1933-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1933-Prix David littérature 1935-Prix d'action intellectuelle - Prix de critique littéraire et de critique d'art 1935-Prix David littérature 1936-Prix David littérature

Cet exemple montre que même un petit projet qui pourrait sembler ne pas nécessiter une base de données lourde gagne à être pensé en adéquation avec un certain nombre de bonnes pratiques (structuration en tables reliées, décomposition de l'objet en différents champs, etc.²⁹). Si, dans ce cas-ci, l'usage de la base de données paraissait unique, la conformation de celle-ci à un schéma suffisamment raisonné pour être réexploitable dans un autre contexte en fait un travail qui n'est pas perdu après son exploitation première.

LA BASE « FIGURATIONS »

La troisième base de données dont nous dirons maintenant quelques mots est associée au projet « Figurations du personnel littéraire en France (1800-1940) » du groupe de recherche GREMLIN (Groupe de Recherche sur les Médiations Littéraires et les Institutions). Si, sur le plan purement informatique, elle est tout à fait comparable aux autres (elle est constituée de tables reliées entre elles), sur le plan des objets traités se marque une différence importante : la base a pour objectif de catégoriser des extraits de textes et de systématiser la lecture d'un corpus littéraire.

L'objectif de ce premier projet du GREMLIN tient en deux propositions : « dégager la socialité de textes qui socialisent la littérature, analyser comment on textualise la dimension sociale de la production de textes³⁰ ».

Dans ses grandes lignes, ce projet de recherche vise à rassembler, par des coupes ciblées, un corpus d'œuvres romanesques françaises mettant en scène la vie littéraire et à analyser les configurations qu'on y met en place et en mouvement. Ceci dans une perspective socio-critique, attentive aux médiations qui unissent ces configurations aux textes, discours et pratiques contemporains³¹.

Outre des outils théoriques adaptés (comme les concepts de figuration et de configuration), il fallait un outil informatique spécifique pour traiter une si vaste question et un si vaste corpus qui, soulignons-le, inclut maintenant des œuvres de la littérature québécoise.



29. Pour une liste de bonnes pratiques pour constituer une base de données, on peut se reporter à l'ouvrage très utile de Claire LEMERCIER et Claire ZALC, *Méthodes quantitatives pour l'historien*, Paris, La Découverte, 2008, en particulier les pages 41-42, mais plus généralement tout le troisième chapitre.

30. GREMLIN, « Fictions, figurations, configurations : introduction à un projet », GREMLIN (dir.), *Fictions du champ littéraire, Discours social*, vol. 34, 2010, p. 3.

31. *Ibid.*, p. 4.

La base de données vise essentiellement à rassembler des données « textuelles », c'est-à-dire à tirer des textes des extraits éclairant les multiples aspects de la figuration littéraire : réflexivité énonciative, qualification des personnages, modes et scènes de sociabilité, rapport aux lieux, à la parole ou à la débauche, intertextualité mobilisée, textes et publications « au second degré », etc. Au fur et à mesure que nous rassemblerons dans la base les résultats de nos lectures annotées, nous pourrions commencer à analyser avec plus d'acuité les constantes, inflexions, points aveugles ou obsessions de notre corpus. Quels sont par exemple les traits récurrents dans la description des éditeurs ? Avec quelle régularité traite-t-on les écrivaines fictives de « bas-bleus » ? [...] Par là, nous espérons tout à la fois élaborer des observations d'ordre général, historique, sur les figurations et configurations littéraires, et nourrir les lectures plus ciblées de certains des textes du corpus. Sur ce plan, on pourrait concevoir la base de données comme une version analytique et ciblée de bases permettant les recherches « plein-texte » comme celle de l'ARTFL.

