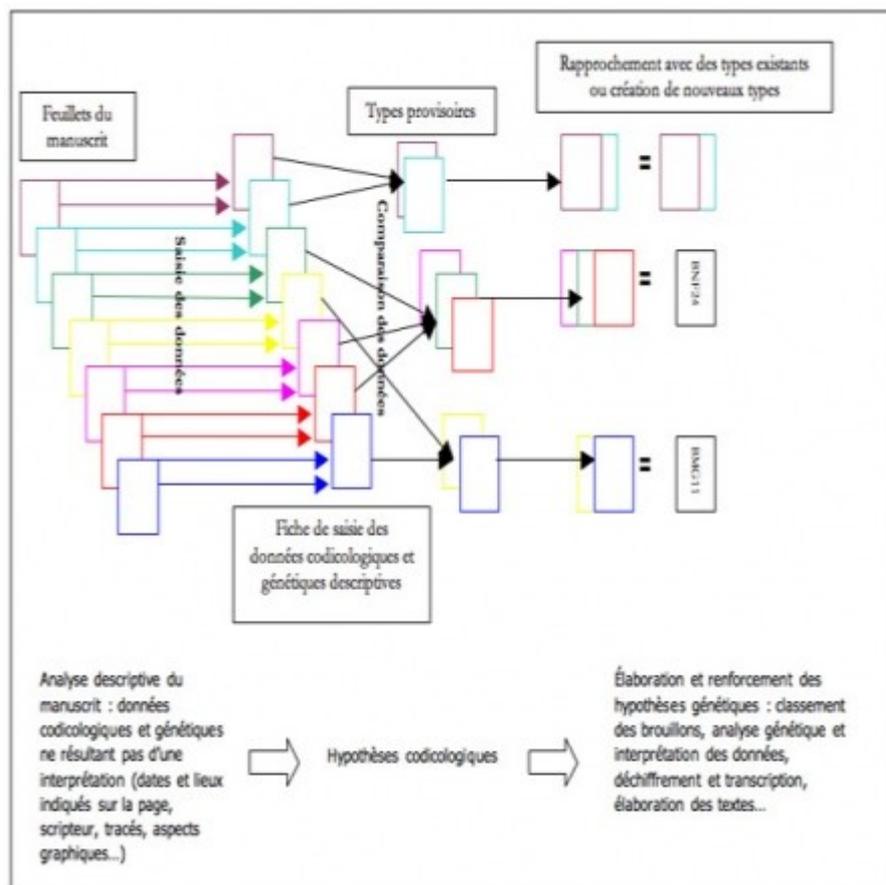


# Un nouvel instrument de travail pour l'analyse des manuscrits : la base de données *MUSE*

Claire Bustarret et Serge Linkès (16 mars 2008)

Fig.1 Processus d'analyse des papiers



Contrairement aux répertoires informatisés qui transmettent des données d'emblée analysées selon les hypothèses des chercheurs, *MUSE* doit permettre de proposer des hypothèses à partir de la multitude des données collectées par le codicologue. En effet, c'est à partir du croisement des données concernant chaque feuillet que pourront être identifiés les différents types de papier, les différentes interventions scripturales. La définition d'un type de papier, et de ses interactions avec les opérations d'écriture, intervient donc au terme de la description codicologique exhaustive d'un ou plusieurs volume(s) manuscrit(s). Le manuscrit dans son ensemble, l'objet intellectuel et l'objet matériel, se trouvent donc au début et à la fin de cette chaîne de données. Au début, car il est l'objet concret soumis à l'observation, à l'analyse matérielle et graphique, et à la fin, car il est l'objet virtuel défini par l'hypothèse résultant du croisement des données.

