



CLASSIQUES
GARNIER

JACQUOT (Clémence), « Du texte aux données. Texte digital, philologie numérique et dispositif d'attention », *Études digitales*, n° 1, 2016 – 1, *Le texte à venir*, p. 41-67

DOI : [10.15122/isbn.978-2-406-06193-9.p.0041](https://doi.org/10.15122/isbn.978-2-406-06193-9.p.0041)

La diffusion ou la divulgation de ce document et de son contenu via Internet ou tout autre moyen de communication ne sont pas autorisées hormis dans un cadre privé.

© 2016. Classiques Garnier, Paris.
Reproduction et traduction, même partielles, interdites.
Tous droits réservés pour tous les pays.

JACQUOT (Clémence), « Du texte aux données. Texte digital, philologie numérique et dispositif d'attention »

RÉSUMÉ – Cet article propose une réflexion sur les transformations du texte littéraire imprimé au texte digital enrichi à des fins de recherche et d'interprétation. Cette transformation au format numérique met en jeu non seulement un changement ontologique de l'objet texte mais aussi une mutation des pratiques phénoménologiques de lecture et de manipulation des "données textuelles" en vue d'une nouvelle approche de l'*attention*.

ABSTRACT – This article offers a reflection on the transformation of printed literary text into enriched digital text for research and interpretation purposes. This transformation into a digital format involves not only an ontological change of the text object but also a mutation of the phenomenological practices of reading and the manipulation of "textual data" with a view to a new approach to attention.

DU TEXTE AUX DONNÉES

Texte digital, philologie numérique et dispositif d'attention

Chaque nouveau support permet de nouveaux traitements, mais ne les détermine pas. Une nouvelle technologie définit précisément un nouveau mode d'interaction entre un traitement et un support¹.

Accordant à juste titre au numérique le statut de « mutation épistémique et archiviologique majeure² », Bernard Stiegler propose d'envisager la réaction de l'esprit humain face à la révolution digitale non tant comme une adaptation de surface au nouveau média, mais bien, par l'extériorisation propre à la « vie de l'esprit », comme un bouleversement ontologique de la pensée elle-même³.

Le support numérique renouvelle en effet le rapport au texte, tant dans sa lecture que dans son écriture et, par extension, les modalités

1 François Rastier, *Arts et sciences du texte*, Paris, PUF, 2001, p. 74-75.

2 Bernard Stiegler, « Pharmacologie de l'épistèmè numérique », in Stiegler, B. [dir.], *Digital Studies. Organologie des savoirs et technologies de la connaissance*, Paris, FYP Éditions, 2014, p. 14.

3 Stiegler reprend notamment la notion d'extériorité technique à Leroi-Gourhan et l'article au concept de « pharmakon » : « *Tout objet technique est pharmacologique* : il est à la fois poison et remède. Le *pharmakon* est à la fois ce qui permet de prendre soin et ce dont il faut prendre soin, au sens où il faut y faire [...]. Cet à la fois est ce qui caractérise la pharmacologie qui tente d'appréhender par le même geste le danger et ce qui sauve. Toute technique est originellement et irréductiblement ambivalente : l'écriture alphabétique, par exemple, a pu et peut encore être aussi bien un instrument d'émancipation que d'aliénation. » (Bernard Stiegler, article « Pharmakon », *Ars Industrialis*. [URL : <http://arsindustrialis.org/pharmakon> ; consulté le 22/02/2015]. Voir aussi Bernard Stiegler, *Prendre soin de la jeunesse et des générations*, Paris, Flammarion, 2012, p. 19 sqq.

d'accès au savoir qui en découlent¹. Au sein de ce que l'on désigne désormais par diverses appellations – *Humanités numériques*, *Digital Studies* ou encore *Knowledge Design* – les disciplines des « arts et sciences du texte² » occupent une place singulière. Leur passage à l'ère du numérique engage un changement de paradigme dans la construction de leur objet d'étude comme dans leurs gammes d'application³.

Le chercheur en littérature, en manipulant et en construisant le texte digital à partir d'un support imprimé, modifie effectivement sa méthode d'étude du texte. Il s'approprie notamment certains outils du traitement automatique des langues et se confronte, grâce à eux, à de « nouveaux observables⁴ », en particulier quantitatifs, qu'il s'agit alors de comprendre et d'interpréter. Ce processus de transformation du texte imprimé en support quantifiable à des fins d'analyse linguistique et herméneutique n'est pas neuf : il remonte à la création des machines mécanographiques et s'est développé au fil des progrès des technologies informatiques, ainsi qu'en témoignent les premiers travaux de lexicométrie et de statistique sur les textes littéraires de Charles Muller et de Pierre Guiraud dans les années 1960⁵.

Cette transformation trouve aujourd'hui une expression un peu différente dans les approches textométriques⁶ et stylométriques⁷, puisque le texte digital – balisé et étiqueté par le chercheur – permet le passage d'une « mesure de mots » à une « mesure des textes⁸ ». Toutefois, en

1 Voir à ce propos : Franck Cormerais, « Humanités digitales et (ré)organisation du savoir », in Le Deuff, O. [dir.], *Le Temps des humanités digitales. La mutation des sciences humaines et sociales*, Paris, FYP Éditions, 2014, p. 129-142 et Milad Doueihy, « Quête et enquête », *ibid.*, p. 7-10.

2 Nous reprenons ici la formule de Rastier dans son ouvrage éponyme (François Rastier (2001), *op. cit.*).

3 François Rastier, *La Mesure et le grain. Sémantique de corpus*, Paris, Champion, 2011, p. 33-34.

4 *Ibid.*, p. 50 *sqq.*

5 Charles Muller, *Étude de statistique lexicale le vocabulaire du théâtre de Pierre Corneille*, Paris, Larousse, 1967 ; Pierre Guiraud *Problèmes et méthodes de la statistique linguistique*, Paris, PUF, 1960. Voir aussi, pour une mise au point théorique : Charles Muller, « Lexicologie, statistique lexicale et critique littéraire » in *Langue française, débats et bilans. Recueil d'articles, 1986-1993*, Paris, Champion, 1993, p. 54-60.

6 Nous renvoyons à ce propos aux articles de Bénédicte Pincemin, en particulier : « Sémantique interprétative et textométrie – Version abrégée », in *Corpus* [en ligne], n° 10, 2011 [URL de référence : <http://corpus.revues.org/2121> ; consulté le 27/03/2014].

7 Voir notamment à ce propos : D.I. Holmes, « The Evolution of Stylometry in Humanities Scholarship », in *Literary and Linguistic Computing*, Volume 13/3, 1998, p. 111-117.

8 François Rastier (2011), *op. cit.*, p. 49.

reprenant sa place dans la dénomination même d'une méthodologie au carrefour de la littérature et de l'informatique, le texte, discrètement, se métamorphose. Construit pour apporter des « données textuelles », il porte les traces des métadonnées qu'on lui attribue pour que l'outil informatique puisse opérer différents types d'analyse systématisée¹.

Le rapport du texte digital à la construction du savoir passe ainsi par des dispositifs de lecture particuliers et surtout dépendants d'une préparation préalable du texte au moyen des opérations de balisage que nous venons d'évoquer. La transformation du texte en corpus numérisé induit dès lors des stratégies renouvelées d'exploration et d'attention² portée au texte. L'interface proposée tant par les éditions numériques savantes que par les logiciels et outils d'analyse textométriques mobilise non seulement les modalités de lecture propre au Web³, mais aussi une grille d'analyse textuelle induite par l'outil lui-même et illustrant de ce qu'on pourrait appeler, avec Yves Citton, une singulière « écologie de l'attention⁴ ».