Par ailleurs, mais de façon plus secondaire, nous espérons aussi obtenir des données plus quantitatives sur les romans « saisis » par et dans la base. Combien de romans, par exemple, mettent en scène, dans leurs premières pages, un jeune homme originaire de la province, monté à Paris et aspirant à la carrière d'écrivain ? Combien de ces romans se terminent sur la folie définitive ou le suicide du héros ? [...] Les résultats n'auront certes qu'une valeur indicative, liée entre autres au choix des romans dépouillés. Néanmoins, ils promettent d'être utiles, croyons-nous. C'est la raison pour laquelle nous envisageons, au terme de nos recherches, de rendre la base de données accessible à l'ensemble des chercheurs, de façon à ce que toute personne intéressée puisse en parcourir les extraits, y faire des requêtes, etc³².

À la lecture de ces extraits, on mesure à quel point la base de données joue un rôle central dans le projet. Cette « mise en série » de la littérature doit permettre une appréhension originale de ces textes, considérés non dans leur isolement, mais comme des éléments d'un vaste ensemble. Pour permettre cette mise en série, il fallait se mettre d'accord sur un protocole de lecture précis. C'est ce protocole qui est au fondement de la base de données.

Structure et contenu

Cette base de données peut être considérée, en fait, comme une version informatisée d'une fiche de lecture (le protocole) orientée en fonction des objectifs

✦ ✦ ✦

32. *Ibid.*, p. 14-15.

du projet. Elle se distingue donc des bases de données textuelles telles que FRANTEXT³³. Il ne s'agissait pas ici, en effet, de reproduire intégralement les textes ciblés par le projet, mais bien plutôt d'en saisir, en les catégorisant et en les commentant, certains extraits jugés significatifs. Soulignons aussi, par ailleurs, le fait que nous n'avons pas jugé pertinente l'application, lors de l'élaboration de cette base de données, des recommandations de la Text Encoding Initiative (TEI), un consortium qui a mis sur pied un standard pour la représentation informatisée de corpus textuels³⁴. Ces recommandations s'appliquent essentiellement, en effet, à des projets d'encodage informatique de textes complets. De plus, nombreux sont ceux qui ont démontré, à l'instar de Yin Liu et Jeff Smith³⁵, les limites et les défauts des recommandations de la TEI lorsque convoquées dans des projets, tel celui du GREMLIN, impliquant une analyse fine des objets textuels ciblés. Ces recommandations, de même que le langage informatique de balisage qui leur est associé, le XML, incitent en effet le chercheur à structurer de façon hiérarchique les éléments textuels encodés, hiérarchisation qui nuit fortement à l'appréhension simple et à l'analyse des réseaux de sens présents au sein des textes ciblés. Nous avons donc décidé d'inscrire nos données textuelles dans une base de données relationnelle qui n'implique pas nécessairement cette hiérarchisation des entités à l'étude.

Cette base est constituée d'une série de tables secondaires associées à une table principale, la table « Textes », dans laquelle se trouvent rassemblées différentes informations factuelles concernant chacune des œuvres ciblées par le projet (auteur, année de publication, éditeur, nombre de pages, etc.). Les tables secondaires contiennent quant à elles différents extraits textuels regroupés en fonction de thématiques communes. On trouve ainsi dans la table « Scènes d'écriture » l'ensemble des extraits des romans ciblés dans lesquels est décrit le labeur d'un personnage écrivain. Certaines de ces tables sont liées entre elles. Ainsi, la table « Scènes d'écriture » est-elle liée à la table « Œuvres projetées », qui contient des extraits où sont présentées des œuvres en préparation. On peut ainsi indiquer à laquelle de ces œuvres recensées dans la table « Œuvres projetées » est liée la scène d'écriture saisie dans la base.

L'interface de consultation et de modification de « Figurations » est modelée sur la structure de la base, chaque onglet donnant accès au contenu

✦ ✦ ✦

33. La base de données FRANTEXT rassemble plus de 4000 textes littéraires, philosophiques, scientifiques et techniques. Son interface de consultation est accessible à l'adresse <http://www.frantext.fr>.

















