



DES BASES DE DONNEES DE THESE : QUEL DEVENIR POUR UNE BASE « ARTISANALE » ?

Florence BONIFAY (U. Lyon 2)
Miriam SPEYER (U. Paris Cité)

L'essor des Humanités numériques a profondément changé la recherche en littérature, a fortiori en littérature de la Première Modernité, ne serait-ce que par l'accessibilité des textes, désormais massivement numérisés par la BnF ou Google books. Naturellement, ce changement se manifeste aussi dans les sujets de thèse : il est possible de s'emparer de corpus plus importants ou d'envisager des corpus constitués sous de nouveaux angles.

L'étude de ces corpus littéraires nécessite le recours à des outils de recherche innovants. La base de données, jadis réservée aux chercheurs en sciences, puis utilisée en linguistique et éventuellement en histoire et en sciences sociales, a fait son entrée dans le champ des études littéraires. Celle-ci ne constitue pas, dans ces cas-là, un « résultat » de la recherche doctorale, mais un instrument parmi d'autres, au même titre que l'outillage traditionnel de l'analyse littéraire. La base ne vise pas à remplacer l'approche littéraire, l'étude stylistique et rhétorique, elle vient la compléter. Dans un premier temps, le développement d'une base de données se justifie par la constitution du corpus d'étude et/ou de l'approche choisie, mais souvent, et c'est le cas dans les deux thèses qui feront l'objet du présent article, la base présente un intérêt pour la communauté des chercheurs. Se pose ainsi, tôt ou tard, la question d'une publication de l'outil, donc de son accessibilité, de sa visibilité et de sa pérennité.

Notre réflexion s'appuiera sur les bases de données relationnelles que nous avons élaborées dans nos thèses, consacrées respectivement aux citations entre poètes entre 1549 et 1586 et aux recueils collectifs de poésies publiés entre 1597 et 1671.

Les citations entre poètes au cours de la période 1549 et 1586

La base de données *RéseauxPoètesXVI* de Florence Bonifay est née dans le cadre de la thèse *Concurrences poétiques. Identités collectives et identités singulières autour de la « Pléiade »* (1549-1586) (Univ. de Lyon, 2016). L'appel d'Emmanuel Buron à ne plus utiliser le « terme équivoque » de « Pléiade » « par lequel les historiens de la littérature désignent l'école de Ronsard »¹, croisé avec la proposition de Caroline Trotot de dater de la Renaissance – dans les traités poétiques des années 1550 qui consacrent « la différenciation entre le poète et l'orateur »² et « reconnaissent la spécificité de l'écriture pratiquée par les poètes »³ – les

¹ Emmanuel Buron, *Dictionnaire des lettres françaises. Le XVI^e siècle*, 2001, Paris, Livre de Poche, p. 953 : « Pléiade. Terme équivoque, et qu'il serait utile d'abandonner, par lequel les historiens de la littérature désignent l'école de Ronsard ».

² Caroline Trotot, « La détermination du champ littéraire entre rhétorique et poétique autour de la Pléiade » dans Pierre Chiron et Francis Claudon (dir.), *Constitution du champ littéraire. Limites, intersections, déplacements*, Paris, L'Harmattan, 2008, p. 192. Voir aussi : Jean-Charles Monferran, *L'École des Muses. Les arts poétiques français à la Renaissance (1548-1610)*. Sébillet, Du Bellay, Peletier et les autres, Genève, Droz, 2011. En ce qui concerne la notion de « spécificité », Bernard Lahire invite bien à faire la différence entre l'autonomie-indépendance vis-à-vis des pouvoirs politiques, religieux ou économiques – qui n'est en fait jamais acquise pour la littérature – et l'autonomie-spécificité « comme domaine particulier d'activité qui se distingue d'autres



prémises de la constitution d'un « champ littéraire » conduisaient à plusieurs interrogations : si l'on ne peut pas parler de « Pléiade », qu'y a-t-il comme collectif ou comme imaginaire collectif qui ne soit pas seulement celui de Ronsard – seul, quasiment, à avoir utilisé le mot « Pléiade » ? Comment les poètes se sont-ils eux-mêmes représentés en groupe ? comment ont-ils mis en scène et désigné la sociabilité poétique de leur temps ? comment ont-ils défendu une « autonomie-spécificité » (Lahire), non pas dans les traités poétiques mais dans les recueils de poésie par le biais de leur mise en scène comme groupe spécifique ? De ces interrogations est née la volonté de construire un corpus qui ne serait pas composé de quelques recueils complets analysés dans leur intégralité, mais qui serait constitué de pièces piochées dans un grand nombre de recueils imprimés, pour leur capacité à témoigner de liens affichés entre poètes. Et ce, sur une longue période pour pouvoir analyser les évolutions de la mise en scène des sociabilités littéraires depuis 1549 (publication de *La Deffence*) à 1586 (année qui suit la mort de Ronsard et voit paraître des *Tombeaux* en son honneur qui offrent un bilan de trente-cinq ans de production poétique non seulement de Ronsard mais aussi de ses contemporains). La base de données *RéseauxPoètesXVI* rassemble alors quelque 2 200 textes issus d'environ 310 recueils et permet d'interroger les liens affichés entre auteurs : qui cite qui ? quand ? dans quel recueil ? en quelle langue ?

Les recueils collectifs de poésies au XVII^e siècle

La base de données relationnelle *Refaire Lachèvre*⁴ de Miriam Speyer est née dans le cadre de la thèse « *Briller par la diversité* » : *les recueils collectifs de poésies au XVII^e siècle* (Univ. de Caen, 2019). Si l'importance des recueils collectifs dans la publication de la poésie au XVII^e siècle est désormais reconnue, ces publications, que Frédéric Lachèvre avait inventoriées dans sa célèbre *Bibliographie* (4 tomes, 1901-1905), n'avaient jamais fait l'objet d'une enquête systématique. L'étude de ces supports de publication exige à la fois la prise en compte d'un échantillon représentatif de recueils et celle d'un échantillon représentatif de pièces. Or, les recueils réunissant souvent plusieurs centaines de pièces, leur nombre exclut une étude systématique des textes qu'ils contiennent. Aussi a-t-il été nécessaire de trouver un moyen d'isoler les textes les plus représentatifs. Dans cette opération, une base de données relationnelle constituée à partir des tables des matières des recueils s'est avérée précieuse : les recueils partageant une partie de leur répertoire, l'outil numérique, qui contient les métadonnées⁵ de plus de 60 recueils et de 10 000 pièces poétiques, permettait d'identifier ces pièces récurrentes, qui constituent en fait les « tubes » du moment, à l'époque de leur publication.

Le développement de ces outils de recherche n'est pas sans influence sur le déroulement du doctorat. Le recensement des textes et leur saisie dans l'outil constituent un investissement temporel notable qui s'ajoute aux recherches et études disciplinaires. Or, l'articulation de ces deux versants dans le temps limité du doctorat ne va pas de soi. Une fois mises en place, ces bases sont confrontées aux difficultés propres à l'outil numérique et à sa publication. Confidentielles au cours du doctorat, elles complètent le manuscrit de la thèse et

domaines » – qui n'a pas attendu le XIX^e siècle pour s'affirmer. Voir Bernard Lahire, *Monde pluriel. Penser l'unité des sciences sociales*, Paris, Seuil, 2012, p. 80.

³ *Ibid.*, p. 206.

⁴ Nom susceptible d'évoluer.