Menacé de céder à la dissémination de son objet et de son attention⁵, l'enjeu de la construction et de l'exploration du corpus digital est d'articuler les entités concrètes qui composent le texte et l'enjeu interprétatif de son analyse. Cette pratique repose sur une reconfiguration des moyens d'analyse par la formulation de requêtes, mais aussi, dans une mesure qu'il convient d'établir et de problématiser, par la conception même et le fonctionnement de l'outil informatique. Celui-ci pose un cadre, met en place un dispositif d'enquête et de construction du savoir⁶

1 Il devient alors un « corpus » informatisé. Le balisage est une opération de formatage du texte.

2 Bernard Stiegler, *De la misère symbolique, t. 1. L'Époque hyperindustrielle*, Paris, Galilée, 2004.

3 Nous renvoyons en particulier, pour le détail de ces modalités, à l'ouvrage collectif de Dominique Boullier, Franck Ghitalla, Pergia Gkouskou-Giannakou, Laurence Le Douarin et Aurélie Neau, *L'Outre-lecture. Manipuler, [s']approprier, interpréter le Web*, Paris, Bibliothèque Centre Pompidou, 2003.

4 Yves Citton, *Pour une écologie de l'attention*, Paris, Seuil, 2014.

5 Le risque de la dissémination est bien documenté et a fait l'objet de nombreuses études en ce qui concerne l'usage du web et le clivage entre *deep attention* et *hyper attention* problématisé par Katherine Hayles dans : « Hyper and Deep Attention : the Generational Divide in Cognitive modes », in *Profession*, p. 187-199 [URL de référence : http://www.english.ufl.edu/da/hayles/hayles_hyper-deep.pdf; consulté le 25/11/2014].

6 Nous l'envisageons ici dans le sens deleuzien du terme comme l'indice d'un « régime » ou d'un « opérateur de visibilité », comme le définit Deleuze dans « Qu'est-ce qu'un

dont le potentiel heuristique se fonde d'une part sur l'émergence de résultats grâce aux requêtes formulées et, d'autre part, sur leur capacité à être interprétés¹.

Or, la formulation des requêtes invite à interroger le point focal adopté pour l'analyse², ainsi que la reconfiguration éventuelle des problématiques de recherche au contact des outils utilisés. Elle soulève ainsi les questions suivantes : comment le texte digital enrichi peut-il conditionner ou, du moins, orienter la construction d'abstractions ou la conception d'objets d'étude tels que l'outil informatique puisse y répondre³ ? Quels liens peut-on en effet établir entre l'ambition heuristique du texte numérique « enrichi » et l'objectif herméneutique visé par la lecture littéraire et stylistique d'un texte ? Quelle valeur objective peut-on, dans ce contexte, attribuer à l'émergence de phénomènes observables ?

L'« ENRICHISSEMENT » DU TEXTE NUMÉRISÉ : UN OUTILLAGE INTERPRÉTATIF ?

La linguistique de corpus, et par extension, les disciplines littéraires fondées sur ses bases méthodologiques, disposent désormais d'un atout

dispositif? », *M. Foucault philosophe. Rencontre internationale. Paris 9, 10, 11 janvier 1988*, Paris, Seuil, 1989.

- 1 Nous nous inscrivons ici dans le prolongement des réflexions récentes sur les *Digital Humanities* synthétisées par David M. Berry dans son article « Introduction : Understanding Digital Humanities » (in : David M. Berry [dir.], *Understanding digital humanities*, Palgrave, Macmillan, 2012, p. 1-20 : « These subtractive methods of understanding culture (*episteme*) produce new knowledges and methods for the control of memory and archives (*techné*). They do so through a digital mediation, which the digital humanities are starting to take seriously as their problematic. » (p. 2).
- 2 La question est notamment posée par Damon Mayaffre dans son article « Les corpus réflexifs : entre architextualité et hypertextualité », in *Corpus* [en ligne], n° 1, 2002 [URL de référence : <http://corpus.revues.org/index11.html> ; consulté le 17/01/2011].
- 3 Cf. Éric Guichard, « L'internet et les épistémologies des SHS », in *Sciences/Lettres* n° 2, 2013 [URL de référence : <http://barthes.ens.fr/articles/Guichard-RSL.html> ; consulté le 02/11/2014] et Tim Hitchcock, « Big Data for Dead People : Digital Readings and the Conundrums of Positivism », in *Historyonics*, 2013 [URL de référence : <http://historyonics.blogspot.co.uk/2013/12/big-data-for-dead-people-digital.html> ; consulté le 02/11/2014]. Ces deux chercheurs posent des questions similaires rapportées à la recherche en histoire et en géographie par exemple.

considérable : l'enrichissement hypertextuel des corpus et l'essor des bases de données numériques. François Rastier montre que ce développement de la linguistique de corpus, fondé par conséquent sur celui de l'outil informatique, entraîne même « un renouvellement inattendu de la philologie, [...] la philologie numérique¹ ». L'ordinateur contribue à redéfinir la nature et les enjeux du support des textes, et, plus précisément, offre de nouvelles techniques pour penser et interpréter ces derniers. À cette première affirmation, ajoutons néanmoins une nuance : si l'outil informatique ne permet pas *stricto sensu* d'interpréter les textes – du moins pas dans le sens herméneutique que l'on attribue généralement au terme dans le domaine littéraire –, il offre, cependant, un nouveau rapport au support textuel grâce à la construction d'un corpus annoté.

UN NOUVEAU CORPUS D'ÉTUDE LITTÉRAIRE ?

Dans le champ des études philologiques, historiques ou littéraires, le terme de « corpus » désigne un ensemble de documents susceptibles d'embrasser de manière cohérente un genre ou une question soumis à une étude scientifique. Le terme désigne un répertoire scientifique d'objets de natures diverses, au nombre desquels le texte peut figurer. C'est ce sens que conserve le mot en informatique, auquel s'ajoutent l'automatisation et la quantification des données observées².

Dans le domaine de la linguistique, le corpus signifie en revanche, selon la définition de Sinclair, « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour être utilisé comme échantillon du langage³ », dont il

1 François Rastier (2001), *op. cit.*, p. 73.

2 La définition générique du mot « corpus » dans le domaine spécifique de l'informatique désigne en effet, dans le CNRTL, un « ensemble de données exploitables dans une expérience d'analyse ou de recherche automatique d'informations ».

3 « A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language. » (John Sinclair, « Preliminary recommendations on Corpus Typology », *Technical Report, EAGLES* (Expert Advisory Group on Language Engineering Standards) [en ligne], mai 1996, p. 4 [URL de référence : <http://www.ilc.cnr.it/EAGLES/pub/eagles/corpora/corpus typ.ps.gz>; consulté le 22/04/2012]). (Nous traduisons).

faut interroger la pertinence et la représentativité. Le texte n'apparaît pas dans cette définition qu'en filigrane ; il s'efface alors derrière les données qu'il peut fournir¹.

Depuis l'essor des corpus enrichis pour la recherche en littérature, les chercheurs des différents domaines des sciences du texte s'efforcent de développer une réflexion épistémologique conjointe sur les délimitations respectives des notions de texte et de corpus. L'articulation du texte et du corpus met d'abord en évidence la transdisciplinarité de l'entreprise. Pour Adam et Viprey, dans leur article de la revue *Corpus*, consacré à cette question :

Si les conditions d'un dialogue de l'analyse textuelle des discours et de l'analyse des données textuelles sont réunies, c'est en raison, d'une part, de leur recentrage sur des données résolument textuelles et, d'autre part, du fait que travaillant sur des textes de plus en plus nombreux, différents, complexes et ouverts, nous exigeons qu'ils soient très solidement établis et que soient réunies les conditions d'un traitement informatisé et d'une circulation scientifiquement contrôlée de cette masse de nouvelles données textuelles. Le traitement informatisé de ces données implique la création de réseaux de collaborations entre des équipes de recherche travaillant à une véritable refonte des sciences et disciplines académiques des textes².