34. Voir, à ce sujet, le site de la TEI : <http://www.tei-c.org/index.xml>.

35. Yin LIU et Jeff SMITH, « A Relational Database Model for Text Encoding », *Digital Studies/Le champ numérique*, n° 12, 2008.

de l'une de ses tables. Ainsi retrouve-t-on dans l'onglet « Positions » l'ensemble des extraits textuels qui mettent en scène les prises de position (esthétiques, littéraires, politiques, etc.) des personnages ciblés par l'analyse. L'illustration suivante présente ainsi les prises de position relativement loufoques de l'un des personnages principaux du *Prix Lacomyne* de Renée Dunan : Paul Le Raive.

ILLUSTRATION 4.
L'INTERFACE DE VISUALISATION ET DE MODIFICATION
DES DONNÉES DE LA BASE « FIGURATIONS »

Figurations-Textes-*Le Prix Lacomyne*-Paul Le Raive

Accueil Textes Gestion des listes Gestion des usagers Recherche Modifications Déconnexion						
Identification Résumés Énonciation Personnages Configurations Textes Sphères Sources						
Catégories Activités culturelles Descriptions Langage Positions Capital Modalités Environnement						
Prises de position						
Type	Thème	Extrait	Début	Fin	Commentaire	
Esthétique	L'art	« L'art n'est pas un cornet de berlingots. »	23	23		 
Littéraire	Finalité de l'écriture	« J'écris pour masquer mon désir d'assassiner »	23	23		 
Littéraire	L'écrivain et son lecteur	« L'écrivain se doit de ne jamais être moins sot que son lecteur »	23	23		 
Littéraire	Le romancier et ses créations	« Le romancier a des devoirs envers ses personnages. »	23	23		 
Littéraire	Poésie	« La poésie est une forme de l'analphabétisme »	23	23		 
Littéraire	Psychologie	« La psychologie ne se porte plus que dans les porcheries »	23	23		 
[Autre]	Inspecteur des beaux-arts	« Il est plus beau d'empoisonner sa famille que d'être inspecteur des beaux-arts. »	23	23		 
[Autre]	La féminité	« On naît mâle, on devient femme. »	23	23		 

Ajouter un enregistrement

Comme il a été mentionné précédemment, cette base constitue essentiellement la version informatisée d'une fiche de lecture assez traditionnelle, orientée vers les enjeux spécifiques du projet. Il semble de ce fait légitime de se questionner sur les bénéfices que le projet peut tirer de cette mise en série d'extraits textuels et, plus largement, sur ce que ce type de lecture peut apporter à l'histoire de la littérature.

Les avantages

L'important potentiel heuristique de cette base découle en fait assez naturellement de la mise en série de la façon dont ces romans représentent le milieu littéraire. C'est en effet la première fois que l'on peut repérer de manière systématique (en fonction du corpus dépouillé, bien évidemment) des régularités dans les modes de figuration des sociabilités littéraires. On peut en effet rapidement, grâce à l'outil de recherche accessible par l'interface

de consultation de la base, comparer divers éléments des textes recensés. Cet outil de recherche, relativement complexe, permet de créer rapidement des sous-corpus d'analyse, sous-corpus qui facilitent la confrontation des romans en fonction de critères précis. On trouve par exemple dans l'illustration suivante la liste de l'ensemble des romans dépouillés dans la base, publiés après 1900, et qui invoquent d'une façon ou d'une autre Rimbaud.

ILLUSTRATION 5.

