



**HAL**  
open science

## Nouvelles configurations : textes, outils, méthodes, et infrastructures de recherche dans les études de lettres

Christof Schöch

► **To cite this version:**

Christof Schöch. Nouvelles configurations : textes, outils, méthodes, et infrastructures de recherche dans les études de lettres. Colloque international: Configuration(s), Jun 2012, Paris, France. hal-00951518

**HAL Id: hal-00951518**

**<https://hal.science/hal-00951518v1>**

Submitted on 24 Feb 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Christof Schöch

## **Nouvelles configurations : textes, outils, méthodes, et infrastructures de recherche dans les études de lettres**

### **Table des matières**

Introduction

1. Configurations textuelles récurrentes – la description romanesque
2. (Re-)configurations méthodiques – une base de données
3. Reconfigurations de la recherche – les infrastructures numériques

Conclusion

### **Introduction**

Le tournant numérique n'est plus aujourd'hui uniquement un phénomène de technologie et de société, mais concerne pleinement l'enseignement et la recherche en sciences humaines et sociales<sup>1</sup>. En effet, la transdiscipline des *humanités numériques* est la manifestation autant que le moteur de modifications en profondeur de nos pratiques de chercheurs. On peut aborder ces modifications par le biais des nouvelles configuration textuelles que les méthodes numériques permettent de découvrir dans les textes littéraires, par les re-configurations que l'existence de données et d'outils numériques produit au niveau des méthodes de la recherche en lettres, et par l'apparition des infrastructures numériques et de leur rôle transformateur pour les conditions de la recherche en sciences humaines. Il importe aujourd'hui de prendre la mesure de ces transformations pour mieux comprendre les chances et les risques qu'apporte le numérique aux études de lettres.

Un projet de recherche sur la pratique descriptive dans un corpus de trente-deux romans français de la seconde moitié du dix-huitième siècle<sup>2</sup>, projet qui a été marqué à plusieurs niveaux par le tournant numérique, devrait permettre de mettre au jour quelques-unes de ces nouvelles configurations. Après une rapide réflexion sur ce qu'on peut entendre par une configuration, je voudrais parler de trois types de configurations nouvelles : au niveau textuel, c'est-à-dire celui de la pratique descriptive dans les romans étudiés, une analyse qui emploie les outils numériques est en mesure de découvrir un certain nombre de configurations textuelles récurrentes insoupçonnées ; au niveau méthodique, l'usage d'une base de données textuelles, employée pour étudier le foisonnement des descriptions romanesques, a produit un certain nombre de re-configurations méthodiques ; enfin, quand on élargit

---

1 Le terme « tournant numérique » a récemment été directement associé aux sciences humaines et sociales par les auteurs du *Manifeste des humanités digitales* ; voir COLLECTIF, « Manifeste des humanités digitales », in *That Camp Paris*, 26 mars 2011, <http://tcp.hypotheses.org/318>. Le terme correspondant anglais, « digital turn », semble en usage dans le contexte des sciences humaines depuis environ l'an 2000.

2 Voir SCHÖCH, Christof, *La description double dans le roman français des Lumières, 1760-1800*, Paris, Éditions Classiques Garnier, « L'Europe des Lumières », 2011.

la vue pour voir au-delà de ce projet spécifique, on peut observer aujourd'hui un certain nombre de reconfigurations plus générales de la recherche dans les études de lettres, liées surtout aux infrastructures numériques. Au cours des réflexions que voudraient être les pages suivantes, on passera donc du projet individuel et concret vers les conditions mêmes de la recherche de manière plus générale.

Qu'est ce que donc une configuration ? Pour approcher de la notion, on peut utilement consulter quelques dictionnaires, numériques en l'occurrence. On découvre ainsi que, pour le *Dictionnaire de l'Académie française*, dans sa cinquième édition de 1798, le terme désigne de manière quelque peu paradoxale la « forme extérieure » des corps, mais surtout « en parlant des parties tenues et insensibles, qui échappent à la vue<sup>3</sup> », de manière à ce que ces configurations n'ont pas une évidence directe mais restent à découvrir ou à démontrer. Le dictionnaire *Littré*, au dix-neuvième siècle, nous indique qu'il s'agit d'un « terme d'astrologie » et insiste surtout sur l'idée que la configuration définit des rapports entre plusieurs unités<sup>4</sup>. On peut en conclure qu'une configuration est un ensemble organisé d'éléments distincts, ce qui implique qu'elle est en même temps plus que la simple somme de ses parties. En outre, il semble que le terme ait une dimension topographique qui permet de l'employer comme une métaphore spatiale.

Plus récemment, le *Online Etymology Dictionary* nous renseigne sur l'étymologie du terme<sup>5</sup>, dans laquelle on retiendra surtout l'idée que, dérivée d'un verbe, *figurare*, la configuration est une entité non pas passive et stable, mais bien une activité de construction. Cette idée est confirmée encore par un sens plus récent du terme sur lequel l'encyclopédie collective en ligne *Wikipédia* nous renseigne<sup>6</sup> : dans le contexte de l'informatique, la configuration des composantes d'un ordinateur ou des paramètres d'un logiciel est le résultat d'une activité de l'utilisateur, ce qui implique aussi bien un choix individuel et une orientation vers le futur, c'est-à-dire vers un usage avec une finalité propre. Dans tous les cas, ces définitions impliquent que la configuration n'est pas forcément une donnée empirique, mais quelque chose à découvrir de manière active et même quelque chose à construire, et, le cas échéant, à déconstruire.

## 1. Configurations textuelles récurrentes : la description romanesque

Avec ces quelques éléments de définition, je voudrais donc maintenant aborder, dans un premier temps, le sujet des configurations textuelles ; il s'agit d'ensembles de phénomènes textuels qui fonctionnent en tant que système et dont les éléments et leurs relations ne sont souvent pas directement visibles ; plutôt, ces structures récurrentes ne deviennent visibles que quand on prend en

---

3 « Forme extérieure, ou surface qui borne les corps, et leur donne une figure particulière. [...] Il s'emploie surtout en physique et en chimie, en parlant des parties tenues et insensibles, qui échappent à la vue. » Article « Configuration », in *Dictionnaire de l'Académie française*, 5e édition (Paris, 1798), édition électronique, Chicago, The ARTFL Project, 2004-2012, <http://artfl-project.uchicago.edu/content/dictionnaires-dautresfois>.