Cette définition met en évidence un sens du « corpus » au carrefour des différentes acceptions que nous citons plus haut : l'artefact que constitue le corpus d'étude littéraire s'intègre au sein d'une démarche scientifique et technique qu'il va nourrir. Explorant des recueils de textes et d'œuvres littéraires choisis en fonction d'une quête interprétative à

-
- 1 Rappelons que la démarche initiale de la linguistique de corpus, qui se développe en Europe à partir des années 1960, est empirique : conçue comme un outil heuristique et pédagogique pour préciser la compréhension des mots dans le discours, elle avait notamment pour objectif de pallier les manques de certains dictionnaires traditionnels et de leur apporter des nuances sémantiques grâce à la contextualisation des occurrences étudiées, mais ne poursuivait aucunement d'ambitions herméneutiques. C'est ce que rappelle Wolfgang Teubert, lorsqu'il écrit : « Initially, corpus linguistics was not concerned with the interpretation of unique occurrences. It did not see itself in the tradition of hermeneutics. Corpus linguistics was developed as a methodology to find out things about a language in general, in order to help with the tasks of applied linguistics. » (Wolfgang Teubert, « Corpus Linguistics : An Alternative », in *Semen* [en ligne], n° 27, 2009, § 20 [URL de référence : <http://semen.revues.org/8914> ; consulté le 22/04/2012].)
- 2 Jean-Michel Adam et Jean-Marie Viprey, « Corpus de textes, textes en corpus. Problématique et présentation », in *Corpus* [En ligne], n° 8, novembre 2009, § 3 [URL de référence : <http://corpus.revues.org/index1672.html> ; consulté le 04/04/2013]. (Nous soulignons).

mener, les chercheurs construisent un corpus informatisé destiné à fournir un outil heuristique à leur analyse. Il leur permet de rechercher des informations, d'éprouver des hypothèses et donc, d'offrir à l'interprétation littéraire, grâce aux logiciels informatiques, de nouveaux observables, à condition d'articuler les données textuelles segmentées et quantifiées à la prise en compte de l'intention esthétique inhérente à l'œuvre étudiée. Dès lors, le corpus ainsi défini est déterminé par l'enrichissement du texte brut à partir duquel il est constitué.

L'ANNOTATION DE CORPUS LITTÉRAIRES

Face aux outils de textométrie qui segmentent, sélectionnent, classent et comptabilisent des résultats, un texte brut parle peu. Pour mener à bien une étude systématique et quantifiée, l'élaboration d'un corpus étiqueté, enrichi de métadonnées est une étape cruciale. L'outil informatique (sous la forme de différents logiciels de traitement et de calcul des données) offre alors un nouveau rapport au support textuel, notamment grâce à la technique de l'annotation¹.

C'est le terme d'annotation qui, malgré son usage déjà considérable de *marginalia*, désigne l'activité à la fois catégorielle et prospective d'étiquetage du texte transformé. Aussi, contrairement à la définition généralement admise, l'annotation ne coïncide plus seulement dans ce contexte avec les « rétentions tertiaires² » qui accompagnent et expliquent le contenu ou la forme du texte sous la forme de notes manuscrites

1 Notre réflexion sur les corpus numériques et leurs conditions d'interprétation s'articule autour de ce que Crozat, Bachimont, Cailleau, Bouchardon et Gaillard définissent comme les deuxième et troisième niveaux du texte numérique. En effet, au delà du premier niveau « théorético-idéal », notre propos est d'analyser le passage du texte aux données textuelles comme *manifestation* et comme condition d'interprétation du texte : c'est ce que ces auteurs nomment respectivement le niveau « techno-applicatif » (ou niveau de manifestation) et le niveau « sémio-rhétorique » (ou niveau d'interaction). Cf. Stéphane Crozat, Bruno Bachimont Isabelle Cailleau, Serge Bouchardon, Ludovic Gaillard « Éléments pour une théorie opérationnelle de l'écriture numérique », in *Supports et pratiques d'écriture en réseau*, Bouhaï, N. [dir.], 2011, URL de référence : <http://www.cairn.info/revue-document-numerique-2011-3.htm> [consulté le 14/05/2015].

2 Nous renvoyons à ce propos à : Bernard Stiegler (2008), *op. cit.*, p. 171 *sqq.*

ajoutées en marge, traces mémorielles et geste d'accompagnement de la lecture. Appliquée au corpus, elle prend par exemple la forme de balises XML intégrées dans le texte pour l'exploration approfondie et les traitements du corpus par des logiciels d'analyse textuelle.

Les annotations peuvent en principe porter sur divers niveaux, domaines et types de données textuelles qu'elles qualifient. Ces méta-données se définissent comme un ensemble structuré d'informations dont la fonction est de décrire le corpus et qui offre différentes ressources et entrées de classement et de traitement de ces informations. Rapportées à l'économie du texte littéraire, on peut ainsi distinguer deux principaux paliers d'annotation. Au niveau macrotextuel, on identifie par exemple les informations structurelles¹ et péri-textuelles² du texte ; au niveau microtextuel (c'est-à-dire au niveau du mot ou du groupe de mots), l'annotation mobilise des informations linguistiques, comme l'identification plus ou moins fine de la classe grammaticale du terme ou de la locution étudiés ou encore sa lemmatisation³. Elle permet également de mettre en évidence des réseaux sémantiques variés (champs sémantiques, champs lexicaux, reconnaissance d'entités nommées⁴, etc.).

Ces métadonnées peuvent s'articuler les unes aux autres sans se faire concurrence et font tantôt l'objet d'un étiquetage automatique, lorsque les logiciels de traitement des données sont suffisamment entraînés, tantôt d'une annotation « manuelle » pour renseigner par exemple des phénomènes sémantiques fins non résolus par les algorithmes ou désambigüiser les cas de balisages problématiques ou inexacts⁵.

1 Il s'agit de balises qui informent sur la structuration du texte et ses différentes unités de lecture : le recueil, la section, le chapitre, le poème, le paragraphe, la strophe, le vers, etc.

2 Nous reprenons le terme à Gérard Genette (*Seuils*, Paris, Seuil, 1987). Les informations péri-textuelles sont hétérogènes, elles concernent les éléments paratextuels situés à l'intérieur même de l'œuvre (les titres et sous-titres, les dates de publication, de rédaction, les supports de publication, les tables des matières, les dédicaces et épigraphes, etc.).

3 Le lemme est la forme graphique conventionnellement choisie pour désigner un mot dans le lexique. Elle réduit ainsi toutes les formes fléchies ou les diverses variations graphiques d'un mot à une entrée unique et constitue alors une information importante dans l'étude de la variété du vocabulaire par exemple.

4 Citons par exemple les noms de personne, de lieux, d'organisation et d'événement.

5 Nous pensons en particulier aux opérations de désambigüisation des étiquettes morpho-syntaxiques (l'identification de cas de conversion – « rien » *vs.* « un rien » – ou d'amalgame – « du, des », par exemple), ou de balises sémantiques (« La Sorbonne » désigne ainsi, selon le contexte, un lieu ou une institution).

La notion de corpus désigne en définitive les processus mêmes de construction d'un objet scientifique. Elle engage la responsabilité des personnes qui le constituent à interroger les conditions de sa création – par l'activité d'annotation notamment – et, surtout, l'objectif visé à travers lui. Il s'agit alors d'un dispositif hypertextuel qui détermine dans une large part la mise en place et l'orientation de stratégies attentionnelles mises en œuvre pour pouvoir l'appréhender. Ces stratégies attentionnelles sont en effet, au moins dans un premier temps, « paramétrées », encadrées par les pratiques induites par les logiciels ou les algorithmes de traitement de données textuelles. Ainsi, les logiciels d'étiquetage morphosyntaxiques reposent sur des catégorisations paramétrées, dont le degré de granularité ou l'identification des classes grammaticales peut même être discutable selon le parti pris de l'analyse¹.

