

Une base de données en sociologie de l'art et de la culture :

pour quoi faire ?

Björn-Olav DOZO (F.R.S.-FNRS – Université de Liège)

Olivier LAPOINTE (Université de Montréal)

Depuis dix ans, les bases de données se sont imposées comme un passage quasiment obligé pour les projets de recherche en sociologie de l'art et de la culture se voulant résolument modernes. IncurSION de la technique au sein de la forteresse des humanités, nouveau moyen de poursuivre le Graal de l'exhaustivité, les bases de données sont de toutes les recherches, et surtout, de tous les financements. Pour certains, leur présence au sein d'un projet justifie quasiment à elle seule les fonds publics concédés.

Mais finalement, à quoi sert une base de données ? Qu'espérer de ces quelques tableaux informatiques reliés ? Cet article entend faire le point sur les espoirs et les difficultés suscités par cet outil et aborder quelques usages génériques 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 développeurs informatiques. Elle constitue une phase cruciale, qui conditionne l'exploitation future de la base de données.

Mais avant cette phase de constitution, il nous semble important de présenter sans jargon ce qu'est une base de données. 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. Commençons par définir grossièrement chaque entité, pour en cerner ses spécificités par rapport aux autres et ainsi préciser pourquoi on la différencie.

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.

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 relations.

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 ci-avant 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

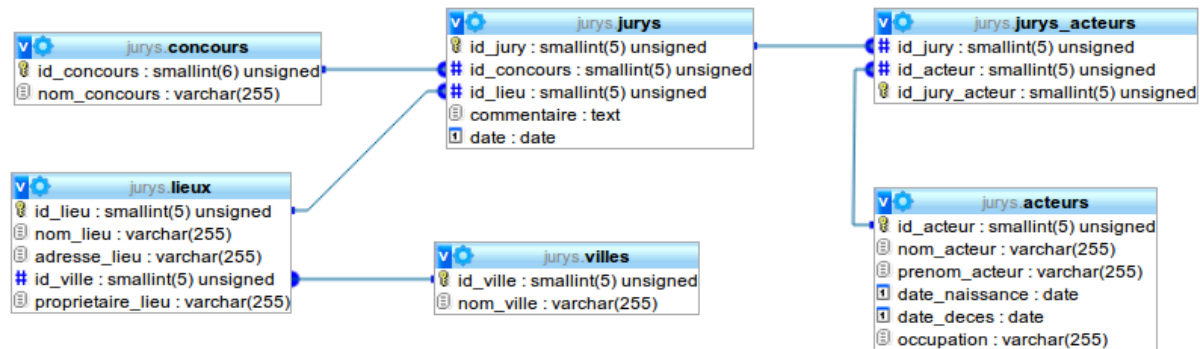
groupes de recherche, à jouer le rôle d'intendant dans le processus de constitution de bases de données. Nous avons de ce fait été régulièrement confrontés aux divers problèmes et questionnements qui peuvent survenir lors de ce processus.

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.

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 ce, 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 assigné 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 encoder à plusieurs reprises ces informations dans la base. Les liens entre deux tables sont inscrits dans des champs dénommés *clés étrangères* dont la valeur est généralement celle de la clé primaire de

l'enregistrement ciblé. On peut observer, à titre d'exemple, sur le schéma qui suit, les interrelations entre les différentes tables d'une des bases de données dont nous discuterons dans cet article, la base « Jurys » :



Légende : Schéma représentant la structure de la base « Jurys ».

On le voit très bien, sur ce schéma, cette base contient six tables reliées de diverses façons. Chaque table contient plusieurs champs. La table « jurys » contient ainsi cinq champs, dont une clé primaire, « id_jury », reliée à une clé étrangère présente dans la table « jurys_acteurs », une clé étrangère faisant le lien entre cette table et la table « concours », « id_concours », une autre clé étrangère qui fait, cette fois, le lien entre la table « jurys » et la table « lieux », « id_lieu », ainsi qu'un champ « date » permettant d'inscrire la date de formation du jury. On y retrouve aussi, enfin, un champ textuel dénommé « commentaire » permettant aux usagers de la base d'y inscrire diverses indications concernant les informations liées à l'enregistrement associé.