⁵ Ces métadonnées concernent le titre du recueil, les titres thématique et rhématique de la pièce, son incipit, la signature, la page à laquelle elle se trouve ainsi que sa forme d'écriture (en vers, en prosimètre ou en prose). La base se comprend comme un reflet fidèle de la présentation imprimée des pièces. Ainsi, si une pièce n'était signée que d'initiales, celles-ci n'ont pas été développées, ni de fausses attributions corrigées. D'éventuelles précisions concernant une pièce ont été consignées dans une entrée à part, consacrée aux remarques.



le livre qu'il va devenir. Ouvrant la voie à des investigations variées, ces bases de données représentent un réel intérêt pour la communauté des chercheurs. Du fait de la plasticité caractéristique des bases de données relationnelles, celles-ci peuvent être mises à contribution dans des enquêtes autres, voire devenir des incubateurs de travaux de recherche à venir. Se pose ainsi la question de leur devenir et de leur pérennité, questions d'autant plus complexes que le jeune chercheur dispose rarement des infrastructures et compétences techniques nécessaires.

À partir des expériences que nous avons pu faire avec les bases *RéseauxPoètesXVI* et *Refaire Lachèvre*, il s'agira, dans le présent article, de questionner ces difficultés rencontrées et de proposer, à défaut de solutions, des pistes de réflexion sur le devenir de ces « bases artisanales » nées dans le cabinet du doctorant.

I LA BASE DE DONNEES DE THESE : LA NECESSITE D'UNE SOLUTION AD HOC

La mise en place d'une base de données relationnelle répond à une spécificité du corpus étudié. Celui-ci, dont la cohérence – comme pour tout travail de recherche – a été vérifiée et justifiée en amont, se caractérise par une grande variété, par la quantité des sources ou encore leur dispersion, qui amènent le chercheur à recourir à des outils de travail numériques. Qu'elle ait ou non été conçue initialement sous la forme d'une liste tabulaire sous Microsoft Excel, la base de données représente d'abord un répertoire, un inventaire des textes à examiner, une forme de bibliographie raisonnée du corpus primaire.

La constitution de cette base de données et la forme qu'elle prend sont influencées par des contraintes institutionnelles inhérentes à la recherche doctorale. Il s'agit d'abord de contraintes temporelles. Par rapport au cadre idéal d'un doctorat en trois ans, la conception de la base prend du temps et peut avoir une incidence sur la durée de thèse, d'autant plus que, une fois développée et nourrie, la base de données ne constitue pas un résultat de la recherche mais simplement un support de travail : toute l'analyse du contenu reste à faire⁶ ! D'un autre côté, une fois opérationnelle, la base permet d'augmenter la rapidité et l'efficacité du traitement analytique des données pendant la phase de rédaction de la thèse. Le bon équilibre doit alors être trouvé entre la fascination pour le chiffre – c'est-à-dire la volonté d'engranger toujours plus de données brutes –, et le goût pour l'analyse littéraire – c'est-à-dire l'entrée dans le texte et dans le temps long de l'étude. En somme, le jeune chercheur doit, à un moment, décider que les données réunies dans la base composent un corpus suffisamment riche et cohérent pour autoriser des analyses pertinentes, et s'imposer de cesser la saisie de données. Pour nous deux, cette phase de création et d'enrichissement d'une base de données a duré entre un an et demi et deux ans, une période qui certes permet de se familiariser avec les textes du corpus et de mettre à l'épreuve la définition de celui-ci, mais qui laisse peu de temps pour une analyse littéraire approfondie.

Une deuxième contrainte pèse sur le doctorant : la contrainte matérielle. L'apprenti chercheur dispose rarement de moyens financiers spécifiques alloués pour un développement informatique précis, contrairement aux projets de recherche collectifs de grande envergure de type ANR ou ERC. Par ailleurs, seules les grandes équipes de recherche bénéficient du soutien d'ingénieurs d'étude et de recherche et, si ingénieur d'étude il y a, il est plutôt mobilisé par des

⁶ C'est là un constat régulièrement soulevé par les théoriciens des humanités numériques : sans une interprétation informée par le chercheur, les données quantitatives n'ont guère de valeur. Franco Moretti écrit ainsi : « La recherche quantitative fournit un type de données qui est idéalement indépendant des interprétations, [...] et c'est ce qui constitue également ses limites : elle fournit des *données*, pas de l'interprétation » (*Graphes, cartes et arbres* (2006), trad. E. Dobenesque, Paris, Les Prairies ordinaires, 2008, p. 41). Voir aussi Christophe Schuwey, *Interfaces. L'apport des humanités numériques à la littérature*, Neufchâtel, Livreo-Alphil, 2019, notamment chap. 8 « Nouvelles visualisations ».



travaux de grande ampleur. En ce qui nous concerne, lorsque nous avons commencé nos thèses, le système D s'imposait donc : apprentissage autodidacte de quelques rudiments en langage informatique, sollicitation d'amis de bonne volonté. Quoique nous ne nous soyons pas découragées, nos débuts balbutiants, puis notre parcours semé de quelques embûches techniques, confirment ces propos de Claire Ducourneau et Anthony Glinoyer : « Lorsqu'elle est une démarche individuelle, la constitution d'une base de données engage des difficultés de départ et des errances potentiellement décourageantes »⁷. Ces contraintes ne sont cependant pas toutes automatiquement des inconvénients. Le fait qu'il s'agisse d'une recherche personnelle, par exemple, permet d'avancer beaucoup plus vite, car cela nécessite peu de concertation, n'engage pas d'attente liée à une chaîne institutionnelle de traitement, ni de lourde élaboration de cahier des charges. En somme, la souplesse prévaut.

Le cas RéseauxPoètesXVI

Puisque l'intérêt se portait vers les réseaux d'adresses entre poètes dans leurs recueils publiés (gageant, par là, d'une certaine dimension publicitaire), il fallait sortir de la classique analyse de recueils pour extraire uniquement les pièces témoignant d'une mise en scène des relations entre pairs. La volonté était d'élargir la lecture bien au-delà d'une poignée de poètes dits de la « Pléiade » pour rendre compte d'un vaste réseau d'échanges. Toutefois, la poésie adressée étant une constante de la poésie de la Renaissance, et le nombre d'auteurs dépouillés n'ayant cessé d'augmenter au cours des premiers mois de recherche pour rendre compte de la diversité des positionnements (poète / poétesse, ancrage à Paris / ancrage en province, générations qui font leurs débuts dans les années 1550 / 1560 / 1570 / 1580), le nombre de pièces à étudier est monté à plusieurs centaines et cela a exigé le recours à un outil informatique.

En effet, l'intention n'étant pas de s'en tenir à des études quantitatives, les textes qui intéressaient le sujet ont été recopiés intégralement (ou copiés-collés, lorsque cela était possible). Ce sont ainsi plusieurs centaines de textes dans lesquels un poète s'adresse à / évoque / traduit un ou plusieurs confrères qui ont été saisis et enregistrés dans des fichiers Word distincts et clairement étiquetés (nom du recueil + nom du poème) ; puis ces fichiers ont été rassemblés dans des « dossiers » par auteur. La masse de données est rapidement devenue ingérable : comment naviguer efficacement entre ces centaines de textes enregistrés dans des fichiers distincts ? Une base MySQL, programmée en PHP par un compagnon informaticien dévoué, a alors vu le jour. Elle permet non seulement un archivage des textes mais également un enregistrement des données relationnelles : qui s'adresse à / évoque / traduit qui ? à quelle date ? dans quel recueil ? publié chez quel éditeur ? en quelle langue ?