LECTURE, INTERPRÉTATION ET RÉGIMES ATTENTIONNELS

LECTURE ET INTERPRÉTATION : DEUX RAPPORTS DISTINCTS AU TEXTE

Dans *L'Avenir des Humanités*, Citton distingue deux rapports au texte littéraire : la lecture et l'interprétation. L'objectif est de cerner les enjeux de l'interprétation et d'en préciser les spécificités. En comparaison, l'activité de lecture apparaît par conséquent sous un jour plus modeste. Elle se définit par une forme de passivité² du lecteur face à l'objet textuel : celui-ci relève, par la lecture, les informations véhiculées par le texte

- 1 Le logiciel *Treetagger* ne distingue pas, dans sa configuration pour le français moderne, les conjonctions de coordination et les conjonctions de subordination (dont le fonctionnement syntaxique respectif est pourtant très différent), mais attribue automatiquement une seule catégorie grammaticale, la conjonction. Seule une ré-annotation manuelle, grâce à un éditeur XML, peut permettre de discriminer ces formes, en vue d'une analyse plus fine des constructions syntaxiques.
- 2 Yves Citton, *L'Avenir des Humanités. Économie de la connaissance ou cultures de l'interprétation ?*, Paris, La Découverte, 2010 : « [...] on peut parler de passivité en ceci que chacun de ces "actes" de lectures ne fait qu'ajouter quelques bits d'informations en stock dont nous disposions originellement, tout en reconduisant à l'identique à la fois notre connaissance du code et les cadres d'analyse qui nous orientent dans notre perception du donné » (p. 45).

sans remettre en cause la « connaissance du code et les cadres d'analyse qui nous orientent dans notre perception du donné¹ ». L'interprétation, au contraire, est caractérisée par l'attention active portée sur le texte. Elle implique « la remise en cause d'au moins un des trois postulats implicites qui gouvernent la lecture conventionnelle² », c'est-à-dire : le code utilisé, la manière de déchiffrer le texte, la pertinence de ce que le texte lui offre, voire (c'est le cas le plus courant) une combinaison de ces trois questions.

Pour illustrer la démarche de l'activité interprétation, Citton reprend une belle description de Deleuze dans ses cours sur le cinéma, qui cherche à traduire le phénomène de « reconnaissance attentive³ » auquel chacun d'entre nous est régulièrement confronté et qui apparaît comme une illustration saisissante de l'activité interprétative. Celle-ci serait alors une manière de former de « petits circuits », de les faire fonctionner en réseau, le temps du tâtonnement, du travail élastique, plastique, de l'attention jusqu'à ce que ces « circuits » invoqués par l'esprit trouvent à s'actualiser, à se dilater.

L'interprétation et l'invention, envisagées de ce point de vue, participent alors du même mouvement, celui du vide, du manque, de la tension initiale de l'esprit – provoquée par la difficulté qu'ont les « circuits » de la mémoire ou de la connaissance à s'enchaîner les uns aux autres – vers la compréhension, vers une reconnaissance proche du concept d'*anamnèse* platonicienne.

Dès lors, deux remarques s'imposent : d'abord, l'interprétation d'un texte qui met en branle ce processus complexe de mise en relation des circuits de l'esprit participe d'une perspective phénoménologique de collecte, de sélection et d'hésitation. Enfin, les régimes de l'attention sollicités par le processus d'interprétation nous apparaissent pluriels, placés sous le signe de la discontinuité et du surgissement. L'attention interprétative portée au texte littéraire envisage en effet ce dernier comme le lieu d'une émergence.

1 *Ibidem.*

2 *Ibidem.*

3 Gilles Deleuze, Cours n° 19 du 18 mai 1982 « Cinéma / Image-Mouvement, novembre 1981-juin 1982 », *La Voix de Gilles Deleuze* [en ligne], site de l'Université Paris-8 Saint Denis [URL de référence : http://www2.univ-paris8.fr/deleuze/article.php3?id_article=157; consulté le 10/02/2015].

L'ATTENTION INTERPRÉTATIVE,
UNE DISPONIBILITÉ FACE À L'OBJET

Le cheminement rhizomatique mis en œuvre par l'interprétation se produit dans un environnement de latence, de suspension¹. Cet état de « disponibilité » à l'objet, de guetteur en repos, trouve une belle définition dans le texte de Simone Weil, *L'attente de Dieu* :

L'attention consiste à suspendre sa pensée, à la laisser disponible, vide et pénétrable à l'objet, à maintenir en soi-même à proximité de la pensée, mais à un niveau inférieur et sans contact avec elle, les diverses connaissances acquises qu'on est forcé d'utiliser. La pensée [...] doit être vide, en attente, ne rien chercher, mais être prête à recevoir dans sa vérité nue l'objet qui va y pénétrer².

Cette latence est l'une des conditions nécessaires de l'interprétation, c'est notamment grâce à elle que devient possible l'appréhension de la polyphonie inhérente à tout texte, et en particulier du texte littéraire. Ainsi, pour Jean-Marie Schaeffer :

Si l'on admet [...] qu'un trait textuel peut produire des effets au niveau de l'attention, donc au niveau de la phénoménologie de la lecture, sans être lui-même attentionnellement traité, on peut dire qu'un des aspects de la relation esthétique au texte réside dans l'exploration attentionnelle consciente de certains traits textuels, qui n'opèrent plus seulement de manière causale mais se trouvent intégrés dans le traitement de l'œuvre comme structure intentionnelle et contribuent ainsi au passage de la lecture monodique à une lecture polyphonique – passage qui semble être la marque par excellence de la lecture esthétique des textes³.

Appréhéné de manière critique, le texte se « défamiliarise » du langage ordinaire : l'attention interprétative portée sur lui l'objective. Elle invite, par le fonctionnement rhizomatique de propagation de l'information ou plutôt de la signification qu'elle produit, à déconstruire pour un temps la progression du texte pour la réagencer sous la forme

1 Yves Citton parle également d'« intervalle », de vide entre la sensation initiale et la réaction différée ou de « vacuole » (*Ibid.* p. 58 et p. 74). Voir aussi : Yves Citton (2014), *op. cit.*, p. 230 *sqq.*

2 Simone Weil, *Attente de Dieu*, Paris, Fayard, 1966, p. 116.

3 Jean-Marie Schaeffer, « Styles attentionnels et relation esthétique », in Jenny, L. [dir.], *Le Style en acte. Vers une pragmatique du style*, Genève, Mētis Presses, 2011, p. 142-143.

des « circuits » évoqués plus haut et que Stiegler nomme, quant à lui, les « rétentions » attentionnelles.

L'attention à la fois diffuse et alertée sur laquelle se fonde l'interprétation opère donc une forme de délinéarisation du texte, créée par la conscience d'une intention esthétique sous-jacente à son écriture. Cette délinéarisation, nous la rapprocherions volontiers de ce que le *Groupe μ* nommait, dans la *Rhétorique de la poésie*, la « lecture tabulaire » du texte¹, c'est-à-dire le « résultat de la superposition des différentes lectures des unités d'un texte² ».

Enfin, lorsqu'on étudie un texte en adoptant un angle d'attaque technique, comme c'est le cas en stylistique par exemple, c'est l'attention interprétative elle-même qui est orientée. L'enjeu de l'interprétation stylistique est avant tout celui d'une attention portée à la forme (celle de son expression comme celle de son contenu). Elle opère un premier acte de filtrage microtextuel – il s'agit de voir *comment* ça s'écrit – et interroge également des relations macrotextuelles qui articulent des phénomènes langagiers ténus (qui semblent glanés à la surface du texte) et les enjeux intentionnels identifiés dans le texte. Ces relations macrotextuelles concernent également les rapports qu'entretiennent respectivement ce que François Rastier nomme le *local* (l'entour immédiat des indices relevés dans le texte) et le *global* (le texte entier, l'œuvre, la culture dont tout cela procède).

On mesure, à nouveau, combien ce type de lecture participe d'un régime attentionnel complexe – disséminer d'abord son attention pour filtrer le texte et synthétiser, faire ensuite converger l'attention pour contextualiser ses remarques, ce que Rastier résume en ces termes :

C'est au palier du texte que la conception commune de la compositionnalité laisse apparaître le plus clairement ses lacunes : en effet, le global y détermine le local et le recompose³.