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 (en langage SQL), soit passer par une *interface* graphique qui simplifie fortement cette opération. Dans des logiciels commerciaux de grande diffusion tels FileMaker et Microsoft Access, cette interface est créée par l'utilisateur au même moment que la base de données et elle est bien souvent, dans l'esprit de cet utilisateur, confondue avec cette dernière. Dans d'autres solutions logicielles, plus robustes, dont Oracle

Database ou MySQL, l'utilisateur doit créer lui-même, du moins, s'il le désire, sa propre interface. Celle-ci permet notamment à des personnes qui ne sont pas familières avec le langage SQL 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 recherche, de construire une interface graphique suffisamment ergonomique et paramétrable pour permettre à chaque chercheur du projet 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 temps de développement. On connaît trop de bases de données qui existent sur les ordinateurs personnels des membres de l'équipe mais qui n'ont jamais été rendues publiques parce que le développement d'une interface spécifique à cet usage n'était pas prévu. Or, à notre sens, cette interface 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 que, sans elles, l'interface publique n'aurait pas grand sens.

La « Base de données sur les métiers du livre au Québec »

Nous nous attarderons tout d'abord à 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). Cette base de données a été développée au départ pour le projet « Les métiers du livre au Québec » (MLQ).

Le projet « Les métiers du livre au Québec » (MLQ) s'inscrit dans la foulée d'une histoire du livre qui, prise dans son acception 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 répondre¹.

Le développement d'une base de données sur internet est le prolongement naturel des activités du GRÉLQ depuis sa fondation. Il récolte en effet depuis le début 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. Celles-ci étaient conservées jusqu'en 2008 au sein de fichiers FileMaker partagés. Les travaux du CIEL² ont persuadé les directrices du GRÉLQ, Marie-Pier Luneau et Josée Vincent, de suivre cette voie : François

¹Cette description correspond au texte de la page de présentation du projet de recherche : <http://www.usherbrooke.ca/grelq/recherche/projets-de-recherche/les-metiers-du-livre-au-quebec/>, consultée le 21 juin 2011.

²Le 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.

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 encodent 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 Québec³ ».

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 un table principale et des tables secondaires qui lui sont reliées. Une des principales adaptations, qui illustre également l’aspect fécond 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

³ 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/>.

même groupe⁴. Le projet « Jurys », dont il sera question plus loin et sur lequel nous travaillions à l'époque du développement de la base MLQ, a attiré notre attention sur les relations interpersonnelles directes et la nécessité de les « coder » d'une manière ou d'une autre au sein de la base MLQ. Cela fut fait dans une table des relations qui permet à l'heure actuelle de lister directement, d'un clic dans l'interface produisant une requête programmée, les 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.

Structure et contenu

Venons-en à la description détaillée de la base de données MLQ. Comme nous l'annoncions, celle-ci tient compte et est structurée en fonction de trois réalités de l'histoire du livre et de l'édition au Québec : 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 diverses informations associées aux enregistrements de ces tables principales. Bien que la base souligne les liens qui existent entre ces différentes entités (relations des acteurs entre eux, des acteurs avec les organismes, des organismes avec les lois, etc.), elle était surtout destinée à devenir un outil de stockage d'informations concernant divers attributs des entités décrites. Ces informations sont de diverses natures et ont, de fait, été regroupées en différentes catégories. Ainsi, par exemple, retrouve-t-on pour chacun des acteurs ciblés 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

⁴ 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.

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 encodées dans la base MLQ.

GRELQ-Individus

Identification	Pseudos	Métiers	Autres métiers	Formations	Résidences	Pnx	Appartenances	Sociabilités	Relations	Sources	Accueil	Déconnexion
Arbour, Paul-Émile												
Début	Fin	Secteur	Métier	Nom détaillé du métier	Organisme	Statut						
1913	1922	Imprimerie	[Indéterminé]	?	Arbour & Dupont liée	Employé						
Commentaire			On mentionne dans "Le Maître Imprimeur" qu'il était connu à cette date comme un chef très compétent.									
Début	Fin	Secteur	Métier	Nom détaillé du métier	Organisme	Statut						
1922	1946	Imprimerie	Directeurs	Gérant de l'imprimerie	Arbour & Dupont liée	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.									
Début	Fin	Secteur	Métier	Nom détaillé du métier	Organisme	Statut						
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.									
Début	Fin	Secteur	Métier	Nom détaillé du métier	Organisme	Statut						
24 février 1933	1946	Imprimerie	Directeurs	Gérant d'imprimerie	Association des maîtres imprimeurs de Montréal	Dirigeant						
Commentaire			Présent à l'assemblée de fondation.									

Légende : Une page de l'interface de consultation de la base MLQ. Ici, la liste des postes occupés par l'imprimeur Paul-Émile Arbour.

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 GRÉLQ⁵, 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.