⁷ Claire Ducourneau et Anthony Glinoyer, « Arpenter la vie littéraire. Les bases de données dans les études sociohistoriques sur l'imprimé et la littérature », dans *Arpenter la vie littéraire*, numéro thématique de la revue *Biens Symboliques / Symbolic goods*, n°2, février 2018, en ligne : <https://www.biens-symboliques.net/21>

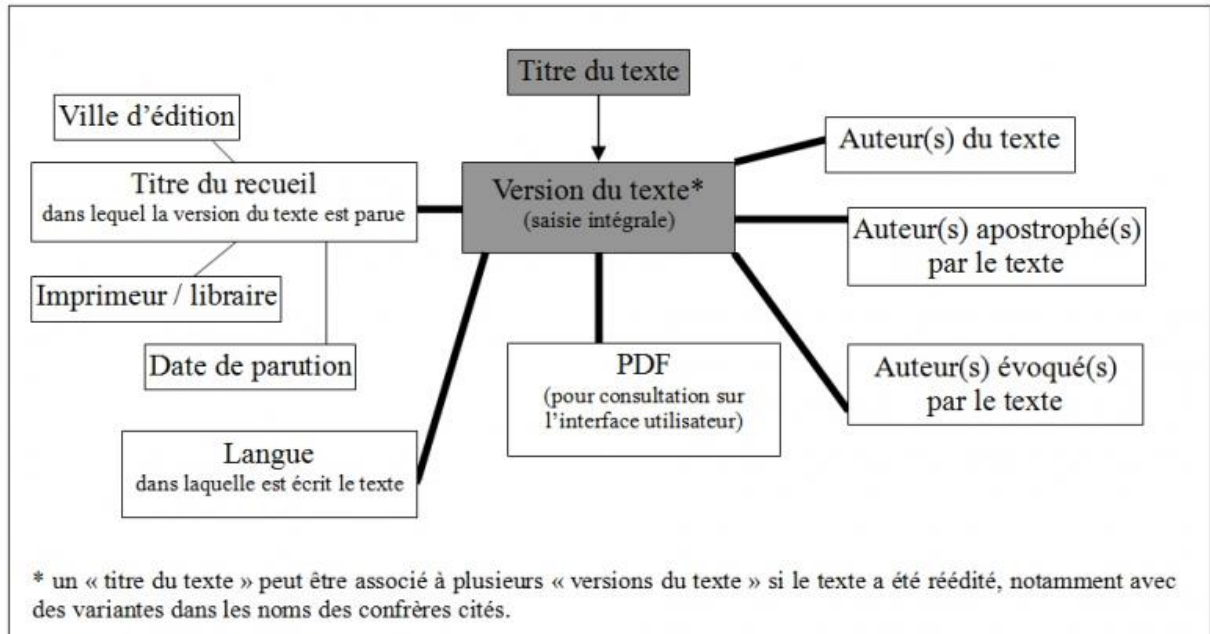


Figure 1. Organisation des données dans la base RéseauxPoètesXVI.

Dans l'interface usager (<https://sites.univ-lyon2.fr/ReseauxpoetesXVI/front/index.php>), on trouve, à l'onglet « Recherche », un moteur de recherche avancé qui permet de faire de multiples requêtes pour naviguer entre les quelque 2 200 textes de la base. De la sorte, *RéseauxPoètesXVI* permet :

- de rendre compte des réseaux de citations entre poètes ;
- de faire des études sur l'évolution des pratiques d'adresse aux pairs dans la carrière d'un auteur ;
- d'accéder à l'ensemble des textes échangés entre l'ensemble des poètes / un nombre choisi de poètes dans les recueils publiés à un instant T (une année précise) ou sur une période qu'on peut définir (ex. : entre 1549 et 1552) ;
- de demander l'ensemble des textes de la base qui citent 3/4/5/6/7/8/9/... noms de confrères pour travailler sur la représentation des poètes en groupe et la symbolique des chiffres ;
- d'accéder à des données (auteur.s ou autrice.s, année.s, recueils, fréquence d'adresse, nombre de noms cités dans les poèmes-listes, palmarès des destinataires les plus importants pour un auteur, etc.) et d'accéder à des textes, donc de lier analyses quantitatives et analyses qualitatives / littéraires.

La saisie des textes (et non pas seulement d'un ensemble de données du texte) s'est, en effet, d'emblée imposée pour garantir la dimension littéraire de l'étude. Ce choix permet, en outre, de donner à lire des textes rares qui n'ont pas d'édition moderne⁸. Car donner à lire, et donc donner accès, a été une intention des premiers instants. La base a ainsi été pensée dès le départ comme un outil à partager après la soutenance de thèse. Si les textes n'avaient pas été saisis, la crainte était celle d'un enregistrement de données assez ésotériques, difficilement utilisables ensuite par la communauté seiziémiste. Ce choix de saisir les textes s'est *a posteriori* révélé judicieux, car les recherches plein texte que cela autorise pallient le manque de certains

⁸ Par exemple, des textes issus des recueils de Jean de Boissières, poète auvergnat originaire de Clermont-Ferrand, dont ont été dépouillées *Les Premières œuvres amoureuses* de 1578, *Les Secondes œuvres poétiques* de 1578 et *Les Troisièmes œuvres* de 1579, dans le cadre de ce travail de thèse.



critères de recherche qui n'ont pas été originellement imaginés et qui ne sont, pour le moment, pas disponibles pour interroger la base (ex. : critère de l'ancrage géographique des mises en scène groupales : les poètes « du Clain » / de Poitiers ; les poètes lyonnais / mâconnais, etc.).

Si l'augmentation du corpus pendant les premiers mois de doctorat a rapidement entraîné la nécessité d'une solution informatique *ad hoc* pour gérer efficacement cette masse de données, en retour, le développement d'un outil informatique performant a favorisé la poursuite gourmande de l'enrichissement du corpus. Ainsi, le dépouillement exponentiel de recueils autorisé par l'outil capable de naviguer entre un grand nombre de données extraites de ces recueils, a facilité la constitution d'un corpus d'étude allant bien au-delà du canon de la « Pléiade » pour explorer les marges. C'est, en effet, un des atouts des humanités numériques souligné par Claire Ducourneau et Anthony Glinoyer : « La mise à disposition d'une quantité très importante de données permet [...] de traiter également le corpus littéraire canonique et des textes méconnus, sans avaliser directement les différences de légitimité existant entre eux »⁹. De la sorte, ce travail de thèse a pu explorer un réseau de sociabilité large avec plusieurs noyaux locaux, par exemple, et en faisant une place aux femmes.

Le cas *Refaire Lachèvre*

Le catalogue des incipit que propose la base *Refaire Lachèvre* permet de rendre compte des recueils collectifs au XVII^e siècle. À cette époque, la compilation constitue en effet le mode de publication majeur pour des nouveautés poétiques. Quand l'inventaire de F. Lachèvre rend bien compte de l'importance de ce genre éditorial, sa forme et sa structure ne permettent pas de suivre la circulation des textes, d'isoler les pièces les plus représentatives, de faire une étude diachronique des formes et des genres poétiques empruntés ou encore d'élaborer des visualisations. Le catalogue des incipit, développé d'abord sous la forme d'un tableau Excel, puis d'une base de données relationnelle en SQL¹⁰, au contraire, ouvre la voie à une exploitation multidimensionnelle du corpus. Celle-ci est plus fidèle à la lecture que l'on faisait de ces recueils à l'époque : les compilations ne se lisaient pas de manière linéaire du début à la fin, mais constituaient plus probablement des répertoires de textes dans lesquels chaque lecteur puisait au gré de ses envies et intérêts.

⁹ Claire Ducourneau et Anthony Glinoyer, *art. cit.*

¹⁰ Le nombre d'informations réunies (> 10 000 lignes) ainsi que les contraintes liées au logiciel lui-même rendaient un traitement des données sous Excel impossible. C'est grâce à l'aide, le conseil et l'expertise informatique de Christophe Schuway que le fichier a pu être transformé en base de données relationnelle sous SQLite.