Du point de vue de la méthode codicologique, l'utilisation de la base entraîne d'importantes conséquences. D'une part, le caractère systématique de la collecte informatisée impose une normalisation des critères descriptifs : nous avons ainsi réduit les options pour l'identification

des couleurs du papier, distingué les données mesurables des évaluations qualitatives et séparé les zones de remarques. La capacité de calcul de l'ordinateur donne en outre accès à l'exploitation de données chiffrées, encourageant le recours aux instruments de mesure tels le micromètre et le perthomètre (les mesures d'épaisseur et de rugosité donnant lieu à des calculs automatiques de moyennes, par exemple). D'autre part, la base de données relationnelle autorise la visualisation comparative d'une quantité de données dépassant largement nos capacités de mémorisation, le stockage permettant une accumulation qui rend la base de plus en plus performante au fur et à mesure du traitement des corpus successifs (consultation des types de papiers déjà inventoriés<sup>13</sup>, des informations historiques sur les fabricants, sur les copistes, recoupements avec des occurrences relevées dans d'autres corpus, etc.).

Enfin, le réseau de relations entre les fichiers constitue un contexte multiple pour chaque groupe de données, et permet de mettre au jour des interactions insoupçonnées entre divers aspects de l'activité de l'écrivain telle qu'elle nous est restituée par le témoin manuscrit. Nous sommes ainsi passés d'une base de références – à l'instar des répertoires de filigranes – à un instrument de recherche permettant non seulement de collecter les données in situ, mais aussi de tester des hypothèses de classement. Examinés conjointement, les indices matériels et les divers états du matériau scriptural révèlent des façons de faire, des gestes, des usages inspirés par le projet créatif et qui, en retour, informent celui-ci<sup>14</sup>. Cette notion d'usages des matériaux et des moyens graphiques au service de la création nous a paru déterminante, c'est pourquoi la base que nous présentons aujourd'hui comme un nouvel outil de travail fonctionnel pour les chercheurs (codicologues, généticiens, ou autres) étudiant des corpus du XVIII<sup>e</sup> au XX<sup>e</sup> siècle s'intitule « *MUSE* » : *Manuscrits, Usages des Supports et de l'Écriture*.

### Choix du logiciel et du mode relationnel

Lorsque le chercheur littéraire souhaite aborder le monde informatique, il doit accepter d'adopter les stratégies en vigueur dans cet univers. Ainsi, tout bon informaticien sait qu'il est impossible d'élaborer une base de données sans passer par l'étape primordiale de la conception du système d'information. Ce système permet non seulement d'organiser les données (organisation des fichiers et des relations entre les fichiers), mais aussi d'envisager les aspects humains de la base de données : utilisateurs en amont et en aval, organisation de la saisie des données, maintenance, sauvegarde, conditions d'utilisation...

Ce système doit donc être élaboré en étroite collaboration par toutes les parties intéressées afin de tenir compte des besoins, des attentes et des contraintes de chacun. Dans le cas de *MUSE*, ce travail a été réalisé en présence d'utilisateurs potentiels étudiant des corpus de nature et d'époques différentes. L'organisation des données dans chaque fichier a également donné lieu à des débats afin de définir les termes descriptifs et critiques à utiliser, les automatismes permettant d'éviter les saisies rébarbatives... La même attention a été portée à la définition des relations entre les fichiers.

L'autre aspect de la mise en place du système d'information concerne le choix du logiciel permettant d'élaborer une base de données. Ce choix représente un enjeu considérable, c'est même l'un des éléments essentiels du cahier des charges. On en tient rarement compte au moment de se lancer dans une telle aventure... Pourtant, ce choix conditionnera la compatibilité, la pérennité, et l'évolution de la base de données. N'oublions pas que le but d'une base est avant tout de communiquer des données, même si son interface peut paraître moins conviviale que certaines interfaces Web très communes aujourd'hui. Notre choix a donc d'abord été conditionné par un souci de compatibilité entre plateformes PC et

Macintosh. Cette base devait non seulement être compatible Mac et PC tout en étant utilisable sur n'importe quel système d'exploitation, elle devait également faciliter un échange dynamique des données.