1 Groupe Mu, *Rhétorique de la poésie. Lecture tabulaire, lecture linéaire*, Paris, Seuil, 1990. Cette idée est partiellement reprise par Jean Peytard lorsqu'il conçoit la lecture comme « déambulation tabulaire ». Cf. Jean Peytard, « Écriture et pointillés de sens : lecture-analyse de deux pages de Proust (La Fin de la jalousie) », *Semen* [En ligne], n° 11, 1999, [URL de référence : <http://semen.revues.org/2911> ; consulté le 11/01/2015].

2 Groupe Mu (1990), *op. cit.*, p. 65.

3 François Rastier, « La macrosémantique », in *Texto!* [en ligne], juin 2002. [URL de référence : http://www.revue-texto.net/Inedits/Rastier/Rastier_Marosemantique1.html ; consulté le 10/02/2015].

L'attention interprétative oscille donc entre un premier mouvement, apavage de ce que Schaeffer nomme le « style cognitif divergent », et un second temps propre à un traitement synthétique cette fois-ci des indices pris dans le texte et rapportés à sa globalité, ainsi que le définit Rastier. Le style divergent, propre à l'exploration interprétative du texte, se caractérise, selon Schaeffer, par « l'importance des traitements descendants, la dé-hiérarchisation, le retard de catégorisation [qui] concourent à faire [du texte] un traitement non économique, aboutissant parfois à une surcharge attentionnelle¹ ». Il aboutit alors à une attention maximalisée portée à chaque strate informationnelle du texte et due à l'« abaissement du seuil de sélectivité attentionnelle² » et une lecture non économique du texte.

LE SUPPORT COMME DISPOSITIF ATTENTIONNEL

La caractéristique de l'interprétation littéraire est qu'elle repose jusqu'à présent sur des régimes attentionnels implicitement ou explicitement (selon le degré polémique du propos) qualifiés de « profonds ». Elle devient parfois dans la démonstration le contre-exemple de la distraction propre à l'« hyper-attention³ », décrite par Katherine Hayles.

Deep attention (ou attention « profonde ») et *hyper attention* (attention « dispersée » ?) correspondent, pour Hayles, à des sollicitations distinctes de l'environnement, du média, du support convoqué⁴. En effet, les pratiques attentionnelles précédemment décrites relèvent, pour l'essentiel, d'une confrontation avec le support imprimé. Poser la question de l'interprétation en environnement numérique, ainsi que

1 Jean-Marie Schaeffer (2011), *art. cité*, p. 145.

2 *Ibid.*, p. 147.

3 Katherine Hayles, « Hyper and Deep Attention : the Generational Divide in Cognitive modes », in *Profession*, 2007, p. 187-199 [URL : http://www.english.ufl.edu/da/hayles/hayles_hyper-deep.pdf]. « Hyper attention, by contrast, is characterized by switching focus rapidly between different tasks, preferring multiple information streams, seeking a high level of stimulation, and having a low tolerance for boredom. [...] Each cognitive mode has advantages and limitations. Deep attention is superb for solving complex problems represented in a single medium, but it comes at the price of environmental alertness and flexibility of response. Hyper attention excels at negotiating rapidly changing environments in which multiple foci compete for attention; its disadvantage is impatience with focusing for long periods on a non-interactive object such as a Victorian novel or complicated math problem. ». Nous renvoyons également à ce propos à : Nicholas Carr, *Internet rend-il bête ?*, Paris, Robert Laffont, 2011.

4 Voir également à ce sujet : Alexandra Saemmer, *Matières textuelles sur support numérique*, Saint-Étienne, Publications de l'Université de Saint-Étienne, 2007, p. 37 *sqq.*

nous aimerions le faire à présent, suppose donc la problématisation de la notion de support numérique.

Il ne s'agit pas alors de cliver l'analyse ou de poser une alternative dont le premier membre serait l'attention profonde, apanage du « cerveau littéraire » et le second, l'hyper-attention qui exclurait ainsi son objet du champ des humanités ou, plus généralement, des connaissances. Au contraire, l'usage d'outils d'exploration et de fouille automatisée des textes littéraires tend plutôt à confirmer l'importance de pratiques complémentaires de la lecture interprétative.

DU SUPPORT IMPRIMÉ AU SUPPORT NUMÉRIQUE Enjeux et conséquences pour l'attention interprétative ?

Notre propos n'est pas ici de rappeler les différences entre la lecture de supports imprimés et la lecture sur écran de supports hypertextuels. elles ont fait l'objet de nombreuses études précises et documentées, mais nous intéresserons davantage ici aux modalités de construction de l'interprétation dans un texte numérique enrichi, construit pour être justement fouillé et analysé automatiquement.

Rappelons, en préambule de cette réflexion sur le texte numérique et les régimes attentionnels sur lesquels il repose, la définition du support que propose Bruno Bachimont :

les propriétés matérielles du support d'inscription conditionnent l'intelligibilité de l'inscription. [Par conséquent,] [...] l'interprétation de l'inscription, ou le sens qu'on lui accorde, dépend de sa structure matérielle et de ses propriétés physiques. La matérialité du support pré-détermine, conditionne, le sens que l'on peut accorder à une inscription¹.

¹ Bruno Bachimont, *Arts et Sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire d'H. D. R., Université de Technologie de Compiègne, 12 janvier 2004, p. 77. [URL : http://www.utc.fr/~bachimon/Livresettheses_attachments/HabilitationBB.pdf, consulté le 02/02/2015] : « Entre déterminisme technique et neutralité, la théorie du support argue que les dispositifs et innovations techniques modifient les conditions de possibilité de la pensée et des échanges sociaux. Cette modification n'est pas une détermination : il n'y a pas de couplage nécessaire entre une innovation technique et une mutation cognitive,

Le traitement technique automatisé des textes et, plus généralement, des contenus culturels et intellectuels souligne de manière accrue l'importance de la pensée du support, *a fortiori* dans la pratique de l'attention interprétative.

LE SUPPORT NUMÉRIQUE :
UN MODE D'ABSTRACTION DU TEXTE ?

Le support numérique est généralement distingué du support imprimé par ses propriétés de dématérialisation et son fonctionnement hypertextuel. Pourtant, qui compose et cherche à analyser le texte digital est justement frappé par le paradoxe de cette supposée « dématérialisation ». Le texte numérique est bien, au contraire, matériel¹, *a fortiori* dans la perspective d'une recherche en linguistique ou en littérature par exemple. La construction d'un texte numérique enrichi à partir d'un texte qualifié de « brut² » renvoie à des opérations de discrimination de formes et de segmentation du corpus de travail en entités concrètes, de taille plus ou moins variable que sont, pour ne citer que des éléments généraux, les mots, les phrases, les énoncés, les textes, etc.

Le support numérique contribue en fait à externaliser à un autre niveau le langage ou, dans la perspective qui est la nôtre, le texte lui-même. La théorie du support déployée par Bachimont apparaît alors plus que bienvenue et invite à problématiser ce qu'il nomme la « réification³ » des contenus culturels. Ce processus est pour lui, ainsi que je le suggérais, une conséquence logique du processus de grammatisation⁴ « où l'expression de son rapport au monde se manifeste à travers des supports matériels qui permettent d'objectiver [l]a représentation du monde⁵ ».

culturelle ou sociale. La technique fonctionne comme moteur du changement, modifiant à la fois la réalité mais aussi les critères d'évaluation de la réalité. » (p. 74).

1 En particulier par la « grammatisation » de l'écriture et, de fait, le changement de support – l'« automatisation » – n'altère pas à mon sens cette propriété. Cf. Sylvain Auroux, *La Révolution technologique de la grammatisation*, Paris, Mardaga, 1993.

2 Les termes utilisés pour les décrire sont dans cette perspective très éclairants. Bernard Stiegler propose, quant à lui, de parler d'« hypermatérialité ». Voir : Bernard Stiegler, « L'attention, entre économie restreinte et individuation collective », in Citton, Y. [dir.], *L'Économie de l'attention. Nouvel horizon du capitalisme ?* Paris, La Découverte, 2014, p. 123-135.