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

Les avantages

Ce rassemblement dans une base de données des informations mises au jour par le projet MLQ facilite, en fait, grandement leur exploitation. Elles 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 prosopographies des milieux visés par le projet. Ainsi, par exemple, trouve-t-on dans l'illustration suivante la liste de l'ensemble des acteurs répertoriés dans la base du GRÉLQ, ayant suivi une formation universitaire et ayant travaillé dans le secteur de l'imprimerie :

GRÉLQ-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

Légende : Un exemple de résultats d'une recherche sur la base MLQ.

La possibilité de croiser les critères de recherche et d'ainsi créer en quelques clics de souris des sous-corpus d'analyse permet aussi, soulignons-le, de mettre en évidence des réalités surprenantes ou, du moins, 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 fructueuse⁶. 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 factorielles des correspondances, dont l'usage heuristique — voire interprétatif en sociologie des champs — s'est répandu en sciences humaines et sociales depuis plus de trente 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 prix.

On mesure avec la base MLQ combien l'utilisation d'une base de données, dans le cadre d'un projet socio-historique, peut amener d'importants profits tant heuristiques qu'herméneutiques. Ces profits ne sont d'ailleurs pas 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 départ⁷. 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 ne pas que les données se retrouvent 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.

⁶ Voir Björn-Olav DOZO, « La présence des juristes dans l'institution littéraire belge. Examen de l'évolution des trajectoires scolaires et professionnelles des écrivains dans l'entre-deux-guerres », *Textyles*, n° 31, 2007, p. 28-46, Björn-Olav DOZO, « Portrait statistique de l'écrivain journaliste en Belgique francophone entre 1920 et 1960 », *Textyles*, n° 39, 2010, p. 123-146 et Björn-Olav DOZO, « Structure de l'espace relationnel des auteurs francophones belges de l'entre-deux-guerres », Marie-Pier LUNEAU et Josée VINCENT, *La fabrication de l'auteur*, Québec, Nota Bene, 2010, p. 183-203.

⁷ 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.phalèse.fr/bdhl/bdhl.php>) est un bon exemple de reconversion d'usage d'une base de données.

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 « Jurys⁸ »). 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 PHVC (« 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 relationnels⁹.

La base de données devait donc mettre l'accent dans un premier temps 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 les participants des jurys associés à différents concours artistiques québécois de 1918 à 1943 (Prix David, Prix de l'ACJC, Prix Europe, etc.) et inséré ces informations dans la base de données. L'analyse de cette base a pu produire différents résultats, que l'on retrouve dans notre article (avec Michel Lacroix) détaillant cette étude :

⁸ Celle-ci est accessible à l'adresse <http://jurys.phvc.ca/> .

⁹ Cette étude a été soumise aux revues *OPUS. Sociologie de l'art* et *Redes* pour un numéro spécial commun sur l'art et les réseaux.

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 faibles¹⁰ », 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, on va 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

Cette base de données, contrairement à celle décrite ci-haut, se caractérise par la grande simplicité de sa structure que nous avons déjà, par ailleurs, présentée brièvement au tout début de cet article. Les diverses informations inscrites dans les quelques tables qui la composent le sont en fonction de son objectif principal : la recension des membres des jurys ciblés. On

¹⁰ 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.

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 inscrites diverses informations sur les acteurs impliqués dans 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 tout à fait constituer des passerelles vers d'autres bases.