4 « Situation relative des corps planétaires les uns par rapport aux autres, et, en particulier, des satellites de Jupiter. » Article « Configuration », *Dictionnaire de français Littré* (Paris, 1863-1877), édition électronique, Paris, Reverso, 2010, <http://littre.reverso.net/>.

5 « From lat. *configuratio*em, derived from *configurare*, « to fashion after a pattern », from *con-*, « together » and *figurare*, « to shape / to represent », from *figura*, « visible form or appearance of a person ». *Online Etymology Dictionary*, éd. Douglas Harper, 2001-2012, <http://www.etymonline.com/>.

6 « Ensemble de caractéristiques techniques qui ne dépendent pas du constructeur mais découlent des choix de l'acheteur et de l'utilisateur [...]. C'est aussi l'activité qui consiste à modifier des paramètres de configuration ». COLLECTIF, Article « Configuration (informatique) », *Wikipédia France*, version du 23 octobre 2012, [http://fr.wikipedia.org/w/index.php?title=Configuration\\_%28informatique%29&oldid=84473553](http://fr.wikipedia.org/w/index.php?title=Configuration_%28informatique%29&oldid=84473553).

compte un nombre plus large d'exemples que celui qu'on trouverait dans tel ou tel roman individuel.

Cette première partie est la plus proche du texte littéraire, et la plus proche également du projet de recherche qui me sert d'appui ici. Ce projet de recherche portait sur la pratique descriptive dans le roman français de la seconde moitié du dix-huitième siècle. C'est une période-clé dans l'histoire de la description, pour plusieurs raisons : on y trouve une grande expérimentation avec des stratégies descriptives dans de nombreux genres d'écriture, dont le roman ; c'est la période qui prépare la transition de la description rhétorique et contrôlée du dix-septième siècle vers la description foisonnante, multi-fonctionnelle, du dix-neuvième siècle ; enfin, c'est une période intéressante et paradoxale pour la description, puisqu'on y observe un contraste saisissant entre l'apparition massive de descriptions partout dans les belles-lettres et les sciences, tandis que dans le roman, la description reste malgré tout un phénomène émergent, au statut encore problématique, et par conséquent en quête de légitimation.

Pour aller vite, on peut noter que le dix-huitième siècle est le siècle de la « description double », c'est-à-dire celui d'une coexistence de deux notions de la description, avec leurs définitions et leurs enjeux propres : la première notion est rhétorique, tandis que la seconde est structurelle<sup>7</sup>. L'enjeu central de la notion rhétorique est le rapport entre description et peinture, c'est-à-dire qu'il s'agit avant tout de l'*ekphrasis* au sens large, donc des qualités picturales de la description. La notion structurelle de la description, par contraste, a pour enjeu central le rapport entre description et narration, autrement dit le statut et les fonctions de la description dans le roman, avec le statut problématiques de la description et des stratégies textuelles de sa légitimation. Dans la suite, je me concentre uniquement sur un aspect textuel précis de la notion structurelle de la description.

L'aspect de la description romanesque sur lequel je voudrais me concentrer est ce que j'appelle la *motivation narrative*, définie de la manière suivante :

Par opposition aux remarques métadiscursives, façon explicite de gérer ou de « négocier » la place de la description dans le roman, la motivation narrative le fait de manière implicite mais tout aussi efficace. Sa fonction est analogue, mais son fonctionnement est très différent : la motivation narrative rend légitimes les passages descriptifs présents dans le roman, mais elle le fait surtout en les intégrant dans le tissu narratif, en ménageant dans l'histoire du roman des occasions prétendues « naturelles » pour leur apparition. Autrement dit, la motivation narrative concerne la mise en place, au niveau de l'histoire, d'événements, de situations ou d'actions qui donnent lieu à des passages descriptifs. Plus précisément, un événement dans l'histoire, l'action d'un personnage ou certaines circonstances de l'histoire déclenchent (le plus souvent), prolongent et conditionnent (parfois) et terminent (plus rarement) le passage descriptif. La fonction principale de la motivation narrative est de fondre le passage descriptif dans la narration de manière à ce que le lecteur ne s'y heurte pas, mais l'accepte comme une suite logique ou un prolongement naturel de la narration<sup>8</sup>.

Les éléments qui entrent dans la motivation narrative sont donc des entités au niveau de l'histoire, c'est-à-dire des personnages, des actions, des événements ou des circonstances fictionnels. Philippe Hamon, qui a le premier formulé une théorie complète des phénomènes sous considération ici, parle d'un « système configuratif » de la description<sup>9</sup>. Chez Hamon, ce système est plus large que la seule motivation narrative décrite ici, et comporte aussi bien des éléments typographiques et structurels (alinéas et divisions en chapitres), des remarques métadiscursives (commentaires explicites) et des

---

7 Pour plus de détails sur la question, voir SCHÖCH, Christof, *La description double*, op. cit., p. 25-45.

8 *Ibid.*, p. 157.

9 HAMON, Philippe, *Du descriptif* (1981), Paris, Hachette, 1993, p. 165-204.

éléments de l'histoire (de nature implicite). Selon Hamon, la fonction primordiale du système configuratif est de marquer la limite entre description et narration, et de déclencher de cette manière ce qu'il appelle une compétence descriptive chez le lecteur. Hamon souligne par ailleurs qu'on trouve surtout des configurations symétriques dans les romans qui forment la base de son modèle, c'est-à-dire les romans du dix-neuvième siècle et tout particulièrement ceux de Zola. Un exemple, cité ici de manière abrégé, permettra de montrer cette structure symétrique :

D'ailleurs, la porte s'était ouverte sans attendre, et la voisine, Mme Jabouille, Mathilde comme on la nommait familièrement, parut sur le seuil. Elle avait trente ans, la figure plate, ravagée de maigreur, avec des yeux de passion, aux paupières violâtres et meurtries. [...] / Un instant, elle resta immobile, fouillant les coins d'un rapide coup d'œil. Une senteur forte s'était répandue, la senteur des simples dont sa robe se trouvait imprégnée, et qu'elle apportait dans sa chevelure grasse, défrisée toujours [...]. / Mais Mathilde riait avec les autres, d'un rire aigu d'impudeur. Carrément, elle était entrée, elle avait refermé la porte<sup>10</sup>.