3 Bruno Bachimont, « Le nominalisme et la culture : questions posées par les enjeux du numérique » in Stiegler, B. [dir.] (2014), *op. cit.*, p. 65.

4 Voir : Sylvain Auroux (1993), *op. cit.*, p. 109.

5 Bruno Bachimont (2014), art. cité, p. 65.

CONSTRUCTION DE CORPUS
ET PROJECTIONS INTERPRÉTATIVES

La construction d'un corpus numérique engage, comme nous le mentionnions plus haut, par la transposition du texte imprimé vers un nouveau support, de nouveaux traitements de ce texte, un nouveau mode d'interaction par la pratique des logiciels de textométrie ou de calculs statistiques par exemple¹. Par ailleurs, l'activité d'annotation de corpus apparaît comme un processus « informé » de traitement textuel, qui participe d'une anticipation de l'attention portée sur telle ou telle qualité ou propriété du texte².

Dans le cadre de projets tels que la conception d'outils d'exploration textuelle ou l'annotation de textes littéraires édités en ligne, le chercheur et le concepteur de l'édition sont donc confrontés à une tension entre deux tendances : celle d'informer ce qui servirait à la recherche immédiate et celle qui vise à construire une structure de recherches ultérieures et qui, par conséquent, peut parfois peiner à circonscrire ses objectifs.

La constitution d'une grille de métadonnées présuppose et prédétermine l'utilisation future du texte enrichi ; elle est effectivement le lieu de projections interprétatives fortes³. Le codage du corpus rend manifestes toutes les intentions de recherche et d'« exploitation » du texte,

-
- 1 Nous renvoyons à ce propos à la citation proposée en exergue, tirée de : François Rastier (2001), *op. cit.*, p. 74-75 : « Chaque nouveau support permet de nouveaux traitements, mais ne les détermine pas. Une nouvelle technologie définit précisément un nouveau mode d'interaction entre un traitement et un support ».
 - 2 Cf. Alan Liu, « Where is Cultural Criticism in the Digital Humanities ? », in *Debates in the Digital Humanities*, Gold, M. K. [éd.], Minneapolis, University of Minnesota Press, 2012, p. 490-509 : « In the digital humanities, cultural criticism – in both its interpretative and advocacy modes – has been noticeably absent by comparison with the mainstream humanities or, even more strikingly, with “new media studies” [...]. We digital humanists develop tools, data, metadata, and archives critically; and we have also developed critical positions on the nature of such resources (e.g., disputing whether computational methods are best used for truth-finding or, as Lisa Samuels and Jerome McGann put it, <http://www.jstor.org/stable/20057521>). »
 - 3 Pour Rastier, dans *Arts et sciences du texte*, l'un des aspects les plus intéressants de cette évolution est sans doute le changement de paradigme mis en œuvre par la linguistique de corpus : « Alors que l'objectif de la représentation des connaissances demandait de capter dans un formalisme le maximum d'“informations”, les nouveaux objectifs d'analyse imposent d'en sélectionner un minimum [...]. Or, ces représentations dépendent des tâches d'interprétation ; on passe alors d'une *problématique de la représentation* à la *problématique de l'interprétation*, et du programme cognitif de *simulation de la compréhension* à l'*aide à l'interprétation* » (François Rastier (2001), *op. cit.*, p. 76-77, nous soulignons). Ce changement

au sens le plus pratique du terme. L'ajout de balises et la construction de logiciels adaptés à des besoins précis de l'interprétation ou, dans une autre perspective, de l'édition critique d'un texte anticipe et oriente les pratiques interprétatives. Le lien entre l'interprétation, l'édition et le balisage du texte apparaît d'ailleurs de manière lumineuse dans le schéma suivant, proposé par Rastier :

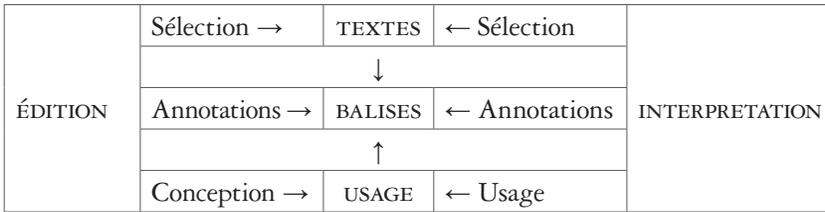


FIG. 1 – Schéma des rapports du balisage, de l'édition et de l'interprétation des textes (*in* : François Rastier (2001), *op. cit.*, p. 81).

Ainsi la « philologie numérique », selon son expression, opère un double mouvement réflexif sur le texte et produit des outils visant à faciliter et à rendre opératoires les critères de lecture des supports qu'ils se donnent comme objet¹.

Cette projection interprétative nuit-elle à l'interprétation elle-même ? Rend-elle la démarche plus problématique ? Nous avançons que l'enrichissement des bases et travail collaboratif sur lequel reposent, par exemple, les futures éditions savantes en ligne doit permettre de faire du texte numérique annoté un support à géométrie variable et susceptible de répondre finalement à des types de sollicitations et de stratégies attentionnelles divers.

LES TRANSFORMATIONS DU TEXTE INITIAL

Construire un corpus d'étude enrichi à des fins herméneutiques est, enfin, un processus fondamentalement ambivalent, notamment, parce qu'il fait intervenir deux objets distincts à interpréter : le texte, en tant

de paradigme engage à notre sens la modification des stratégies attentionnelles portées sur les corpus.

1 Voir à ce propos la description par Rastier des différentes étapes de la « mise en texte » numérique dans : François Rastier (2001), *op. cit.*, p. 82 *sqq.*

que substance entière et matière propre à susciter l'attention interprétative décrite plus haut, mais aussi et surtout les résultats de l'enquête menée sur le texte. Ce processus a déjà cours lorsqu'il s'agit d'un support imprimé, mais il a des implications plus importantes dans le cas présent, en particulier parce qu'il est explicité par les requêtes successives formulées dans le logiciel de traitement des données textuelles et parce que ces requêtes¹ produisent alors un autre état du texte, qui diffère souvent et dans sa présentation et dans son contenu réel du texte initial².

La « plasticité » du texte digital s'illustre par l'alternance souple et aisée d'un mode de visualisation du texte à un autre. Concordanciers, index, chiffres, nuages de mots, graphiques, classifications sont en effet autant d'états du texte initial, dont nous donnons ici quelques exemples. Chaque représentation présentée ci-dessous présente un état particulier des textes poétiques de Guillaume Apollinaire, tirés de l'édition savante en ligne « HyperApollinaire » mise en œuvre par le LABEX Obvil³. Cet état est le résultat d'une requête portant sur les substantifs du lexique et leur distribution relative dans l'œuvre poétique d'Apollinaire⁴.

-
- 1 Les logiciels de textométrie offrent de ce point de vue un bon exemple du type d'enquêtes que l'on peut mener sur un corpus, grâce à des requêtes par mot (chaîne de caractère), par lemme ou par catégorie grammaticale. Le traitement des résultats et leur classement par fréquence ou par spécificité illustrent également de manière remarquable les différentes visualisations du texte rendues possible par son formatage en données textuelles. Nous renvoyons pour la question de la manipulation du texte et de ses conditions d'inscription à : Yves Jeanneret, « Écriture et médias informatisés », in *Histoire de l'écriture. De l'idéogramme au multimédia*, Christin, A.-M. [dir.], 2012, p. 395-402.
 - 2 Ce que Rastier exprime ainsi : « Considérer un texte comme une chaîne de caractères serait le réduire à sa seule substance graphique, encourager son traitement séquentiel ou plus exactement déterministe (au moyen d'une fenêtre de lecture déplacée linéairement, comme c'est souvent le cas dans les analyseurs syntaxiques) ; enfin le couper de ses entours local (sa situation) et global (la culture dont il procède). Ce serait aussi réduire le langage au seul plan du signifiant ; encore ne s'agirait-il que des mots, à quoi correspondent dans le meilleur des cas les chaînes de caractères ». Cf. François Rastier (2002), art. cité.
 - 3 Nous renvoyons pour plus d'informations au corpus HyperApollinaire sur : <http://obvil.paris-sorbonne.fr/corpus/apollinaire/>. À propos du projet du LABEX Obvil, consulter la page : <http://obvil.paris-sorbonne.fr/projets/hyperapollinaire>.
 - 4 L'édition critique d'Apollinaire a été structurée et annotée en langage XML-TEI de manière à constituer une véritable base documentaire facilitant l'extraction de sous-ensembles (dans notre cas, le corpus poétique par opposition au corpus critique ou prosaïque apollinarien). Ce balisage rend ainsi possible la prise en compte des données textuelles et de laisser de côté pour cette analyse les notes de l'apparat critique (consultables dans l'édition en ligne).