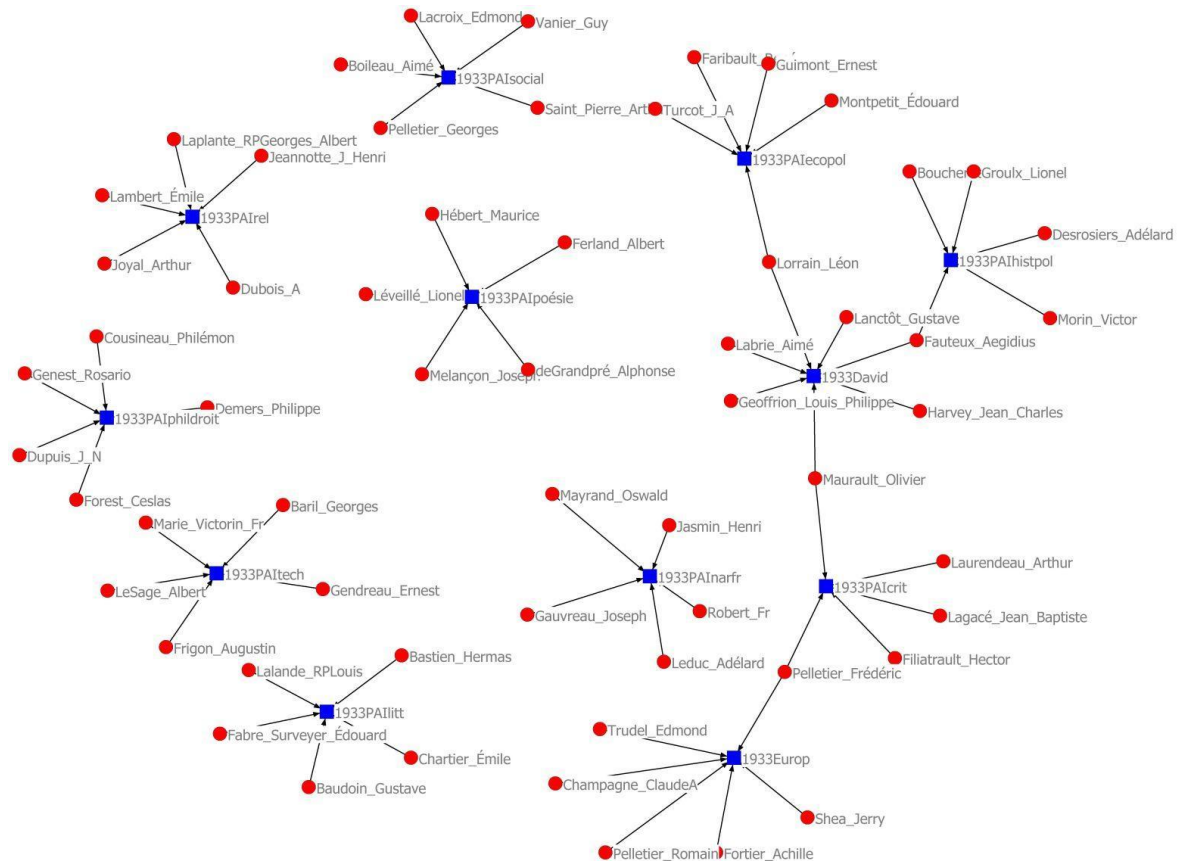
Il semble légitime de s'interroger, dès lors que l'on prend en compte la simplicité de la structure de cette base et des informations qui y ont été inscrites, sur la pertinence d'une telle entreprise. De simples fiches papier consignait, pour chacune des éditions des prix, les informations nécessaires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

Les avantages

En fait, ce qu'il faut bien comprendre ici, c'est que 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 « instituant »), tels 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 voir, à titre d'exemple, dans l'illustration suivante, un graphe relationnel montrant l'ensemble des relations entre les jurés des différents prix culturels canadiens-français de 1933 :



Légende : Graphe des réseaux formés par les membres des jurys des prix culturels et artistiques canadiens-français de 1933. PAI est l'acronyme de Prix d'action intellectuelle (de l'ACJC). PAIhistpol : Prix d'histoire et de politique, PAIlitt : Prix de littérature, PAIrel : Prix de littérature et sciences religieuses, PAItech : Prix de travaux scientifiques et techniques, PAIecopol : Prix d'économie politique, PAIcrit : Prix de critique littéraire et de critique d'art, PAInarfr : Prix de narration française, PAIsocial : Prix de sciences sociales, PAIActFr : Prix du Concours dramatique de l'Action française, PAIphil droit : Prix de philosophie et de droit, David : Prix David, Europ : Prix d'Europe.

En-dehors de 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, et ce, 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 était membre le sociologue Arthur Saint-Pierre.

PHVC-Jurys

Visualisation Recherche Gestion B.D. Wiki

Visualisation des acteurs répertoriés

Saint-Pierre, Arthur ▼

Prénom:	Arthur
Nom:	Saint-Pierre
Date de naissance:	1885-09-30
Date de décès:	1959-01-01
Occupation(s):	Journaliste, sociologue, professeur
Jurys liés:	1923-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu 1925-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu 1926-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu 1927-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu 1928-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu 1929-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu 1930-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu 1933-01-01-Prix d'action intellectuelle - Prix de sciences sociales-Inconnu

Légende : L'interface de visualisation des informations associées aux acteurs recensés dans la base « Jurys ».

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.¹¹). Si dans ce cas-ci, l'usage de la base de données paraissait unique, la conformation de

¹¹ 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, coll. « Repères », 2008, en particulier les pages 41-42, mais plus généralement tout le troisième chapitre. Pour

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 (ce sont des 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 textes¹² ».

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 sociocritique, attentive aux médiations qui unissent ces configurations aux textes, discours et pratiques contemporains¹³.

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.

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

d'autres exemples de constitution de bases de données, voir aussi Benoît HABERT, *Construire des bases de données pour le français*, t. 1, Paris, Ophrys, 2009.