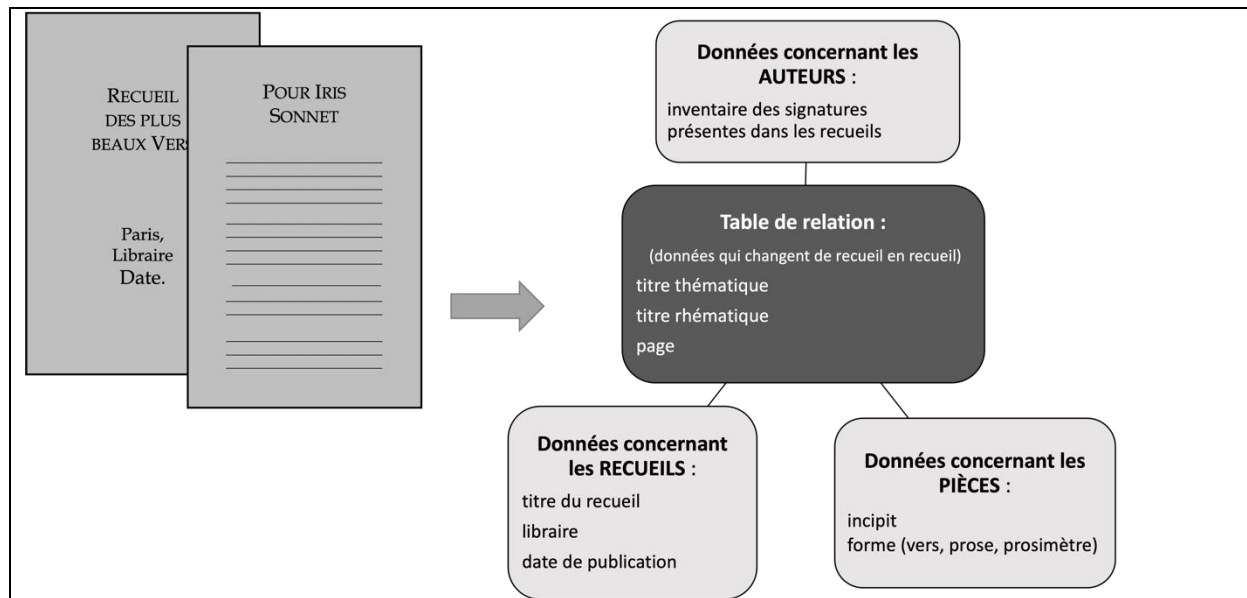


Figure 2. Schéma d'organisation de la base Refaire Lachèvre.

La transformation du logiciel de tableur vers la base de données sous *SQLite* entraîne cependant une perte de confort dans l'utilisation : les recherches sous *SQLite* ne se font pas à travers une interface *WYSIWYG*¹¹, un moteur de recherche, mais à l'aide de commandes en langage SQL. En plus du développement de la base, il était nécessaire de disposer d'une formation minimale à ce langage informatique pour pouvoir interroger la base¹².

Le langage SQL (*structured query language*) permet de formuler des requêtes multicritérielles et offre une souplesse nettement supérieure à celle qu'offrent les logiciels de tableur. Comme souligné en introduction, les divers recueils collectifs partagent une partie de leur répertoire. À une époque où la production d'un livre imprimé constitue un investissement considérable, la réimpression d'un ensemble de textes est un signe de leur succès. Pour connaître les pièces les plus représentatives du corpus, il était donc nécessaire d'identifier celles qui avaient été imprimées le plus souvent. Avec la base de données en *SQLite*¹³, ce repérage (par le recours à des fonctions comme *count()* et *group_concat()*) devenait aisé. À l'aide des fonctions *min()* et *max()* appliquées aux dates de publication, il est de plus possible de connaître la « durée de vie » de chacune des pièces et d'identifier, partant, des modes poétiques. L'étude proprement littéraire et stylistique des pièces ainsi isolées m'a permis de définir des styles poétiques représentatifs de certaines périodes du XVII^e siècle (le style des années 1600-1620, le style des années 1630, etc.). Ce travail, complété par des comparaisons chiffrées des genres poétiques pratiqués pendant cette même période, offre dès lors la possibilité de réécrire d'une certaine manière l'histoire de la poésie du XVII^e siècle : l'étude ne prend alors pas seulement en compte les œuvres et les auteurs canonisés, mais aussi des *minores* et des « oubliés », et ce sans établir de hiérarchie entre eux¹⁴.

¹¹ « *What you see is what you get* ». Une interface *WYSIWYG* dispense l'utilisateur de connaissances informatiques, l'affichage lors de la saisie correspondant au rendu final.

¹² À défaut de formations doctorales appropriées, cet apprentissage s'est fait en autodidacte à l'aide de tutoriels disponibles en ligne.

¹³ L'avantage de *SQLite* réside notamment dans la particularité que l'ensemble des données est stocké sur le disque dur de l'utilisateur, ce qui permet de travailler hors ligne, et limite le risque de piratage. La plupart des autres logiciels de gestion de base de données en SQL, p. ex. *MySQL*, recourent au stockage en ligne et nécessitent, partant, une connexion internet pour l'exécution des requêtes.

¹⁴ Il s'agit là d'une des grandes chances offertes par les humanités numériques, comme le note aussi Christophe Schuwey. L'exploration de grands corpus par des moteurs de recherche « sape [...] l'autorité du canon, de



La thèse « *Briller par la diversité* » met aussi l'accent sur la circulation de certaines pièces en particulier et les changements qu'elles peuvent subir lorsqu'elles passent d'un éditeur à un autre (changement de titre, présence/absence de signature, place dans le recueil, ...). Pour le milieu du siècle, ont pu être identifiées les pièces qui se lisaient tant à Paris qu'en province et en Hollande et celles qui résistaient à l'« exportation » hors de la capitale, ce qui a permis de mieux cerner les attentes et les goûts des lecteurs en fonction d'un lieu géographique. Mais il ne s'agit là encore que d'une petite part des possibilités offertes par la base de données *Refaire Lachèvre*.

À partir des données encodées, d'autres requêtes sont possibles. Là où le répertoire de F. Lachèvre attirait l'attention sur l'auteur et la présence ou absence des signatures, la base de données laisse au chercheur la liberté de choisir son point d'entrée dans le corpus, qui peut être l'auteur, mais aussi le libraire, la ville de publication, le genre poétique, la forme d'écriture (vers, prose, prosimètre) ou une pièce en particulier. En fonction de la formulation de la requête, la base de données en *SQLite* permet à l'utilisateur d'identifier des sous-corpus pour ensuite les étudier d'un point de vue littéraire. C'est pourquoi l'outil présente un intérêt évident pour la communauté des chercheurs en littérature du XVII^e siècle.

Nous l'avons dit, le travail solitaire et artisanal, en dehors de toute contrainte institutionnelle, permet un gain de temps initial qui peut faire fi d'une étape lourde de concertation et de validation collectives. Nos bases de données témoignent de la capacité de doctorantes, dans le cadre d'un travail individuel, à mettre en relation sur un temps relativement court une quantité de données massive. Toutefois, selon l'adage : si « seul on va plus vite, à plusieurs on va plus loin ». Ce gain de temps initial et cette productivité indéniables peuvent, *a posteriori*, occasionner quelques interrogations lorsqu'il s'agit d'ouvrir la base à un usage collectif, car la pérennité d'un tel travail personnel, comme sa publication ne vont pas de soi.

II LA BASE DE DONNEES DE THESE APRES LA THESE (1) : QUELS PUBLICS ?