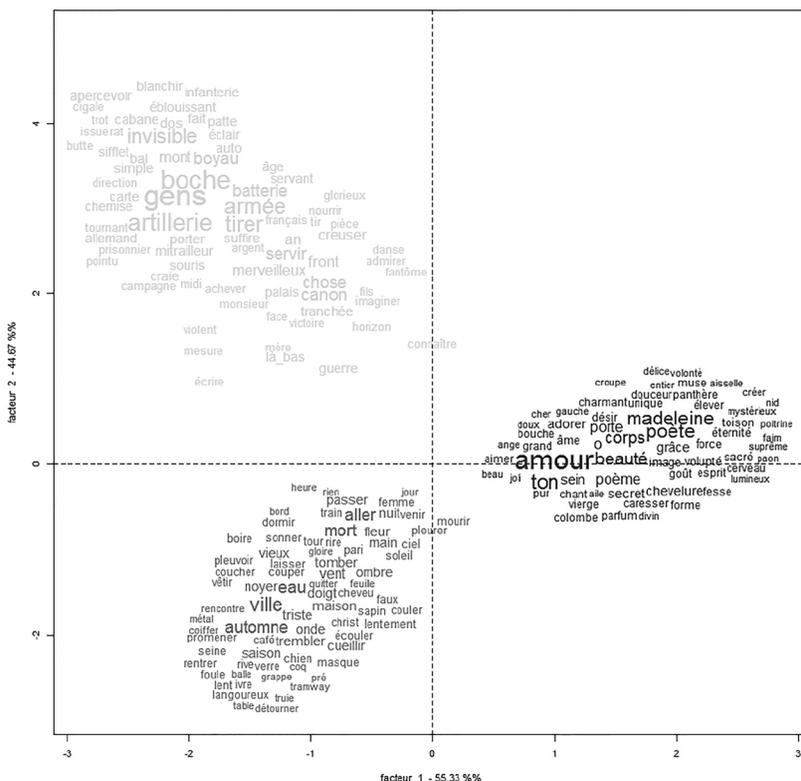


FIG. 4 – Visualisation en AFC des mots lexicaux des œuvres poétiques d’Apollinaire¹.

1 Sur une classification automatique du corpus, en fonction de critères purement statistiques, cette visualisation en AFC montre les mots lexicaux spécifiques des thématiques isolées par la machine, selon la méthode de segmentation et de classification Reinert. Pour plus d’informations à ce propos, nous renvoyons au site du logiciel *Iramuteq* qui décrit la Méthode Reinert : http://www.iramuteq.org/documentation/html/copy8_of_2-5-1-3-options-supplementaires-de-lanalyse-statistique-textuelle-sur-chaque-forme.

Ces images donnent à *voir* des données textuelles, interrogent les voisinages de termes (comme le font les nuages de mots ou les AFC par exemple) et matérialisent des différences et des coïncidences. L'analyse des similitudes en donne un bon exemple en mettant en évidence, par l'intermédiaire de « branches », des liens entre les termes du lexique. Les dislocations du texte s'offrent comme image dans son immédiateté et posent néanmoins immédiatement la question de son interprétation. Face au calcul statistique ou à ces types de représentation du texte initial, le chercheur en littérature est parfois dérouté, voire découragé par la tâche même à entreprendre : interpréter ce qu'il voit. Ces sentiments, Bachimont en définit ainsi l'origine :

En permettant une manipulation automatique du contenu, le numérique fait sortir le contenu culturel hors du contrôle interprétatif où était soumis jusqu'à présent tout contenu grammatisé¹.

Les interfaces et les modes de visualisation des résultats traités par la machine diffèrent l'interprétation directe et imposent un mode opératoire distinct de l'interprétation sur support papier : il s'agit d'abord de comprendre les résultats, d'en analyser la validité, le degré de pertinence et de signification. Par les actions de collation, de comparaison, de transformation, le contenu du texte change d'état et se transforme en fait à analyser.

Les diverses visualisations du texte permises par ces logiciels mettent en lumière le rôle des données textuelles, des entités formatées pour le support numérique. Elles ne donnent finalement *que* les formes identifiées par des balises. L'opération interprétative tend, quant à elle, à leur restituer le rôle de « trace » à lire, comprendre et resituer dans le texte.

CONCLUSION

Le passage du texte à un corpus numérique annoté (traité, par exemple, traité par des logiciels de textométrie) nous semble bien décrit l'expression forgée par Jeffrey Schnapp, « Knowledge design », puisqu'il rend explicite les modifications opérées sur notre manière d'appréhender le texte par le passage à un environnement numérique.

¹ Bruno Bachimont (2014), art. cité, p. 66.

Les outils informatiques du traitement de corpus scindent en effet le parcours interprétatif du texte, balisé par les métadonnées du corpus. Ainsi, celles-ci orientent bien la conduite de la quête interprétative menée à travers le texte et articulent les deux faces des objets numériques que Lev Manovich décrit comme le processus de « transcodification¹ ». La démarche d'enquête linguistique sur le texte digital nous semble donc évoluer d'une « philologie numérique » à une philologie *du* numérique, où deux niveaux d'analyse sont constamment sollicités.

Par les interfaces de représentation des résultats de l'enquête, « le rapport linguistique au monde de la culture, la langue et l'expression du contenu [...] est remplacé par un rapport perceptif et qualitatif² », écrit encore Bachimont, de sorte que l'attention sollicitée par la visualisation des résultats et l'attention interprétative ont parfois du mal à s'articuler. Cette difficulté reste à notre sens une des pistes à explorer et un défi à relever par les éditeurs de texte en ligne.

En ce qui concerne l'apport interprétatif de ces outils d'exploration et de calcul du texte, dont l'abord n'est *a priori* ni forcément aisé, ni particulièrement évident, ils nous apparaissent, à l'usage, non seulement ouvrir des pistes de lecture littéraire et stylistique, mais aussi – en qualité de *pharmaka*, tantôt alliés, tantôt poisons – alerter le chercheur sur les enjeux mêmes de sa recherche et le contraindre ainsi à définir et à expliciter plus clairement que jamais les postulats et les objectifs de ses explorations du texte.

La quantification d'un fait de langues, par exemple, peut posséder une véritable fonction heuristique, qui postule la capacité d'un corpus annoté à faire émerger et saillir des éléments observables et remarquables. Le travail de l'interprète réside dans l'articulation du quantitatif et du qualitatif pour construire et expliciter la signification de son enquête. Le calcul statistique, notamment, permet de mettre au jour des « coalitions d'indices », propices à nourrir l'herméneutique.

L'outil spécifique que proposent les statistiques textuelles impose de ne plus suivre (ou en tout cas, plus seulement) la démarche traditionnelle du chercheur en sciences humaines. La technique change effectivement les pratiques mêmes de l'enquête. Le second enjeu heuristique

1 Lev Manovich, *The Language of New Media*. Cambridge, Massachusetts, London, England, The MIT Press, 2001.

2 Bruno Bachimont (2014), art. cité, p. 70-71.

de l'interprétation assistée par l'ordinateur se trouve résumé ainsi par Mayaffre :

Le risque bien connu est d'abord de toujours trouver ce que l'on cherche [...].
Le risque ensuite est d'ignorer des faits essentiels au corpus, qui passeront inaperçus faute d'avoir été pressentis comme dignes d'intérêt¹.