LES RÉSULTATS D'UNE RECHERCHE DANS LA BASE « FIGURATIONS »

Figurations-Recherche

The screenshot shows a web interface for 'Figurations-Recherche'. At the top, there is a navigation menu with buttons for 'Accueil', 'Textes', 'Gestion des listes', 'Gestion des usagers', 'Recherche', 'Modifications', and 'Déconnexion'. Below this, there are tabs for 'Simple', 'Avancée-Textes', and 'Avancée-Personnages'. The 'Critères de recherche' section includes:

- Identification :**
 - Publication-Intervalle :** Après 1900
- Intertextualité #1 :**
 - Quoi? :** Rimbaud

The 'Résultats de la recherche' section indicates '5 textes concernés' and displays a table with the following data:

Année	Titre	Auteur
1904	<i>Le Salon de Madame Truphot</i>	Fernand Kolney
1925	<i>Les Faux-Monnayeurs</i>	André Gide
1926	<i>Une jeune fille à la page</i>	J.-H. Rosny Aîné
1931	<i>Les Soeurs Hortensias</i>	Henri Duvernois
1934	<i>Les arrivistes ... et les autres</i>	J.-H Rosny

At the bottom of the results, there is a button labeled 'Nouvelle recherche'.

On s'imagine très bien l'intérêt d'un tel outil de recherche qui permet de naviguer rapidement au sein d'une masse textuelle extrêmement volumineuse. De plus, les données saisies permettent le recours à des outils d'analyse spécialisés qui peuvent faire apparaître diverses réalités insaisissables autrement. De la même manière que pour les autres bases, l'analyse factorielle ou l'analyse structurale des relations sociales, parmi d'autres outils (comme la classification hiérarchique ascendante), pourraient apporter un éclairage complémentaire. On peut, de plus, imaginer tout ce que pourrait révéler l'exploration par des outils d'analyse lexicographique, tel Alceste³⁶, de l'impressionnante masse d'extraits textuels regroupés dans cette base.

✦ ✦ ✦

36. Alceste est un logiciel de statistique textuelle développé par la société Image, en collaboration avec le CNRS.

Cette approche des textes par les outils informatiques, loin de fourvoyer la sociocritique dans un positivisme béat, lui offre de nouveaux défis, en soumettant au regard sociocritique une masse de textes très importante, qui rappelle les dimensions des corpus traités par Marc Angenot³⁷. Cette masse de textes, loin d'apporter des réponses définitives, impose au contraire de nouvelles questions, demandant l'adaptation des outils d'interprétation auxquels les littéraires sont habitués. Nous y voyons un défi immense pour l'histoire de la littérature : la confrontation aux textes qu'elle a toujours ignorés, faute d'outils adéquats pour les traiter, ces petits textes illégitimes, oubliés. Les bases de données comme celle du GREMLIN représentent une possibilité unique pour affronter ce type de questions.

LA BASE « ÉMERGENCE » ET LA DIFFUSION DES DONNÉES RECUEILLIES DANS LES BASES DE DONNÉES EN SOCIOLOGIE DE L'ART ET DE LA CULTURE

Nous avons exploré dans cet article trois cas de figure, trois exemples de la façon dont les bases de données peuvent venir appuyer des projets de recherche en sociologie ou en histoire de l'art et de la culture. Soulignons-le, les projets décrits, bien qu'ils mobilisaient trois types d'approches distinctes, ont été menés par des individus issus d'une seule et même discipline : les études littéraires, dans leur acception la plus large. Or, à la façon dont nous avons pu tenter une première mise en pratique avec la base « Jurys », la construction et l'utilisation d'une base de données dans le cadre de projets de recherche pluridisciplinaires peut déboucher sur d'importants profits heuristiques. De fait, il nous semble pertinent de mentionner ici le fait que nous avons participé, dans le cadre du projet de recherche « Penser l'histoire de la vie culturelle » (PHVC), à l'élaboration d'une autre base de données pluridisciplinaire, plus ambitieuse que la base « Jurys », soit la base de données « Émergence ». Celle-ci regroupe un très grand nombre d'informations concernant le milieu culturel canadien-français de la première moitié du XX^e siècle. Jusqu'à présent, les informations inscrites dans cette base l'ont été par des membres de différents groupes de recherche liés à PHVC et issus de différentes disciplines, dont les projets « La vie littéraire au Québec », « Caricatures et satires graphiques à Montréal » (CASGRAM) et « Regards croisés sur la réception critique francophone et anglophone des arts de la

✦ ✦ ✦

37. Voir, en particulier, Marc ANGENOT, 1889. *Un état du discours social*, Longueuil, Le Préambule, 1989.

scène à Montréal entre 1900 et 1950 : musique, théâtre, danse ». On y trouve des informations tirées de différentes sources, dont des articles de périodiques, des monographies et des fonds d'archives, concernant les œuvres et les acteurs associés aux milieux littéraires, musicaux et artistiques du Québec de la première moitié du XX^e siècle. À moyen terme s'ajouteront aux données déjà inscrites dans cette base des informations sur les milieux du cinéma, de l'architecture et du théâtre.