De la même façon que les fiches réalisées à la main risquent de ne plus faire qu'un tas de papier dépourvu de signification après la retraite ou la mort de leur(s) concepteur(s) ou conceptrice(s), les bases de données informatisées peuvent devenir inutilisables faute de machine en état de les lire ou, pire, disparaître une fois le projet de recherche terminé¹⁵.

Ce problème de la pérennité des outils de recherche soulevé par Claire Ducourneau et Anthony Glinoyer se présente de manière particulièrement nette dans le cas des bases de données de thèse. Au plus tard à l'issue des études doctorales se pose la question de leur devenir. À défaut d'une valorisation, les bases risquent en effet de n'être que d'éphémères outils personnels, et ce alors qu'elles constituent le résultat d'un investissement intellectuel et temporel important. Or, leur valorisation ne paraît pas seulement souhaitable, mais nécessaire. Outils intrinsèquement liés aux conclusions exposées dans le manuscrit, les bases de données réalisées constituent un gage scientifique. Elles permettent la contextualisation de la méthode de travail et la vérification des résultats. C'est pourquoi il est indispensable de les rendre accessibles aux membres du jury de thèse. Par la suite, une publication est à souhaiter : la base complète l'ouvrage que deviendra la thèse une fois publiée.

l'histoire littéraire traditionnelle et des hiérarchies établies, puisqu'elle met sur le même plan des ouvrages bien connus et des fonds de bibliothèques oubliés ». (*Interfaces, op. cit.*, p. 79).

¹⁵ *Ibid.*



Plus généralement, les bases artisanales peuvent aussi avoir un intérêt pour des recherches qui dépassent le cadre et la problématique de la thèse. Du fait de sa souplesse constitutive, la base de données relationnelle se prête à une exploitation multidimensionnelle. Les données engrangées peuvent être interrogées selon des angles et des questions de recherche innovantes qui diffèrent de la problématique de la thèse et présentent, de ce fait, un intérêt pour la communauté des chercheurs.

Le cas *RéseauxPoètesXVI*

RéseauxPoètesXVI, par exemple, est un des outils de travail qu'a mobilisé Mathilde Vidal pour sa thèse sur l'étréne¹⁶. Elle l'a aidée à constituer son corpus d'étude. Michèle Clément, autre exemple, a mobilisé les ressources de la base pour préparer un travail sur Jacques Grévin qui explore, entre autres, ses échanges avec Jean Tagault. Ces usages de la base en ligne, suite à son ouverture publique, confortent la conceptrice dans l'idée qu'il était pertinent de penser d'emblée *RéseauxPoètesXVI* pour un usage collectif et donnent de l'énergie pour surmonter toutes les difficultés techniques liées à sa pérennisation sur le web¹⁷.

Le cas *Refaire Lachèvre*

La base « Refaire Lachèvre » est en cours de publication. Étant donné qu'elle n'est donc pour l'instant pas accessible à la communauté des chercheurs, des requêtes spécifiques ont été effectuées parfois sur leur demande. Le travail de Mathilde Vidal sur l'étréne, là aussi, en est un exemple. Ces interrogations des collègues, thématiques, historiques ou génériques, sont en même temps une ressource précieuse pour mieux prendre connaissance des requêtes susceptibles d'être formulées par d'autres, et ainsi la réflexion sur le développement de la base en vue d'une publication.

Le catalogue des incipit que propose *Refaire Lachèvre* a permis d'enrichir les dossiers concernant certaines pièces, leur circulation et leur(s) éventuelle(s) réutilisation(s) dans d'autres contextes de publication en particulier (transmission manuscrite, insertion dans des fictions narratives, des pièces dramatiques, etc.). La base n'étant pas accessible en ligne et son utilisation exigeant, de plus, la maîtrise du langage SQL, il était à chaque fois nécessaire de procéder à la requête à la place des collègues. La fréquence et la diversité de leurs interrogations confirment l'intérêt que présente une publication de la base en ligne, complétée alors par une interface *WYSIWYG* proposant un moteur de recherche avancé.

Si une telle utilisation publique et partagée présente donc un intérêt certain, sa mise en œuvre n'est pas sans présenter des difficultés. Les bases *RéseauxPoètesXVI* et *Refaire Lachèvre* ont été conçues en fonction des besoins propres aux sujets de thèse. Contrairement à des bases de données conçues dans le cadre de projets recherche, comme celle des *Joyeuses Inventions*, elles n'ont pas été élaborées d'abord en vue d'une utilisation par la communauté scientifique. Il arrive ainsi que les données demandent des remaniements plus ou moins importants pour gagner en ergonomie et assurer la lisibilité des informations en vue du projet de thèse et de ses contraintes, notamment temporelles.

Si *RéseauxPoètesXVI* a déjà été mise en ligne, plusieurs remaniements seraient nécessaires pour en faciliter l'usage collectif. Des collègues ont ainsi pointé le fait qu'une requête sur un mot clé ou une expression est un peu fastidieuse. En effet, le résultat de la requête s'affiche sous la forme d'une liste de titres de textes sur lesquels il faut cliquer pour lire

¹⁶ Mathilde Vidal, « *Bon jour, bon an, et bonne estreine* » : *Poétique de l'étréne en vers de Marot à Scarron*, thèse s. d. Jean Vignes soutenue le 30/09/2019, Université Paris-Diderot.

¹⁷ Voir supra, III.



le contenu afin de retrouver ledit mot clé ou ladite expression. Afficher directement, en résultat, les extraits des lignes contenant le mot clé ou l'expression recherchés pourrait être plus commode. Le chercheur pourrait ainsi rapidement déterminer si le mot en contexte répond à sa curiosité ou non.

Par ailleurs, au départ, la base était pensée comme une ressource d'informations qui ne devait toutefois pas dispenser de se reporter aux éditions modernes ou anciennes du texte lors de l'exercice de citation dans la thèse. Ainsi, aucune harmonisation éditoriale n'a été pensée en amont. Lorsque le texte connaît une édition moderne de référence, c'est cette édition qui a été recopiée, suivant les choix de transcription faits par l'éditeur scientifique. Lorsque le texte ne connaît pas d'édition moderne, il a été choisi une saisie diplomatique de l'édition ancienne, avec toutefois les dissimulations u/v, i/j et le développement du tilde. L'ensemble des 2 200 textes de la base mériterait donc une sérieuse révision en vue d'une harmonisation qui opérerait pour une saisie diplomatique systématique à partir d'une édition ancienne de référence.

Ce sont autant de faiblesses qu'il faudrait corriger pour renforcer la scientificité de la base et son ergonomie.

Pour le cas de *Refaire Lachèvre*, la forme de certaines données encodées est à repenser. Les recueils n'ont pas été encodés avec les références bibliographiques complètes, mais à l'aide d'abréviations personnelles qui, si elles sont parfaitement claires et univoques pour la conceptrice, ne le sont pas pour l'utilisateur peu familier de la base. Une table supplémentaire, associant les abréviations aux références bibliographiques complètes a déjà été préparée, mais il serait nécessaire de la joindre à la base. Afin de faciliter l'accès aux recueils et aux pièces qu'ils contiennent, la base ne répertoriant pas les textes mais seulement les incipit, il serait en outre désirable de proposer, pour les recueils disponibles dans des bibliothèques numériques, un lien vers celles-ci.