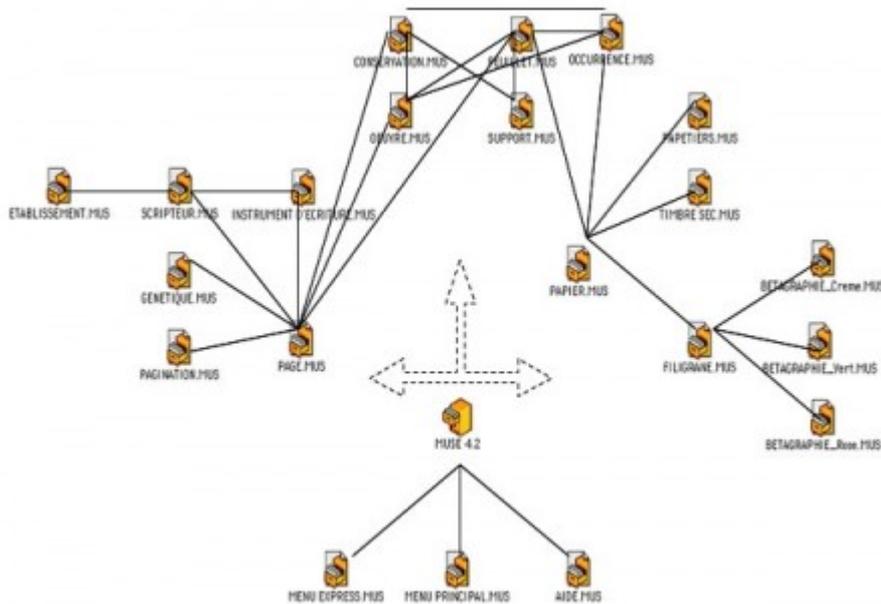
Cette condition explique que nous ayons choisi d'utiliser un logiciel du commerce et non de développer une application à partir d'un langage de programmation. Les logiciels de bases de données compatibles Mac/PC étaient assez rares à l'époque de la gestation de *MUSE* ; nous avons opté pour le logiciel *File Maker Pro* qui semblait le plus adapté à notre projet puisque la société qui le proposait avait comme principal souci cette compatibilité – descendante et ascendante. De plus, le dynamisme de cette société laissait supposer que ce logiciel ne risquait pas de disparaître du jour au lendemain, comme c'est souvent le cas dans le monde informatique. Il faut également signaler que ce logiciel était accompagné d'un module permettant de créer des applications autonomes libres de droit à partir de la base de données réalisée, ce qui nous a permis de multiplier les versions de *MUSE* sur différents postes de travail.

Afin de mieux comprendre les besoins auxquels doit répondre la base, suivons pas à pas la démarche descriptive préalable à tout travail génétique<sup>15</sup>. En abordant un corpus manuscrit, qui a été constitué d'après le catalogue de l'institution de conservation ou à partir des travaux déjà existants, le ou la codicologue commence par inventorier les supports qui composent les différents volumes ou liasses : les ensembles, reliés ou non, constituent-ils des séquences homogènes ou des recueils hétérogènes du point de vue matériel (cahiers, carnets ou feuilles volantes, formats in-folio ou in-quarto, qualité des papiers, modalités d'assemblage, etc.) ? De même on distinguera les documents autographes des interventions allographes, mains de copistes, secrétaires ou familiers de l'écrivain, et l'on relèvera d'éventuelles marques d'origine privée ou institutionnelle concernant le classement du corpus (foliotation, pagination, annotations marginales, feuillets intercalaires). Une seconde étape (voir fig. 1) consiste à repérer les différents types de papier utilisés : chaque feuillet est donc mesuré et décrit dans le détail, y compris, le cas échéant, son filigrane, afin de regrouper progressivement sur une même fiche toutes les occurrences d'un même type de papier, les occurrences douteuses pouvant rester en attente d'une identification ultérieure (types « indéterminés »). Quant aux aspects graphiques, chaque page écrite fait l'objet d'une analyse scripturale : couleur de l'encre, nature du ou des instrument(s) d'écriture, caractéristiques courantes ou exceptionnelles de la disposition spatiale, numérotation unique ou multiple. Dans un troisième temps, le chercheur recourt aux diverses sources qui lui permettront de mieux situer les phases de rédaction dans l'espace et le temps : sources biographiques et études littéraires d'une part, ouvrages sur l'histoire du papier et répertoires de filigranes d'autre part. C'est en procédant à une comparaison systématique des données codicologiques, scripturales et historiques que l'on peut obtenir une synthèse descriptive suffisamment complète (identification, répartition et modalités d'utilisation, classement) pour servir de base à l'analyse génétique proprement dite des manuscrits examinés<sup>16</sup>.