On peut décrire le fonctionnement de ce passage descriptif comme la configuration symétrique de deux suites d'éléments de la motivation narrative. (Voir la figure 1, ci-dessous, pour une visualisation de cette configuration, dans laquelle la position des éléments dans le texte forme l'axe vertical, et les trois types des éléments de la motivation narrative forment l'axe horizontal.)

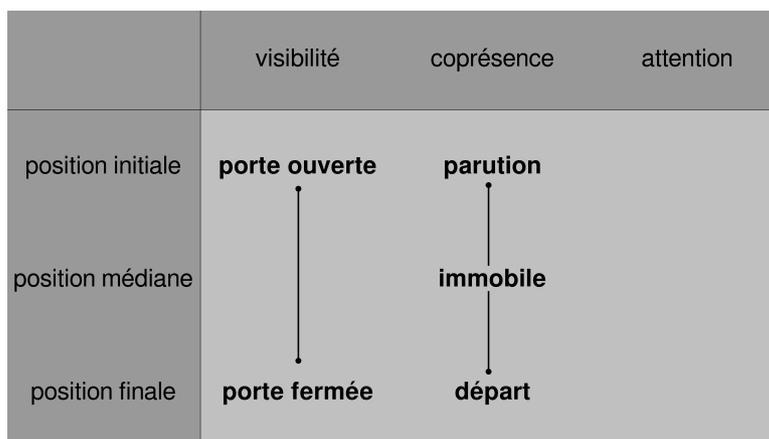


Fig. 1 : Configuration symétrique de la motivation narrative, chez Zola.

Une première série d'éléments met en rapport l'ouverture de la porte au début avec la fermeture de la porte à la fin du passage ; une seconde série met en rapport l'apparition de Mathilde, son immobilisation sur le seuil de la porte, enfin le fait qu'elle quitte le seuil de la porte pour entrer dans la pièce. La double série symétrique construit, toujours selon Hamon, une sorte de cadre autour de la description, ce qui implique à la fois une délimitation de la description par rapport à la narration, et l'activation de la compétence descriptive spécifique chez le lecteur qui est parfaitement conscient de cette transition de la narration à la description.

Quand on approche les romans du dix-huitième siècle avec ce modèle symétrique, force est de constater que, si l'on trouve bien quelques jolies exemples symétriques, ce qui domine largement sont des configurations asymétriques et qui semblent, à première vue, former tout simplement des occurrences peu systématiques ou même chaotiques. Or, cette projection d'un modèle sur un corpus nouveau risque de créer une méprise du fonctionnement de la description romanesque dans le contexte esthétique spécifique du dix-huitième siècle, qui donne une fonction et une forme spécifique

10 ZOLA, Émile, *L'Œuvre* (1885/86), éd. d'Antoinette Ehrard, Paris, Flammarion, 1974, p. 127-128.

à la motivation narrative dans le roman de cette époque.

Ce qui n'était dans un premier temps qu'une intuition de lecture a pu être éprouvé, appuyé sur un ensemble massif d'exemples, grâce aux méthodes et outils numériques. Une première précaution méthodique a été de décomposer le système de Hamon ; d'une part, en traitant séparément deux aspects du "système configuratif" de Hamon, les remarques métadiscursives d'une part, la motivation narrative de l'autre ; et d'autre part, de faire éclater les grandes configurations symétriques proposées par Hamon pour arriver aux composantes minimales qui les forment. On peut considérer, en effet, que les entités qui entrent dans la motivation narrative sont des personnages, des actions, des événements et des circonstances des événements. En outre, ces entités peuvent concerner la coprésence entre l'objet décrit et le personnage porteur de la perspective narrative, ou bien la visibilité de l'objet décrit, ou encore l'attention que porte le personnage à l'objet décrit. Enfin, ces entités peuvent apparaître en position initiale, médiane ou finale par rapport à la description en question. (La définition et la délimitation de ces entités et de ces catégories a elle-même été le résultat d'un long processus, inductif et itératif, de travail avec les textes, et ce au sein d'un cadre de travail numérique ; j'y reviendrai.)

La deuxième étape méthodiquement centrale de l'analyse a été de documenter dans une base de données, pour l'ensemble des environ 1500 descriptions relevés dans les trente-deux romans étudiés, les entités de motivation narrative qui y sont présentes, avec leur nature et leur position. Grâce à cette base de données, il a été possible de constater quels types de configurations des entités de la motivation narratives sont les plus récurrentes dans les romans étudiés. Ces configurations récurrentes, qui existent dans un rapport fonctionnel avec la légitimation de la description dans le roman, forment la motivation narrative. Il se trouvent qu'il y a bien des configurations récurrentes, mais que celles-ci, au lieu d'être symétriques, tendent à être asymétriques. L'exemple suivant représente bien le genre de motivation narrative asymétrique que l'on peut trouver dans le roman du dix-huitième siècle :

Je me levai sans oser souffler, et fus me placer à un secrétaire noir, marqué de taches rouges en bois incrusté, et qui jouaient le sang à faire horreur. J'ouvre ce secrétaire, quel spectacle, grand Dieu ! l'encre dans un crâne d'ivoire, les chandeliers... des faisceaux d'ossements, portant une petite bougie lugubre, et les tablettes couvertes d'une série de poisons en fioles et étiquetés ! Je me trouvai mal [...] <sup>11</sup>.

On constate dans un premier temps qu'il y a trois éléments de motivation narrative : la narratrice note qu'elle se place proche du secrétaire (en position initiale, on trouve donc un mouvement d'approche entre sujet et objet, donc un élément de coprésence) ; ensuite, en position médiane, la narratrice note qu'elle ouvre le secrétaire et voit ce qu'il y a dedans, ce qui représente un élément concernant la visibilité ; enfin, la narratrice note qu'elle se trouve mal suite à son exposition aux objets qu'elle vient de nommer, ce qui termine la description, et représente une instance d'un élément concernant l'attention du sujet à l'objet (attention). Autrement dit, on trouve un type d'élément différent à chaque position, configuration illustrée par la figure 2.

---

11 RÉVÉRONI SAINT-CYR, Jacques Antoine de, *Pauliska ou la perversité moderne* (1795), éd. Michel Delon, Paris, Desjonquères, 1991, p. 92.