Autre limite actuelle de la base : la forme des incipit. La base ayant été, dans un premier temps, spécifiquement conçue pour l'identification des pièces récurrentes, il était nécessaire que les pièces présentant un même incipit puissent être reconnues automatiquement. Afin de pallier les différences graphiques présentes dans les recueils d'origine, la modernisation de l'orthographe et de la ponctuation s'est imposée, qui a conduit à un nivellement des variantes. Ainsi, un incipit qui apparaît sous différentes versions a systématiquement été encodé avec l'ensemble des variations identifiées. La fameuse pièce de Bertaut, par exemple, dont l'incipit varie entre « Des maux si déplorables... » et « Les Cieux inexorables... », se trouve dans la base sous la forme « Des maux si déplorables (Les cieux inexorables) ». Si ce choix permet de trouver toutes les occurrences de la pièce, quel que soit l'incipit effectif, il ne permet pas de connaître la leçon présente dans tel ou tel recueil en particulier, et le chercheur doit, pour l'instant, se reporter au livre en question. Un enrichissement de la base par une colonne répertoriant les incipit non-modernisés, donc l'incipit effectif, est en cours de réalisation.

Ces exemples tirés de *RéseauxPoètesXVI* et *Refaire Lachèvre* montrent que la publication d'une base de données, quelle que soit sa mise en œuvre technique, nécessite une véritable réflexion sur la formalisation des données et la manière dont l'utilisateur y accède. En aucun cas, en effet, la publication d'une telle base de données ne peut se passer d'un paratexte explicitant les choix scientifiques de l'outil ou le corpus encodé. Nées dans un contexte de recherche précis, ces bases de données, même enrichies ultérieurement, ne visent pas l'exhaustivité, et présentent, par conséquent, des limites dont l'utilisateur doit être conscient pour tirer pleinement profit des résultats de sa recherche.

C'est ainsi que la base en ligne *RéseauxPoètesXVI* comporte un onglet « Présentation du projet » qui explique le contexte de création de la base, ses fonctionnalités et pointe ses



limites afin que le chercheur ne soit pas induit en erreur. On y trouve, par exemple, un avertissement sur l'interprétation des résultats fournis par la base : à la suite de la liste des auteurs qui ont été dépouillés de manière exhaustive, on peut lire « Tous les autres auteurs présents dans le corpus ont été dépouillés de manière partielle. Ce caractère partiel de certains dépouillements doit être pris en compte dans l'analyse des résultats fournis par cette base de données. ». C'est ainsi que si on recherche « tous les textes de... Jean Le Frère de Laval... adressés à... ou citant... Pierre de Ronsard », un résultat nul n'indique pas forcément une absence de référence à Ronsard dans les œuvres de Laval : c'est tout simplement que les œuvres de Jean Le Frère de Laval n'ont pas encore été dépouillées et saisies de manière exhaustive.

Mais la conception et la préparation d'une base de données en vue de sa publication n'exige pas seulement des choix scientifiques clairs concernant les données encodées. Ce travail implique également des choix technologiques qui, sans trahir la cohérence de l'outil, sont en mesure d'assurer son ergonomie et sa pérennité.

III LA BASE DE DONNEES DE THESE APRES LA THESE (2) : IDENTITE ET DEFORMATIONS

Concernant les outils numériques exploités, tout projet doit aujourd'hui choisir entre deux possibilités : l'utilisation d'une plateforme de publication générique (en France par exemple les solutions soutenues par HumaNum, comme Heurist ou Omeka, ou les structures développées par l'équipe de l'Obvil, par exemple)¹⁸, ou le développement d'une plateforme de publication *ad hoc* (par exemple ncd17¹⁹ ou le répertoire du théâtre imprimé d'Alain Riffaud²⁰). Une plateforme générique permet de se rattacher à des collectifs ; elle offre un hébergement peu coûteux, assure une certaine pérennité (aussi grâce à la maintenance technique) et garantit une relative visibilité. Toutefois, elle impose également des contraintes plus ou moins fortes liées à son cahier des charges. Inversement, un développement *ad hoc* offre plus de souplesse pour personnaliser l'objet numérique, mais il est onéreux et hasardeux en termes de visibilité et de pérennité²¹. À l'issue de sa soutenance, le jeune docteur qui veut faire connaître sa base de données de thèse doit donc faire un choix quant au lieu de valorisation de son travail numérique.

Parallèlement, il peut se poser la question du maintien de sa base dans les bornes fixées par le travail de doctorat, ou d'une modification de ses contours, éventuellement pour la fondre dans des projets voisins de plus grande envergure. Sortir la base de données d'un usage personnel tout entier tendu vers la soutenance d'une démonstration singulière, pour la faire rejoindre une dynamique collective de constitution et valorisation de ressources numériques utiles à une communauté de chercheurs n'est toutefois pas chose aisée. D'une part, parce que cela engage le deuil d'une certaine unité heuristique à laquelle s'est fortement attaché le jeune chercheur pendant ses années de doctorat. D'autre part, parce que la convergence intellectuelle entre le corpus du doctorat et les corpus des projets collectifs émergents est difficile à ajuster, malgré des accointances de départ. Enfin, parce que l'interopérabilité entre la

¹⁸ <https://www.huma-num.fr/> (consulté le 6 janvier 2020), <https://obvil.sorbonne-universite.fr/> (consulté le 28 octobre 2021).

¹⁹ <https://www2.unil.ch/ncd17/> (consulté le 6 janvier 2020).

²⁰ <https://repertoiretheatreimprime.yale.edu/> (consulté le 6 janvier 2020).

²¹ Christophe Schuwey aborde ce problème, qui n'est d'ailleurs pas propre au contexte scientifique français, du point de vue de la pérennité de l'interface : si celle-ci n'est pas conçue à l'aide d'une technologie normée et reconnue, elle est condamnée à disparaître à court ou moyen terme. Voir à ce sujet aussi Christophe Schuwey, *Interfaces, op. cit.*, chap. 10 « Les nouvelles missions des bibliothèques ».



solution technique développée pour la base de données de thèse et celle des projets collectifs en cours de développement est rarement évidente.

Le cheminement de nos deux bases de données « artisanales » et nos interrogations sur leur devenir illustrent ces deux dilemmes.

Le cas *RéseauxPoètesXVI*

RéseauxPoètesXVI a d'emblée été pensée comme une base de données utilisable par la communauté seiziémiste à l'issue de la soutenance de la thèse. Aussi a-t-il tout de suite été décidé de la déposer sur le web sous la forme d'un site autonome, dont l'accès a été d'abord verrouillé pendant les années de doctorat. C'est donc le choix d'un développement *ad hoc* qui a été fait, rendu possible par les compétences d'un compagnon informaticien. Le site web *RéseauxPoètesXVI* avec son interface permettant de naviguer entre les textes, appuyé sur la base de données en *back office*, a d'abord été déposé sur un hébergeur gratuit réputé fiable... qui a disparu du jour au lendemain, occasionnant la perte de deux mois de travail. *RéseauxPoètesXVI* a alors été déposé sur le serveur de l'université de rattachement, Lyon 2, moyennant quelques réajustements pour que l'interface soit compatible avec les contraintes du serveur de l'université. L'URL d'accès est donc <https://sites.univ-lyon2.fr/ReseauxpoetesXVI/front/index.php>. Cette migration a assuré une certaine sécurité mais n'a pas écarté toute vulnérabilité : ce serveur institutionnel a en effet, pendant quelques mois, fait l'objet d'attaques répétées de hackers, entraînant une suspension des services pour les sites non « natifs » comme *RéseauxPoètesXVI* ; une panne de climatisation de la salle des serveurs a également interrompu l'accès à la base en ligne. Par ailleurs, si l'hébergement sur le serveur de Lyon 2 garantit la pérennité, cela ne va pas sans l'obligation d'en suivre les évolutions techniques : une migration des serveurs de Lyon 2, et c'est le code informatique de la base qui doit être corrigé pour qu'elle continue à être affichable. Exercice de routine pour qui s'y connaît en la matière ; montagne de perplexité pour une thésarde littéraire qui n'a pas ces compétences, et qui a perdu en cours de route les compétences annexes du compagnon informaticien de départ... Enfin, le dépôt sur le serveur de Lyon 2 crée un lien fort à une institution et pose la question du devenir de l'hébergement de la base en cas de mutation professionnelle. Bénéficiant désormais de l'aide technique d'un ingénieur d'étude du laboratoire IHRIM, et suivant le mouvement d'investissement de la plateforme HumaNum soutenu par ce même laboratoire, Florence Bonifay a décidé de basculer *RéseauxPoètesXVI* sur HumaNum. L'opération est en cours et occasionnera un changement d'URL pour accéder à la base.