¹² GREMLIN, « Fictions, figurations, configurations : introduction à un projet », GREMLIN (dir.), *Fictions du champ littéraire, Discours social*, volume 34, 2010, p. 3.

¹³ *Ibid.*, p. 4.

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 de 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, etc¹⁴.

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 comme des individus isolés, 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 du projet. L'interface de consultation et de modification des informations de la base de données est divisée en différentes sections et sous-sections qui correspondent directement aux diverses sections de cette fiche de lecture fournie aux différents membres du projet. Bien qu'elle contienne un certain nombre d'informations factuelles concernant l'auteur et l'œuvre (année d'édition, éditeur, etc.), cette base de données est essentiellement composée d'extraits textuels tirés des œuvres étudiées. Ainsi, par exemple, retrouve-t-on dans l'onglet « Positions » de la section du site consacrée aux personnages romanesques recensés par le projet, l'ensemble des extraits textuels tirés des romans du corpus et qui mettent en scène les prises de position (esthétiques, littéraires, politiques, etc.) de tel ou tel personnage.

¹⁴ *Ibid*, p. 14-15.

Figurations-Textes-Le Prix Lacomyne-Paul Le Raive

Accueil					
Textes					
Gestion des listes					
Gestion des usagers					
Recherche					
Modifications					
Déconnexion					
Identification					
Résumés					
Énonciation					
Personnages					
Configurations					
Lieux					
Textes					
Sphères					
Sources					
<input type="radio"/> Catégories <input type="radio"/> Activités culturelles <input type="radio"/> Descriptions <input type="radio"/> Langage <input checked="" type="radio"/> Positions <input type="radio"/> Capital <input type="radio"/> Modalités <input type="radio"/> Environnement					
Type ▲ ▼	Thème ▲ ▼	Extrait	Début ▲ ▼	Fin ▲ ▼	Commentaire
Esthétique	L'art	« L'art n'est pas un cornet de beringots. »	23	23	Une des nombreuses (et ridicules) opinions énoncées par Paul Le Raive dans un interview accordé au journal <i>Tant et Plus</i> .
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

Légende : L'illustration présente la page de visualisation des informations sur les prises de position des personnages romanesques ciblés par le projet. On y découvre une série de prises de position relativement loufoques énoncées par Paul Le Raive, l'un des deux principaux protagonistes du roman *Le Prix Lacomyne* de Renée Dunan.

Comme mentionné, cette base constitue la version informatisée d'une fiche de lecture assez traditionnelle, orientée vers les enjeux spécifiques du projet. Quel bénéfice le projet peut-il tirer de cette mise en série de textes et, plus largement, qu'est-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. Ainsi, par exemple, trouve-t-on 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.






Figurations-Recherche

Accueil	Textes	Gestion des listes	Gestion des usagers	Recherche	Modifications	Déconnexion
Simple	Avancée-Textes	Avancée-Personnages				

Critères de recherche :

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

Résultats de la recherche : 5 textes concernés

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

[Nouvelle recherche](#)

Légende : Les résultats d'une recherche dans la base « Figurations ».

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, l'encodage des données permet 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 pourraient ou ont pu apporter un éclairage complémentaire. On peut, de plus, imaginer tout ce que pourrait révéler l'exploration par des outils d'analyse lexicographique de l'impressionnante masse d'extraits textuels regroupés dans cette base.

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

Angenot¹⁵. 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 celles 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 figures, trois exemples de la façon dont les bases de données peuvent venir appuyer des projets de recherche en sociologie 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 à l'élaboration, dans le cadre de notre participation au projet de recherche « Penser l'Histoire de la vie culturelle », d'une autre base de données pluridisciplinaire, plus ambitieuse que la base « Jurys », 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 vingtième 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

¹⁵ Voir, en particulier, Marc ANGENOT, 1889. *Un état du discours social*, Longueuil, Le Préambule, 1989.

francophone et anglophone des arts de la scène à Montréal entre 1900 et 1950 : musique, théâtre, danse ». On y trouve des informations sur les œuvres et les acteurs associés aux milieux littéraires, musicaux et artistiques du Québec de la première moitié du vingtième 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 accessible 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 cette base « Émergence » est de rendre accessible des informations, de communiquer 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, à notre avis, 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 schéma¹⁶ 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 en sociologie de l'art de la culture présentement en activité. L'adhésion d'une partie de la communauté universitaire à ce paradigme, ainsi qu'aux diverses technologies qui lui sont associées, dont le standard RDF (Resource description framework) et les microdonnées, permettraient 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 participants. 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 de l'art et de la culture présentes et à venir.

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