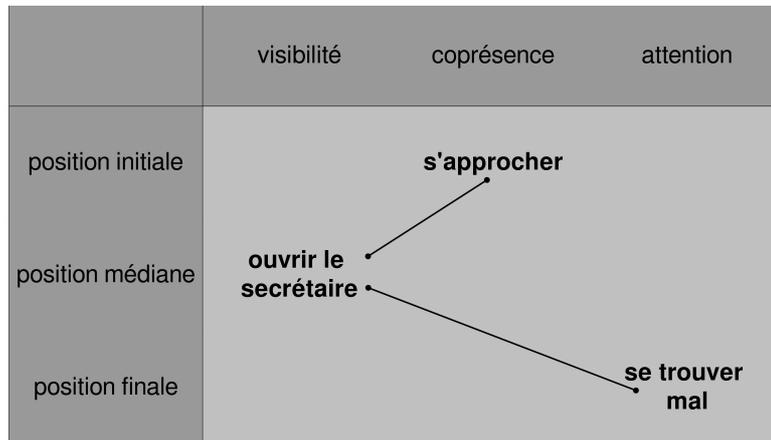


Fig. 2 : Configuration asymétrique de la motivation narrative, chez Révéroni Saint-Cyr.

On voit tout de suite que le schéma est très différent de celui représentant la description tiré de *L'Œuvre* de Zola. Cependant, ces exemples n'en sont que deux parmi des milliers qu'on pourrait choisir, et n'ont donc en eux-mêmes aucun droit à être pris comme représentatifs ou même typiques. C'est ici qu'intervient le regard large sur les presque 1500 exemples des romans du dix-huitième siècle, qui permet de voir quelques grandes tendances (voir fig. 3), dont une en particulier concernant la distribution des éléments de motivation narrative selon leur position.

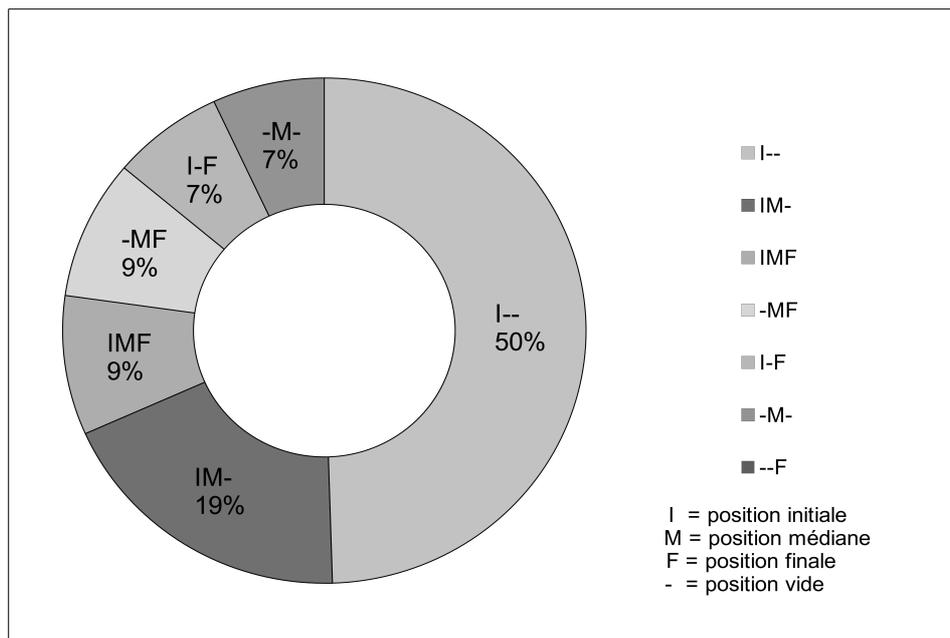


Fig. 3 : Distribution des éléments de la motivation narrative selon leur position (n = 1486).

Abstraction faite de la nature des éléments de la motivation narrative, donc avec un regard uniquement pour leur position, cette vue d'ensemble montre en premier lieu que la configuration largement la plus fréquente, avec 50%, est aussi la plus simple : c'est celle qui consiste en un élément de motivation narrative en position initiale seulement, qui déclenche en quelque sorte la description. En l'absence d'un élément correspondant en position finale, il s'agit forcément d'une configuration asymétrique. En deuxième position, avec 19%, vient la configuration dans laquelle une description comporte des éléments de motivation narrative en position initiale et médiane, mais pas finale. Par contraste, les deux configurations symétriques ne constituent qu'une plus modeste proportion des exemples, avec 9% pour le cas où toutes les trois positions sont occupées, et 7% pour

le cas où les positions initiale et finale sont occupées.

Sur la base de ce type d'informations, on peut analyser dans davantage de détail les configurations récurrentes. On peut constater, par exemple, que les configurations avec un élément de motivation narrative en position initiale et finale comportent souvent des éléments de nature différente dans ces deux positions, ce qui réduit l'effet symétrique et d'encadrement de la motivation narrative. De manière générale, il apparaît que la fonction de la motivation narrative dans les romans du dix-huitième siècle sert parfois, certes, à délimiter ou à encadrer la description, mais que dans la plupart des cas, elle sert plutôt à ménager ou à assouplir la transition entre description et narration.

## **2. (Re-)configurations méthodiques : une base de données**

La disponibilité, sous forme numérique, d'un large corpus structuré et annoté d'exemples permet donc de découvrir avec une plus grande facilité des configurations textuelles récurrentes qui pourraient facilement, sinon, passer inaperçus. Dans ce contexte, on pourrait légitimement se demander si le numérique n'est pas ici qu'un outil simple et neutre, un outil qui rend les opérations habituelles plus aisées et plus rapides, certes, mais sans changer en profondeur les modalités mêmes de la recherche. On pourrait le croire, surtout parce qu'en aucun lieu n'est ici intervenue une approche proprement computationnelle, c'est-à-dire automatisée, statistique, ou probabilistique. L'ensemble des exemples a été détecté à travers une lecture traditionnelle des romans, leur classification a été faite un par un, de manière pour ainsi dire manuelle et avec des catégories sémantiques et/ou structurelles, c'est-à-dire qualitatives.