Le retour au contexte est primordial, de même que l'interrogation constante sur les conditions d'émergence du résultat observé. L'usage de ces outils pour nourrir l'interprétation forme également le chercheur à une perpétuelle vigilance face au rapport entre la formulation de ses requêtes et les résultats obtenus, ainsi que les conditions de leur intelligibilité. La pratique de l'interprétation ainsi définie apparaît donc comme un dispositif d'enquête et de construction du savoir qui suscite un état d'alerte, bien différent des régimes attentionnels de l'interprétation propre au seul support imprimé.

Clémence JACQUOT
Université Paris-Sorbonne
Université Pierre-et-Marie-Curie

1 Damon Mayaffre « L'analyse du discours assistée par ordinateur [Séminaire de formation, Alexandrie, 1-11 décembre 2009] », *Ressource d'Enseignement*, n. p. [en ligne], 2009, p. 13 [URL de référence : eprints.aidenligne-francais-universite.auf.org/19/; consulté le 17/06/2012].

BIBLIOGRAPHIE

- ADAM, J.-M. et VIPREY, J.-M. (2009), « Corpus de textes, textes en corpus. Problématique et présentation », in *Corpus*, n° 8 [URL : <http://corpus.revues.org/index1672.html>].
- AUROUX, S. (1993), *La Révolution technologique de la grammatisation*, Paris, Mardaga.
- BACHIMONT, B. (2004), *Arts et Sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire d'HDR, U. T. C., [URL : http://www.utc.fr/~bachimon/Livresettheses_attachments/HabilitationBB.pdf].
- BACHIMONT, B. (2014) « Le nominalisme et la culture : questions posées par les enjeux du numérique » in Stiegler, B. [dir.], *Digital Studies. Organologie des savoirs et technologies de la connaissance*, Paris, FYP Éditions, p. 63-78.
- BERRY, D. M. [dir.] (2012), *Understanding digital humanities*, Palgrave, Macmillan.
- CITTON, Y. (2010), *L'Avenir des Humanités. Économie de la connaissance ou cultures de l'interprétation ?*, Paris, La Découverte.
- CITTON, Y. (2014), *Pour une écologie de l'attention*, Paris, Seuil.
- CORMERAIS, F. (2014), « Humanités digitales et (ré)organisation du savoir », in : Le Deuff, O. [dir.], *Le Temps des humanités digitales. La mutation des sciences humaines et sociales*, Paris, FYP Éditions, p. 129-142.
- CROZAT, S., BACHIMONT, B., CAILLEAU, I., BOUCHARDON, S., GAILLARD, L. (2011), « Éléments pour une théorie opérationnelle de l'écriture numérique », in *Supports et pratiques d'écriture en réseau*, Bouhai, N. [dir.], URL de référence : <http://www.cairn.info/revue-document-numerique-2011-3.htm> [consulté le 14/05/2015].
- DELEUZE, G. (1982), Cours n° 19 « Cinéma / Image-Mouvement », *La Voix de Gilles Deleuze*, site de l'Université Paris-8 [URL : http://www2.univ-paris8.fr/deleuze/article.php?id_article=157].
- DOUEIHI, M. (2014), « Quête et enquête », in Le Deuff, O. [dir.], *Le Temps des humanités digitales. La mutation des sciences humaines et sociales*, Paris, FYP Éditions, p. 7-10.
- GHITALLA, F., BOULLIER, D., GKOUSKOU-GIANNAKOU, P., LE DOUARIN, L. et NEAU, A. (2003), *L'Outre-lecture. Manipuler, [s']appropriier, interpréter le Web*, Paris, Bibliothèque Centre Pompidou.
- GROUPE MU (1990), *Rhétorique de la poésie. Lecture tabulaire, lecture linéaire*, Paris, Seuil.
- GUIRAUD, P. (1960), *Problèmes et méthodes de la statistique linguistique*, Paris, PUF.
- HAYLES, K. (2007), « Hyper and Deep Attention : the Generational Divide

- in Cognitive modes », *Profession*, p. 187-199 [URL : http://www.english.ufl.edu/da/hayles/hayles_hyper-deep.pdf].
- HOLMES, D. I. (1998), « The Evolution of Stylemetry in Humanities Scholarship », *Literary and Linguistic Computing*, volume 13/3, p. 111-117.
- LIU, A. (2012), « Where is Cultural Criticism in the Digital Humanities ? », in *Debates in the Digital Humanities*, Gold, M. K. [éd.], Minneapolis, University of Minnesota Press, 2012, p. 490-509
- MANOVICH, L. (2001), *The Language of New Media*. Cambridge, Massachusetts, London, England, The MIT Press.
- MAYAFFRE, D. (2002), « Les corpus réflexifs : entre architextualité et hypertextualité », in *Corpus*, n° 1, [URL : <http://corpus.revues.org/index11.html>].
- MAYAFFRE, D. (2009), « L'analyse du discours assistée par ordinateur [Séminaire de formation, Alexandrie, 1-11 décembre 2009] », Ressource d'Enseignement, [URL : <http://eprints.aidenligne-francais-universite.auf.org/19/>].
- MORETTI, F. (2005), *Graphs, Maps, Trees : Abstract Models for Literary History*, Londres, Verso.
- MULLER, C. (1993), *Principes et méthodes de statistique lexicale*, Paris, Champion.
- PINCEMIN, B. (2011), « Sémantique interprétative et textométrie – Version abrégée », in *Corpus* n° 10 [URL : <http://corpus.revues.org/2121>]
- RASTIER, F. (2001), *Arts et sciences du texte*, Paris, PUF.
- RASTIER, F. (2002), « La macrosémantique », in *Texto!* [URL : http://www.revue-texto.net/Inedits/Rastier/Rastier_Marcosemantique1.html]
- RASTIER, F. (2004), « Enjeux épistémologiques de la linguistique de corpus », *Texto!* [URL de référence : http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html].
- RASTIER, F. (2011), *La Mesure et le grain. Sémantique de corpus*, Paris, Champion.
- SCHAEFFER, J.-M. (2011), « Styles attentionnels et relation esthétique », in Jenny, L. [dir.], *Le Style en acte. Vers une pragmatique du style*, Genève, Métis Presses, p. 138-149.
- SINCLAIR, J. (1996), « Preliminary recommendations on Corpus Typology », *Technical Report, EAGLES* (Expert Advisory Group on Language Engineering Standards) [URL : <http://www.ilc.cnr.it/EAGLES/pub/eagles/corporal/corputyp.ps.gz>].
- STIEGLER, B. (s.d.), article « Pharmakon », *Ars Industrialis* [URL : <http://arsindustrialis.org/pharmakon>].
- STIEGLER, B. (2004), *De la misère symbolique, 1. L'Époque hyperindustrielle*, Paris, Gallilée.
- STIEGLER, B. (2012), *Prendre soin de la jeunesse et des générations*, Paris, Flammarion.
- STIEGLER, B. (2014), « Pharmacologie de l'épistèmè numérique », in Stiegler, B.

[dir.], *Digital Studies. Organologie des savoirs et technologies de la connaissance*, Paris, FYP Éditions, p. 13-26.

STIEGLER, B. (2014), « L'attention, entre économie restreinte et individuation collective », in : Citton, Y. [dir.] (2014), *L'Économie de l'attention. Nouvel horizon du capitalisme ?*, Paris, La Découverte, p. 123-135.

WEIL, S. (1966), *Attente de Dieu*, Paris, Fayard.

LABEX Obvil, projet *HyperApollinaire* : <http://obvil.paris-sorbonne.fr/projets/hyperapollinaire>.

Logiciel *Iramuteq* : <http://www.iramuteq.org/>.