Parallèlement à l'expérimentation de ces fragilités et mutations techniques qui obligent à de perpétuelles migrations et mises à jour, on peut s'interroger sur le devenir scientifique de *RéseauxPoètesXVI*. Faut-il lui conserver son amplitude de corpus de thèse ? Faut-il plutôt en élargir les contours ? Doit-elle rester rattachée au nom de sa fondatrice, ou serait-il préférable qu'elle aille se fondre dans un projet collectif ? Trois pistes semblent explorables.

La première consiste à conserver l'identité forte de la base et à en garder le pilotage, en encourageant les contributions collectives pour enrichir son contenu. Quelques propositions d'ajouts de textes ont été reçues, par exemple de la part de Nina Mueggler ici chaleureusement remerciée. Toutefois, globalement, ce genre d'incitation suscite peu d'élans spontanés. De belles bases collaboratives, comme la base ELR (Éditions lyonnaises de Romans au XVI^e siècle) initiée par Pascale Mounier et Mathilde Thorel au début des années 2010, s'enrichissent grâce à l'intelligence collective, mais la plupart du temps suite à des demandes personnelles effectuées auprès des uns et des autres par les gestionnaires de la base.

La deuxième piste consiste à fondre cette base dans un projet collectif émergent ou à l'articuler à un projet collectif émergent. *RéseauxPoètesXVI* a ainsi suscité l'intérêt d'Élise Rajchenbach pour son édition numérique en cours des œuvres de Charles Fontaine et d'une



cartographie de ses réseaux. Il s'est toutefois rapidement avéré qu'il n'y avait pas vraiment de pertinence à fondre les deux objets car *RéseauxPoètesXVI* embrasse plusieurs dizaines de poètes sans qu'il y ait un centre unique du réseau, alors que le projet Fontaine veut rendre compte du réseau de relations de Charles Fontaine uniquement. *RéseauxPoètesXVI* ne peut être qu'une base amie de la future base des réseaux de Fontaine.

Enfin, une troisième piste empruntable serait que la conceptrice de *RéseauxPoètesXVI* engage elle-même un projet collectif à partir de cette base, ou du moins des textes contenus dans cette base, en changeant peut-être l'organisation, l'ergonomie, les choix concernant l'interrogation des données et l'affichage des résultats (par exemple, en adossant la base à un logiciel de visualisation des réseaux sous forme de graphes). Au-delà du tournant technique, ce projet collectif pourrait étoffer l'ambition heuristique de *RéseauxPoètesXVI*, d'une part en élargissant ses bornes historiques et/ou géographiques, d'autre part en ne la restreignant pas à la catégorie peut-être bancale des « poètes » mais en l'ouvrant à la catégorie plus large des « auteurs », permettant d'enregistrer les réseaux d'échange et d'adresses entre poètes, traducteurs, historiens, géographes, bref tout lettré qui publie. Cela ne ferait au fond que radicaliser les petites pistes déjà présentes dans *RéseauxPoètesXVI* puisqu'on y croise déjà ici ou là François de Belleforest cosmographe, Gabriel Chappuys²² traducteur, ou encore Simon Goulart²³ théologien. Toutefois, un projet d'une telle envergure ne peut être envisagé qu'avec un poste adéquat (MCF/CNRS).

Le cas Refaire Lachèvre

Les échanges avec des chercheurs spécialistes du XVII^e siècle ont montré qu'il existe une véritable demande pour la publication de la base, car les données qui y sont consignées permettent d'éclairer des recherches portant sur d'autres corpus/des corpus apparentés ou en fonction d'angles différents. À l'heure actuelle, la publication se heurte essentiellement à deux écueils : celui de l'hébergement et celui, partiellement lié, de l'interface de recherche.

En tant que membre de l'équipe *Joyeuses Inventions*, qui est en train d'élaborer la base *Lachèvre numérique XVI*, Miriam Speyer dispose en principe de la possibilité d'intégrer *Refaire Lachèvre*, sous le nom de *Lachèvre numérique XVII*, dans le projet. Le grand avantage de cette solution serait la possibilité de fondre *Refaire Lachèvre* dans une infrastructure numérique existante dont le suivi et la maintenance sont assurés, et cela à faible coût humain et financier. Or, les présupposés de recherche des deux projets divergent, comme le font aussi les corpus. Tandis que *Lachèvre numérique XVI* se construit, du moins actuellement, autour du noyau formé par les éditions du *Trésor des joyeuses inventions* et accorde une grande importance aux rééditions (en effet importantes au XVI^e siècle), *Refaire Lachèvre* recense les recueils collectifs généraux du XVII^e siècle et n'inventorie que certaines rééditions (le paysage éditorial des recueils vivant, autour de 1597, une reconfiguration). Une mise en commun des deux bases exigerait ainsi des transformations importantes de la base de données de thèse au risque de la priver de sa cohérence. S'il n'y a pas de doute que les compilateurs de recueils collectifs de

²² « Sonet dudit G. Chappuys Thourangeau. A Monsieur Buttet. » dans Harangue de Charles Paschal, Sur la mort de tres-vertueuse Princesse, Marguerite de Valois fille de François premier de ce nom Roy de France, qui fut espouse de tres-illustre Prince Emanuel Philebert, Duc de Savoye, Prince de Piedmont, adressee à son altesse. Traducte de Latin en François par Gabriel Chappuys Thourangeau. A Paris, Chez Jan Poupy, ruë S. Jacques à l'image S. Martin. 1574. Avec Privilege du Roy.

²³ « Au Seigneur Du Bartas » dans *La Sepmaine, ou creation du monde*, de G. de Salluste, Seigneur du Bartas. Genève, Jean Baptiste Pinereul, 1578. [édition pirate] : « Au lecteur, S.G.S. [Simon Goulart Senlisien] » dans *La Sepmaine, ou creation du monde*, de G. de Saluste seigneur du Bartas. Reveue, augmentee, & embellie en divers passages par l'Auteur mesme. En ceste Quinzieme edition ont este adjoustez l'argument general, & amples sommaires ou commencement de chasque livre, annotation en marge, & indices propres pour l'intelligence des motz & matieres de tout l'œuvre, par S. G. S. Imprimé pour Jaques Chouet. 1581. [15^{ème} édition]



poésie du XVII^e siècle ont subi l'influence des pratiques du siècle précédent, le fonctionnement et les visées de leurs publications présentent des différences notables.

Pour assurer la cohérence interne de l'outil, une publication indépendante paraît ainsi plus judicieuse. Mais celle-ci rencontre d'autres difficultés. En tant que solution autonome, une telle publication exige des développements informatiques personnalisés qui demandent un investissement financier certain, difficile à soutenir pour un jeune chercheur sans attaches institutionnelles fixes²⁴. Ce manque de liens avec l'institution constitue aussi un handicap à propos de la question de l'hébergement et de celle de la visibilité. En effet, afin d'assurer la fiabilité et la notoriété de l'outil, il serait souhaitable que la base de données soit hébergée sur le serveur d'un organisme de recherche reconnu, chose difficile pour un chercheur à l'issue du doctorat. Plusieurs possibilités sont actuellement à l'étude, dont l'hébergement sur une future plateforme pilotée par Christophe Schuwey à l'université de Yale ou l'affiliation au projet *CORNUM – Corpus et éditions numériques en Normandie*, porté par les universités de Caen et de Rouen et financé par la région Normandie.