La base de travail doit donc permettre la collecte systématique, dans les divers fichiers réservés à chaque aspect du manuscrit (voir fig. 2), des données concernant les papiers, les filigranes et timbres secs qui les marquent, les scripteurs et les instruments d'écriture qu'ils emploient, les auteurs et les œuvres traitées, les volumes ou autres unités matérielles (« supports ») examinés et les feuillets qui les composent. Enfin chaque feuillet comprend nécessairement deux pages, écrites ou vierges, paginées ou non, se prêtant à l'analyse génétique des opérations d'écriture. L'enjeu d'une telle répartition des informations consiste à rationaliser la saisie mais aussi et surtout à permettre les recoupements entre divers fichiers en réponse à des requêtes multicritères.

## Structure de la base, relations entre les fichiers

Fig. 2 : Interactions et relations entre les fichiers de *MUSE*

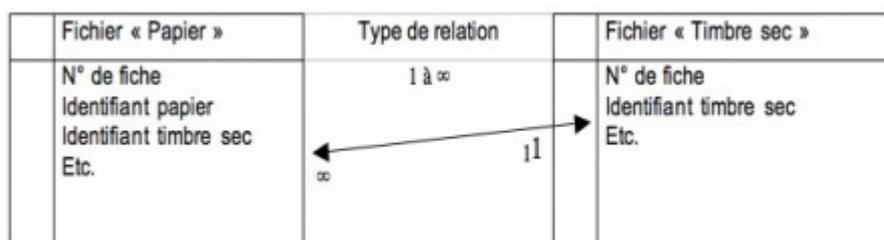


D'un point de vue logiciel, *MUSE* devait pouvoir gérer les intégrités référentielles entre les différents fichiers qui la composent, afin d'éviter tous risques de doublons ou d'erreurs d'attribution (une base de données qui se contenterait d'effectuer le tri de fiches ne pourrait en aucun cas être garante de la validité des recherches).

Ainsi la base de données que nous avons mise au point est répartie en plusieurs fichiers, chacun comportant un certain nombre de données spécifiques. Les relations établies entre les fichiers agissent principalement lorsque l'on ajoute ou supprime des enregistrements<sup>17</sup>. Elle est la garante de la validité des recoupements futurs... Le respect de ces règles nous garantit donc, par exemple, que chaque folio décrit ne sera associé qu'à un seul type de papier. Ainsi, lors d'une interrogation sur le type de papier vélin « *Chambellan* », *MUSE* fournira uniquement la liste des folios rédigés sur ce type de papier, en excluant les autres, puisque seules certaines des entrées du fichier « Occurrences » sont liées à la fiche unique donnant la description de ce type dans le fichier « Papier ». La multiplication des fichiers dans une base de données relationnelles permet, contrairement à ce que l'on pourrait croire, d'accroître la souplesse d'utilisation et surtout la rapidité du traitement des informations.