Cependant, les modifications qu'introduit la présence d'une base de données contenant les exemples de descriptions avec leurs mot-clés sont tout sauf négligeables. En effet, si de telles découvertes deviennent possible grâce à l'existence d'une base de données d'exemples, cette base de données intervient activement, en quelque sorte, dans le processus de recherche, dont elle modifie non seulement les procédures concrètes, mais encore l'approche méthodologique. Cette approche devient plus formalisée, plus inductive, et son caractère inductif et itératif acquiert davantage de relief. Au final, on pourrait dire que les rapports entre texte et théorie, et entre texte et chercheur, se modifient, l'outil intervenant dans cette relation pour permettre au chercheur une nouvelle appréhension du texte, et une nouvelle manière de modéliser un aspect précis du texte pour finalement construire une typologie, puis une théorie de son fonctionnement.

Pour illustrer mon propos, je voudrais brièvement rapporter ce que la base de données de descriptions contient, et surtout, quel processus a mené à l'existence de cette base de données. La base de données répertorie, dans ce projet, toutes les descriptions identifiées dans trente-deux romans français parus entre 1760 et 1800. Les auteurs représentés sont autant des auteurs incontournables (comme Rousseau ou Diderot) que des auteurs majeurs (comme Rétif de la Bretonne ou Sénac de Meilhan) et de second voire troisième rang (comme Tiphaigne de la Roche, Dorat ou Bellin de la Liborlière)<sup>12</sup>. Tandis que les thèmes des romans sont variés (romans philosophiques, libertins, sentimentaux, utopiques, etc., sont représentés), leurs formes narratives sont limités au roman épistolaire, au roman-mémoires et au roman à la troisième personne.

---

12 Pour un répertoire complet des ouvrages contenus dans le corpus, voir SCHÖCH, *La Description double*, op. cit. p. 347-351.

Sur la base d'une définition des deux notions de la description courantes à l'époque, toutes les descriptions dans les trente-deux romans ont été identifiées et répertoriées, avec leur références précises et le texte des descriptions, dans la base de données. Cette base de données contient au final environ 2000 descriptions ainsi que 500 instances de lexique pertinent et 240 remarques métadiscursives. Elle comporte en outre 120 mots-clés différents associées aux descriptions, chaque description étant classifiée à l'aide d'environ 10 à 15 mots-clés (par exemple, décennie et forme narrative du roman, objet de la description, éléments et positions de la motivation narrative). La figure 4 montre l'interface d'utilisateur de la base de données, au moment où une requête avec plusieurs mots-clés combinés vient d'être effectuée.

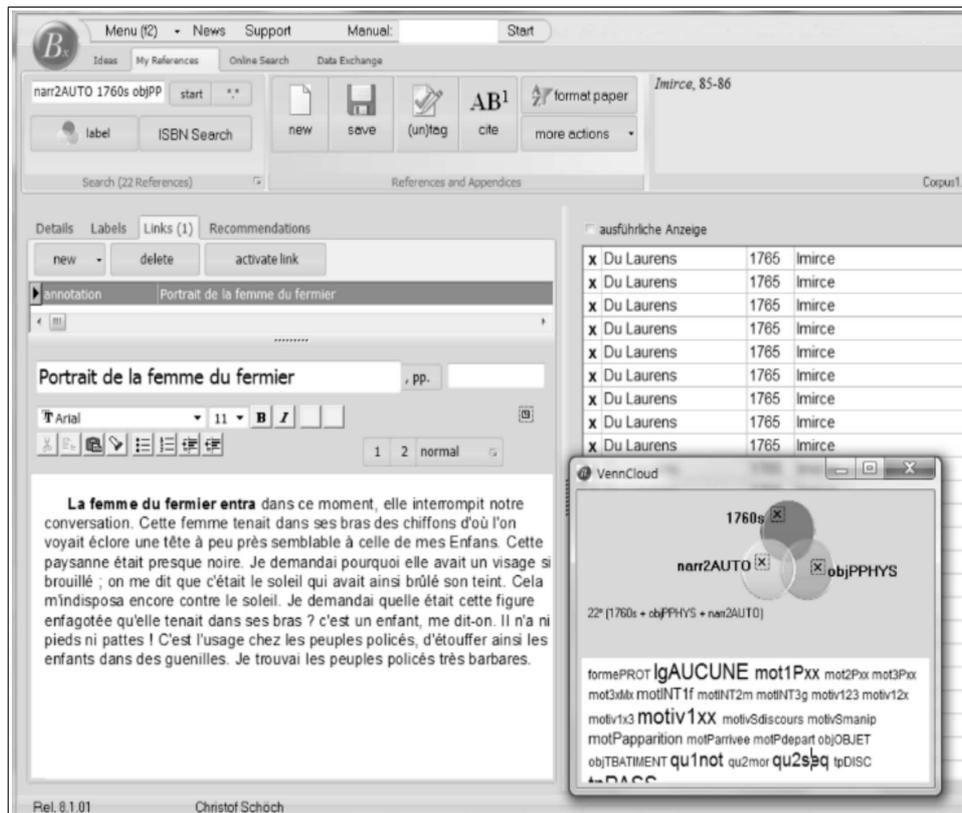


Fig. 4 : L'interface d'utilisation de la base de données Bibliographix<sup>13</sup>.

En bas, à la droite, on voit la requête effectuée, avec trois mots-clés pris ensembles : la décennie (années 1760), l'objet de la description (portrait physique) et le type de narration (autodiégétique, c'est-à-dire roman-mémoires). À droite, au-dessus de la requête, on voit la liste des entrées trouvées, avec l'exemple choisi. À gauche, en bas, on voit le texte de l'exemple. Cela montre que la base de données permet, dans un premier temps, de retrouver rapidement, parmi les 2000 exemples, ceux avec une certaine combinaison de propriétés, ce qui change déjà profondément la pratique quotidienne et l'accès aux matériaux sur lesquels on travaille.