Une troisième voie serait l'enrichissement de la base, en y ajoutant d'une part les textes poétiques ocrés et, de l'autre, les recueils publiés au-delà de la borne chronologique de 1671. À cela s'ajouterait l'incorporation des textes en vers et en prosimètre publiés dans le *Mercure galant*²⁵. Un tel travail d'envergure, qui pourrait se faire en collaboration avec le projet *Mercure galant* (OBVIL/IREMus) piloté par Anne Piéjus, ne saurait cependant se faire sans l'infrastructure et les moyens d'un poste de titulaire (MCF/CNRS) ou, le cas échéant, dans le cadre d'un post-doc.

On le voit, un avantage de la base de données numérique est l'extraordinaire plasticité de l'outil. Nos réalisations individuelles de doctorantes ne sont pas figées mais s'avèrent un tremplin, une pâte que l'on peut remodeler, que ce soit pour enrichir le contenu, pour diversifier les façons de l'interroger ou pour améliorer l'affichage des résultats, et ce dans le but d'en faire des outils de travail intéressants pour la communauté des chercheurs. Si nous n'avons pas reçu de financements spécifiques, ni d'aide technique de la part de nos laboratoires dans la phase de développement de nos bases artisanales de thèse, cette conclusion nous amène toutefois à approuver la position développée par Christophe Schuwey dans son ouvrage *Interfaces. L'apport des humanités numériques à la littérature* : il ne s'agit pas seulement de produire des données et une interface pour les manipuler, il faut aussi entretenir la pérennité des données (ce qui est au cœur des préoccupations des humanités numériques) et des interfaces (ce qui est un enjeu davantage oublié)²⁶, de sorte que les institutions de financement devraient prévoir dans l'établissement des budgets une ligne consacrée à l'entretien et à l'évolution sur le long terme des réalisations en humanités numériques²⁷.

Dans nos cas, l'absence de soutien financier et technique pour la dimension numérique de nos travaux solitaires – et donc artisanaux – s'est révélée une opportunité pour développer des compétences informatiques, dont la maîtrise d'un ou de plusieurs langages de programmation. Toutefois, nous nous réjouissons que les choses aient évolué depuis nos premières années de doctorat au début des années 2010. La place qui revient aux humanités numériques dans les études de lettres croît. Il y a encore quelques années, un jeune chercheur n'était ni outillé, ni facilement accompagné pour développer des projets dans ce domaine. Un séminaire comme celui organisé par l'association « Les Têtes chercheuses » (doctorants de

²⁴ Au moment où la base « Refaire Lachèvre » a été développée, le logiciel de gestion de bases de données *Heurist* n'était pas encore à la disposition des doctorants français.

²⁵ Comme le *Mercure galant* prend, de maints points de vue, la suite du recueil collectif, un tel élargissement de la base demeure cohérent avec le projet initial.

²⁶ Christophe Schuwey, *Interfaces, op. cit.*, chap. 10 « Les nouvelles missions des bibliothèques », p. 107-110.

²⁷ *Id.*, p. 111.



l'ED₃LA de Lyon) en 2017-2018²⁸ est un des révélateurs de cette sensibilité des jeunes chercheurs à de nouveaux enjeux de la recherche, ainsi que de leur volonté de réfléchir ensemble et de mobiliser des interlocuteurs compétents pour pallier peut-être un certain manque dans la formation doctorale. Désormais, un cadre institutionnel émerge. D'une part, les humanités numériques intègrent certaines formations, comme c'est le cas à l'Université de Caen où des cours dans ce domaine font partie de la formation initiale disciplinaire en lettres depuis 2017/2018. On voit également naître des formations spécifiques comme la bi-licence Lettres/Informatique ouverte à Sorbonne Université en 2016²⁹, la Licence Humanités numériques ouverte à Caen à la rentrée 2018³⁰, le Master Humanités numériques de Rouen³¹ ou encore le Master Humanités numériques de l'Université Lyon 2³² – ce dernier n'étant toutefois accessible qu'en double diplôme tant l'hybridité des savoirs est importante en la matière (Humanités numériques/Histoire, Humanités numériques/Info-Com, Humanités numériques/Lettres, etc.). D'autre part, on voit apparaître des pôles « Humanités numériques » dans les laboratoires avec des ingénieurs d'étude compétents pour accompagner les chercheurs, comme c'est le cas par exemple dans le laboratoire IHRIM de Lyon. D'ailleurs, au-delà de ces initiatives locales, le MESRI offre un soutien national à ces enjeux de la recherche. C'est ainsi que le deuxième plan pour la science ouverte (2021-2024) prévoit, entre autres, de créer *Recherche Data Gouv* pour accompagner les chercheurs par le biais d'« ateliers de la donnée » et de « centres de références thématiques qui définissent les pratiques de gestion, de description et d'ouverture des données propre à une discipline ou un domaine de recherche »³³.

De la sorte, les réalisations fleurissent et vont certainement se multiplier dans les années à venir. Reste à leur assurer une visibilité – c'est-à-dire les faire connaître pour qu'elles soient utilisées le plus largement possible – et à leur assurer une pérennité. La solution d'une technologie générique et d'une plateforme d'accueil unique fait courir le risque d'une uniformisation réductrice pour la pensée innovante. Un effort d'intervalorisation, en revanche, semble souhaitable, par le biais de la citation des « bases amies » sur nos interfaces en ligne par exemple. La mise en relation de ces bases, différentes, mais souvent complémentaires, permettrait alors de leur assurer la visibilité qui leur manque individuellement. Nous ne pouvons donc, pour finir, que saluer l'esprit fédérateur du colloque organisé par l'équipe des *Joyeuses Inventions* qui a donné lieu à ces actes, après avoir réuni de nombreux porteurs de projets en humanités numériques dont les trouvailles et les questionnements ont pu converger.

²⁸ « Métamorphoses du document et pratiques de recherche à l'ère des Humanités numériques » : <https://teteschercheuses.hypotheses.org/3879> (consulté le 6 janvier 2020).

²⁹ <https://obvil.sorbonne-universite.fr/formation/bi-licence> (consulté le 25 avril 2022).

³⁰ <http://ufrhss.unicaen.fr/disciplines/humanites-classiques-et-numeriques/> (consulté le 25 avril 2022).

³¹ <https://masterhrouen.wordpress.com/> (consulté le 6 janvier 2020).

³² <https://icom.univ-lyon2.fr/formation/humanites-numeriques> (consulté le 25 avril 2021).

³³ <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/> (consulté le 5 novembre 2021).



BIBLIOGRAPHIE

- DUCOURNEAU Claire et GLINOER Anthony, « Arpenter la vie littéraire. Les bases de données dans les études sociohistoriques sur l'imprimé et la littérature », dans *Arpenter la vie littéraire*, numéro thématique de la revue *Biens Symboliques / Symbolic goods*, n°2, février 2018, en ligne : <https://www.biens-symboliques.net/211>
- HAYLES N. Katherine, *Lire et penser en milieu numérique*, trad. C. Degoutin, Grenoble, Ellug, 2016.
- MORETTI Franco, *Graphes, cartes et arbres* (2006), trad. E. Dobenesque, Paris, Les Prairies ordinaires, 2008.
- SCHUWEY Christophe, *Interfaces. L'apport des humanités numériques à la littérature*, Neufchâtel, Livreo-Alphil, 2019.