En effet, tous les fichiers étant liés entre eux et respectant l'intégrité référentielle (un seul enregistrement est lié à un autre enregistrement ou à plusieurs, selon la nécessité), nous pouvons créer pour les besoins de la saisie une fiche principale qui fera apparaître côte à côte différentes « tables » (listes totales ou partielles des rubriques figurant dans les fichiers liés) reliées, ce qui permettra d'enregistrer sur une seule fiche toutes les informations concernant une page.

Fig. 3 : Relation entre deux tables de la base *MUSE*



La figure 3 présente le type de relations qui est mis en œuvre dans la base de données *MUSE*. Les fiches « Papier » et « Timbre sec » sont reliées par le champ « Identifiant timbre sec » qu'elles ont en commun. Les informations concernant le timbre sec seront saisies une fois pour toutes dans le fichier « Timbre sec » et simplement rappelées lors de la saisie, dans la table « Papier », des informations concernant les caractéristiques du papier, parmi lesquelles figure ce timbre. La relation mise en œuvre dans ce cas est du type « 1 à ∞ » : en effet, théoriquement, chaque papier ne peut revêtir qu'un seul timbre de papetier, alors qu'un timbre, qui est une marque de fabricant, peut être apposé sur différents types de papier. En effet, à la différence du filigrane, le timbre sec est apposé après la fabrication du papier industriel. Il n'est donc pas un critère suffisant pour assurer l'identification du type de papier, mais il permet d'en repérer la provenance, indice qui peut s'avérer déterminant comme nous avons pu le constater lors de l'analyse de manuscrits stendhaliens. On retrouve dans le manuscrit de *Lamiel*, rédigé dans les derniers moments de la vie de Stendhal, plusieurs sortes de papiers vélins employés à des périodes différentes (de 1839 à 1842), qui portent un timbre identique permettant d'en identifier la provenance commune : « Chambellan, Papetier, 1 rue du Bouloi à Paris ». Dès lors, le recoupement des informations concernant le timbre sec, le papier, le scripteur (recours à un ou plusieurs copistes) et le texte nous a permis d'affirmer que Stendhal avait dicté soixante-dix pages de son roman à Paris en mai 1839 au copiste Bonavie, le copiste de *La Chartreuse de Parme*, alors que l'on pensait jusqu'ici que ce roman avait été dicté à Rome à un secrétaire inconnu en octobre 1839<sup>18</sup>.

Ainsi, les bases de données relationnelles permettent non seulement d'économiser la saisie redondante d'une même donnée, mais surtout d'optimiser le recoupement des données lors de requêtes multiples. Un système d'information correctement « pensé », répartissant dans un fichier précis les données concernant un seul et même type d'entité (l'instrument d'écriture, le type de papier, le scripteur...), permet d'envisager les requêtes les plus insolites. On a pu ainsi mener une enquête sur les habitudes d'écriture de Stendhal dans le manuscrit de *Lamiel* et s'apercevoir du rôle systématiquement réservé à tel instrument d'écriture : la mine de plomb par exemple, est uniquement utilisée pour corriger et annoter une version que Stendhal amplifie lors d'une nouvelle dictée.

Du fait de la complexité des relations possibles entre les supports et les campagnes d'écriture (un même feuillet peut porter plusieurs états génétiques, les brouillons d'une œuvre peuvent être répartis entre plusieurs volumes, une même version peut comporter l'intervention de plusieurs scripteurs sur les mêmes supports ou sur des supports différents, etc.), la création d'une base de données adaptée à la problématique des manuscrits d'auteurs nécessite les mêmes précautions, la même rigueur qu'une base utilisée au sein d'une entreprise dans des domaines considérés comme plus techniques. Cependant, il est évident que dans le cas de la littérature et plus généralement des sciences humaines, il convient de faire preuve de souplesse par rapport aux règles établies par les informaticiens. On ne peut cependant pas s'y soustraire totalement, car, indépendamment des habitudes de travail et de la part importante dévolue à l'interprétation dans les disciplines littéraires, la « machine » informatique ne peut fonctionner (pour le moment) que dans un cadre formel systématique, donc contraignant.