Or, les véritables enjeux d'une telle base de données ne résident pas dans une consultation plus rapide des textes et de leur annotations, mais bien dans le fait que cela change la nature même de la méthode d'analyse. En effet, on se rend compte rapidement, surtout dans le processus de construction progressive de la base de données et des mots-clés, que l'ordinateur devient, d'un simple outil de stockage et de découverte des textes, un véritable *outil heuristique* : Willard McCarty emploie, dans un ouvrage fondamental sur la question de la modélisation numérique des données

13 Logiciel Bibliographix, version 7, <http://bibliographix.de>.

textuelles, ce terme de « heuristic device », et ce qu'il formule de manière générale concernant les humanités numériques est précisément le cas avec une telle base de données : « computing affects analysis in the humanities beyond simply fetching, counting and formatting data [...] - in other words, it affects analysis itself rather than its scope, speed or convenience<sup>14</sup>. »

Dans le processus de construction de la base de données, c'est-à-dire pendant la phase où les descriptions nouvellement identifiées ont été progressivement ajoutées à la base de données et classifiés sous diverses perspectives à l'aide d'un système plus ou moins dynamique et évolutif de mots-clés, on constate que la base de données permet ou même exige une certaine formalisation des catégories en même temps qu'une constante adaptation des catégories et de leur emploi. En fait, on peut dire que l'établissement progressif de la base de données suit lui-même un mouvement herméneutique, qui oscille en permanence entre une pensée à partir de l'ensemble et du résultat, c'est-à-dire le nombre limité de catégories qui cherchent à décrire des caractéristiques des descriptions existant dans la base de données, et une pensée à partir des exemples individuels, c'est-à-dire les nombreuses descriptions elles-mêmes dont il s'agit de comprendre et de décrire le fonctionnement et la fonction. Toute nouvelle description ajoutée à la base de données devient une mise en question potentielle du système des mots clés, système auquel la description résiste, de manière à ce que le système des mots-clés est en permanence dérangé, doit être adapté, nuancé et affiné. En fin de compte, il s'établit un processus qui prend les descriptions comme base d'une typologie établie de manière inductive, de sorte qu'on peut parler d'une modélisation du fonctionnement des descriptions par le système des mots-clés (voir fig. 5).

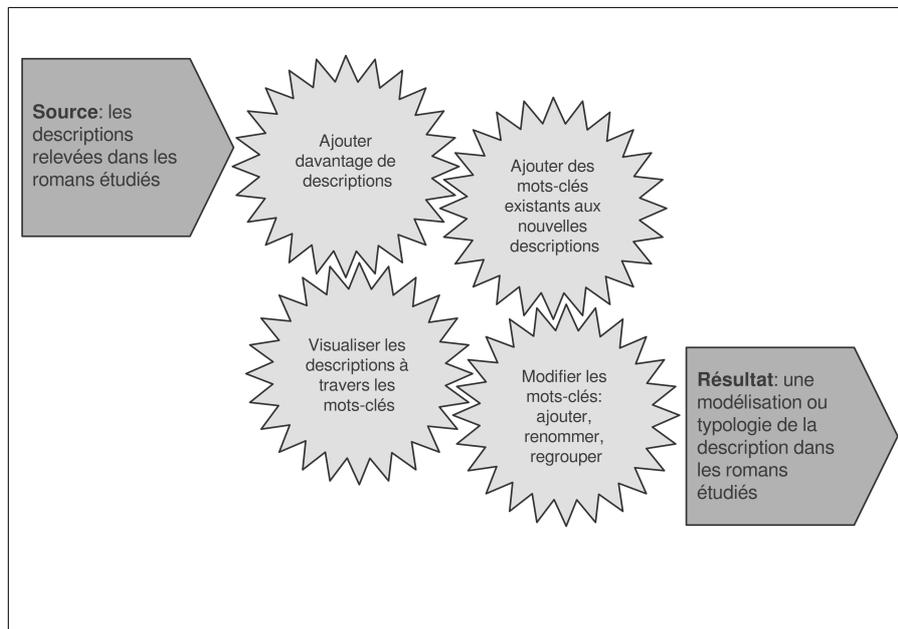


Fig. 5 : Le processus itératif de l'établissement de la base de données.

De manière analogue, dans le processus de consultation de la base de données, le rapport entre texte et théorie est profondément modifié. On peut parler véritablement d'une sorte de lutte avec les matériaux qui ne se laissent jamais véritablement cantonner dans une classification donnée ou expliquer par une typologie ou une théorie unique et cohérente. La résistance des matériaux textuels envers une prise en charge par une théorie est bien plus grande, dans ce cas de figure, que dans une

14 MCCARTY, Willard, *Humanities Computing*, Basingstoke & New York, Palgrave Mcmillan, 2005, p. 207.

approche traditionnelle des mêmes matériaux, puisque la base de données pointe toujours vers des exemples qui échappent à la règle qu'on a voulu établir ; paradoxalement, les exemples s'individualisent sous l'emprise des catégories. Comme le dit encore Willard McCarty : « the massed evidence mechanically returned resists the projection of argument into evidence by giving us more than simply the examples which work<sup>15</sup>. »

L'outil devient de cette manière un médiateur actif ou en tout cas résistant, intervenant comme un tiers entre le texte et le chercheur<sup>16</sup>. Il ne faudrait certes pas s'adonner à l'illusion que le livre imprimé donnerait un accès naturel, direct et complet au texte, puisque ce dernier réalité échappe en réalité, par sa linéarité, à toute emprise globalisante. Cependant, l'outil numérique modifie le rapport au texte, dans la mesure où il en restreint et enrichit l'accès tout à la fois : il rend l'accès plus restreint, parce que les textes s'y trouvent découpés en morceaux, isolés de leur contexte narratif plus large, et coupés de la linéarité constitutive du texte intégral du roman. Cette restriction opère donc avant tout une certaine décontextualisation. Mais l'outil enrichit en même temps l'accès au texte, parce que chaque morceau du texte se trouve connecté, par la requête, et de manière flexible dépendant de chaque requête, aux autres morceaux de texte similaires selon la définition de la requête. Ce que la lecture et l'annotation au crayon d'un livre ne font qu'amorcer ou suggérer, la base de données le rend explicite, manifeste et surtout externalisé. Autrement dit, l'outil et sa capacité de trier les morceaux de texte selon des critères toujours différents, produit des configurations ad hoc d'exemples, et opère donc également une recontextualisation. D'une certaine manière, la base de données fonctionne ici comme un outil permettant une pratique de lecture qui a été appelé « deformance<sup>17</sup> » (mot-valise formé, en joignant *deformation* et *performance*, par Jerome McGann et Lisa Samuels) ou encore « deforming humanities<sup>18</sup> » (concept développé récemment par Stephen Ramsay) : l'idée essentielle est que plutôt qu'une décomposition, de tels outils permettent une manipulation, transformation ou reconfiguration qui permet et même demande ensuite une interprétation. C'est en tout cas une telle re-configuration du texte qu'est l'extraction de toutes les descriptions pour les réunir dans une base de données, base qui en brise la linéarité et le contexte narratif pour construire des rapports toujours nouveaux et mobiles.