Cette base a ceci d'intéressant, entre autres choses, qu'elle permet d'explorer le caractère bien souvent pluridisciplinaire des carrières des acteurs recensés, de rendre compte du fait qu'il n'était pas rare, à cette époque, qu'un même acteur participe de différents champs culturels simultanément. Sans nécessairement permettre d'aplanir l'ensemble des difficultés associées à un investissement réellement pluridisciplinaire de notre objet de recherche, il nous paraît évident que cette base constitue un outil au fort potentiel heuristique. Elle permet, en effet, de rendre accessibles à des chercheurs d'une autre discipline des informations qui, bien trop souvent, ont été cantonnées dans leur discipline d'origine. Ainsi, par exemple, pouvons-nous affirmer que la confrontation, dans cette base de données, des informations associées à la carrière littéraire d'Albéric Bourgeois avec celles concernant sa carrière de caricaturiste a permis une réévaluation de son œuvre de caricaturiste, et vice-versa.

On constate que l'un des grands intérêts de la base « Émergence » est de rendre accessibles des informations, de faire circuler celles-ci entre les disciplines. Ce décloisonnement disciplinaire, qui constitue en soi un idéal, devrait aussi, à notre avis, être doublé d'un effort de diffusion des informations recensées dans les différentes bases de données en sociologie de l'art et de la culture présentement en activité. Il est en effet bien trop courant qu'une base de données construite afin d'appuyer un projet de recherche se trouve délaissée, inutilisée, sans que l'on ait pensé à en faire profiter l'ensemble de la communauté universitaire.

Plusieurs solutions à ce problème seraient envisageables. Nous pourrions, par exemple, profiter du regain d'intérêt pour le paradigme du « Web sémantique » provoqué récemment par l'annonce de l'adoption par les principaux moteurs de recherche d'un schéma³⁸ commun facilitant la sémantisation des documents Web, afin de mettre en branle un projet d'interconnexion des informations rassemblées dans les diverses bases de données

✦ ✦ ✦

38. Voir, à ce sujet, le site Web de ce projet regroupant Google, Yahoo et Bing : <http://schema.org>.

courantes en sociologie de l'art de la culture. L'adhésion d'une partie de la communauté universitaire québécoise à ce paradigme ainsi qu'aux diverses technologies qui lui sont associées, dont le standard RDF (resource description framework) et les microdonnées, permettrait de mettre en place une interface de recherche capable d'effectuer des requêtes sur plusieurs bases de données à la fois, à la condition que ces bases aient subi au préalable un processus de standardisation. La mise en place d'un schéma commun permettant l'interopérabilité de ces bases pourrait malheureusement, à notre avis, s'avérer extrêmement difficile en ce qu'elle requiert une véritable coordination des efforts de tous les participants³⁹. Il serait sûrement plus réaliste, puisqu'il nous semble tout de même crucial de mettre en œuvre de telles entreprises favorisant l'accessibilité des données rassemblées dans des bases bien trop souvent closes sur elles-mêmes, d'envisager la mise sur pied d'un site de référencement et de diffusion des différentes bases de données en sociologie et en histoire de l'art et de la culture présentes et à venir.

✦ ✦ ✦

39. Voir, à ce sujet, Kate BYRNE, « Putting Hybrid Cultural Data on the Semantic Web », *Journal of Digital Information*, vol. 10, n° 6, 2009.