### **3. Reconfigurations de la recherche – les infrastructures numériques**

L'avènement du numérique dans la recherche va bien au-delà de projets individuels tels que celui qui a été au centre de mon attention jusqu'ici. Ce projet, terminé depuis un moment, représente même, par sa petite taille et son professionnalisme limité, une modalité de la recherche numérique qui n'est plus tout à fait à la hauteur du moment. Ce vers quoi pointent seulement ce genre de projets sont, en fin de compte, des modifications qui concernent de manière assez générale la recherche en sciences humaines. Ces modifications produisent de nouvelles tensions dans le champ de la recherche en sciences humaines. Par exemple, il s'établit un rapport presque entièrement inédit (en dehors de quelques tentatives au moment du structuralisme, mais sans l'atout du numérique) entre la recherche herméneutique et qualitative des sciences humaines (le « close reading » mais aussi la synthèse

---

15 *Ibid.*, p. 128.

16 *Ibid.*, p. 24 : « By nature modelling defines a ternary relationship in which it mediates epistemologically between modeller and modelled, between researcher and data or between theory and the world. »

17 SAMUELS, Lisa, and MCGANN, Jerome, « Deformance and Interpretation », in *New Literary History* 30.1, 1999, p. 25-56.

18 RAMSAY, Stephen, *Reading Machines. Toward an Algorithmic Criticism*, Urbana, Univ. of Illinois Press, 2011.

contextualisante), d'une part, et la recherche analytique et quantitative (le « distant reading » et la recherche de structures récurrentes), d'autre part.

Le tournant numérique, comme moteur d'un renouveau méthodologique dans les sciences humaines, crée également de nouveaux rapports entre le chercheur individuel et les autres disciplines. En effet, l'image du chercheur isolé, dans la confrontation solitaire entre un individu et les textes, n'est désormais plus adéquate, si elle l'a jamais été. Le travail en solitaire devient de moins en moins praticable et, bien sûr, souhaitable. Certes, comme nous le rappelle la célèbre phrase attribué à Bernard de Chartres par Jean de Salisbury, nous sommes tous depuis toujours des nains juchés sur des épaules de géants. Or, avec l'avènement du numérique, la spécialisation toujours plus grande entre en conflit avec l'interdisciplinarité fondamentale d'une activité de recherche qui allie l'approche herméneutique des sciences humaines aux formalisations et aux calculs de l'informatique et de la statistique.

Chaque chercheur doit trouver une solution qui convient à sa situation particulière face à ces tensions, mais il incombe aux infrastructures numérique de les expliquer et de les amortir. Le rôle des infrastructures numériques est de faciliter l'emploi des nouvelles méthodes numériques qui peuvent maintenant s'appuyer de plus en plus sur un patrimoine disponible sous forme numérique. Deux types d'acteurs entrent dans une nouvelle configuration du champ de la recherche pour former une telle infrastructure numérique pour la recherche : d'un côté, les acteurs relativement établis et convertis au numérique, qui sont concernés avant tout avec l'archivage à long terme, la facilité de découvrir les contenus grâce aux métadonnées, et la dissémination du patrimoine culturel ainsi que des résultats de la recherche ; les acteurs majeurs dans ce champ sont bien sûr les bibliothèques nationales et européennes, plus que jamais numériques, comme la Bibliothèque nationale de France ou Europeana, mais également des initiatives nouvelles comme les la bibliothèque numérique Deutsche Digitale Bibliothek en Allemagne, la plateforme de diffusion et de publication de la recherche OpenEdition en France, ou les méta-catalogues comme le Karlsruher Virtueller Katalog<sup>19</sup>. De l'autre côté, il y a les initiatives dont l'objectif est de mettre à disposition des chercheurs une infrastructure technique, méthodologique, informationnelle et sociale pour la recherche numérique en sciences humaines ; les deux acteurs oeuvrant au niveau national et européen dans ce domaine sont DARIAH (Digital Research Infrastructure for the Arts and Humanities) pour les sciences humaines de manière plus générale et CLARIN (Common Language Resources and Technology Infrastructure) pour les ressources linguistiques<sup>20</sup>. Ces infrastructures numériques pour la recherche en sciences humaines sont vouées à se transformer de projets qu'ils sont peut-être encore aujourd'hui en des véritables institutions européennes pour la recherche, afin de pouvoir remplir leurs fonctions à long terme<sup>21</sup>.

Les acteurs comme DARIAH et CLARIN doivent devenir des intermédiaires entre le langage et les méthodes littéraires et le langage et les méthodes informaticiennes. Il semble exister, dans ce contexte, une méprise face aux méthodes numériques, qui voudrait voir dans de tels approches des

---

19 Voir les sites web de ces institutions : <http://gallica.bnf.fr/>, <http://europeana.eu/>, <http://www.deutsche-digitale-bibliothek.de/>, <http://www.openedition.org/>, <http://www.ubka.uni-karlsruhe.de/kvk.html>.

20 Voir <http://dariah.eu> ainsi que <http://clarin.eu>.

21 Ceci dans le cadre de leur mission défini par le *European Strategy Forum on Research Infrastructures* (ESFRI ; voir [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)), et fortes de leur statut légal de *European Research Infrastructure Consortium* (ERIC ; voir [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=eric](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric)).

principes nouveaux et extérieurs aux sciences humaines, comme la formalisation, l'automatisation et l'empirisme, principes qui chasseraient en quelque sorte l'humain de sciences humaines : cette composante humaine seraient des compétences comme la production de sens individuelle mais transparente ou l'interprétation fondée méthodiquement et contextuellement. Il paraît qu'il y ait ici une double méprise : d'une part, parce que les nouvelles méthodes numériques ne remplacent pas l'interprétation, mais la soutiennent ou la nuancent et, certes, parfois, la mettent en question. D'autre part, parce que les informations produites par les nouvelles méthodes d'analyse numérique ne sont pas des résultats ou des réponses toutes faites, mais simplement des vues particulières sur les textes et des résultats intermédiaires qui ont besoin d'être interprétés et contextualisés tout comme les textes eux-mêmes. En effet, la « modélisation » théorisée par Willard McCarty<sup>22</sup> ou la « formalisation » décrite par Joris van Zundert<sup>23</sup> sont des intermédiaires entre le qualitatif et le quantitatif, entre l'herméneutique et les sciences.

De même, la dépendance sur les infrastructures des données de la recherche que sont les bibliothèques et les archives se double maintenant de l'appui nécessaire sur les infrastructures de recherche numériques, à travers le développement et la mise à disposition de nouveaux outils de manipulation et d'analyse des textes, d'images et d'objets ainsi que de services de traitement et de stockage de données de recherche. Pour que les textes et les outils puissent s'allier de manière fructueuse pour la recherche, des infrastructures numériques doivent analyser et recommander des standards et des modalités d'échange de données. Les innombrables outils développés un peu partout dans le monde pour les humanités numériques ne sauront réaliser leur capacité de révolutionner les pratiques de recherche que s'ils sont liés aux données, tant au sein d'infrastructure numériques que grâce aux formats standard des données, et si les compétences requises pour leur bonne utilisation sont enseignés dans les écoles et les universités.

L'avènement du numérique portait avec soi la promesse d'une désintermédiation, qui mettrait tout un chacun dans la position d'accéder, de manipuler et de comprendre la totalité du patrimoine culturel et même d'y contribuer ; en réalité, nous vivons en même temps une certaine reintermédiation, justement parce que les moyens techniques et les compétences méthodologiques qui seraient nécessaires pour réaliser la désintermédiation doivent également s'apprendre.

## Conclusion

Tout se tient, d'une certaine manière, mais d'une manière extrêmement souple et complexe : on découvre de nouvelles configurations micro-structurelles au niveau des textes grâce à la disponibilité d'outils comme, par exemple, les bases de données ; ces nouveaux outils génèrent de nouvelles méthodes de modélisation et de formalisation qui soulèvent la question de la nature même de la recherche en sciences humaines et de leur rapport à l'informatique et la statistique ; enfin, les infrastructures numériques doivent rendre possible et fructueux ces nouvelles configurations et ces re-configurations, et être utilisés au mieux par les chercheurs.

Mais les rapports entre les projets de recherche individuels et les infrastructures numériques est complexe, comme le rappelle Amy Friedlander dans un article récent sur l'état actuel et les enjeux

---

22 McCARTY, op. cit., *passim*.

23 VAN ZUNDEERT, Joris, et al., « Cultures of Formalization. Towards an Encounter of Humanities and Computing », in *Understanding Digital Humanities*, Barry, David (dir.), London, Palgrave Macmillan, 2012, p. 279-294.

futurs de la recherche numérique en sciences humaines et sociales :

The current landscape in digital scholarship is replete with examples of bottom-up enterprise ; the open question is whether and how to stimulate large-scale coherence without stymieing individual enterprise, frustrating existing self-organization, or threatening the individualism that traditionally characterizes humanities research. The infrastructure itself is so costly and the potential gains from collaborative research are so appealing that some form of loose coordination seems appropriate<sup>24</sup>.

L'enjeu pour les infrastructures numériques sera, en tout cas, d'être des acteurs fiables de services à long terme tout en facilitant la recherche innovatrice et toujours en devenir. L'enjeu du numérique pour les chercheurs sera, en revanche, de prendre la mesure des nouvelles configurations textuelles et méthodiques que le numérique permet. Surtout, il s'agira de nous approprier le numérique, à toutes les étapes de la recherche, pour façonner activement les nouvelles configurations de données, d'outils et de questions, pour que le numérique serve au mieux nos besoins de chercheurs en sciences humaines.

## Références

COLLECTIF, Article « Configuration (informatique) », *Wikipédia France*, version du 23 octobre 2012, [http://fr.wikipedia.org/w/index.php?title=Configuration\\_%28informatique%29&oldid=84473553](http://fr.wikipedia.org/w/index.php?title=Configuration_%28informatique%29&oldid=84473553).

COLLECTIF, « Manifeste des humanités digitales », in *That Camp Paris*, 26 mars 2011, URL : <http://tcp.hypotheses.org/318>.

*Dictionnaire de l'Académie française*, 5e édition (Paris, 1798), édition électronique, Chicago, The ARTFL Project, 2004-2012, <http://artfl-project.uchicago.edu/content/dictionnaires-dautresfois>.

*Dictionnaire de français « Littré »* (Paris, 1863-1877), édition électronique, Paris, Reverso, 2010, URL : <http://litre.reverso.net/>.

FRIEDLANDER, Amy. « Asking Questions and Building a Research Agenda for Digital Scholarship », in *Working Together or Apart. Promoting the Next Generation of Digital Scholarship*, Washington, CLIR, 2009, p. 1-15, URL : <http://www.clir.org/pubs/abstract/reports/pub145>.

HAMON, Philippe, *Du descriptif* (1981), Paris, Hachette, 1993.

MCCARTY, Willard, *Humanities Computing*, Basingstoke & New York, Palgrave Mcmillan, 2005.

*Online Etymology Dictionary*, HARPER, Douglas (dir.), 2001-2012, URL : <http://www.etymonline.com/>.

RAMSAY, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. Urbana: Univ. of Illinois Press, 2011.

RÉVÉRONI SAINT-CYR, Jacques Antoine de, *Pauliska ou la perversité moderne* (1795), DELON, Michel (éd.), Paris, Desjonquères, 1991, p. 92.

SAMUELS, Lisa, and MCGANN, Jerome, « Deformance and Interpretation », in : *New Literary History* 30.1, 1999, p. 25-56.

SCHÖCH, Christof, *La description double dans le roman français des Lumières, 1760-1800*, Paris, Éditions Classiques Garnier, « L'Europe des Lumières », 2011.

VAN ZUNDERT, Joris, ANTONIJEVIC, Smiljana, BEAULIEU, Anne, VAN DALEN-OSKAM, Karina, ZELDENRUST, Douwe, ANDREWS, Tara, « Cultures of Formalization. Towards an Encounter of Humanities and Computing », in *Understanding Digital Humanities*, BARRY, David (dir.), London, Palgrave Macmillan, 2012, p. 279-294.

ZOLA, Émile, *L'Œuvre* (1885/86), EHRARD, Antoinette (éd.), Paris, Flammarion, 1974.

---

24 FRIEDLANDER, Amy, « Asking Questions and Building a Research Agenda for Digital Scholarship », in *Working Together or Apart. Promoting the Next Generation of Digital Scholarship*, Washington, CLIR, 2009, p. 12, URL : <http://www.clir.org/pubs/abstract/reports/pub145>.