

CORPUS

Corpus

8 | 2009

Corpus de textes, textes en corpus

Corpus de textes, textes en corpus. Problématique et présentation.

Jean-Michel Adam et Jean-Marie Viprey



Édition électronique

URL : <http://journals.openedition.org/corpus/1672>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 15 novembre 2009

Pagination : 5-25

ISSN : 1638-9808

Référence électronique

Jean-Michel Adam et Jean-Marie Viprey, « Corpus de textes, textes en corpus. Problématique et présentation. », *Corpus* [En ligne], 8 | 2009, mis en ligne le 22 janvier 2012, consulté le 19 avril 2019.

URL : <http://journals.openedition.org/corpus/1672>

© Tous droits réservés

Corpus de textes, textes en corpus. Problématique et présentation.

Jean-Michel ADAM
Université de Lausanne

Jean-Marie VIPREY
Université de Franche-Comté

En prenant l'initiative de ce volume de *Corpus*, nous avons choisi de procéder à un appel assez large, afin de faire le point sur des tendances actuelles des travaux sur corpus ancrés dans la textualité des faits de discours. En nous associant, nous pensions attirer des chercheurs d'horizons différents et notre attente n'a pas été déçue. Les propositions furent très nombreuses et de sélection en sélection, en regrettant quelques contributions abandonnées pour des raisons de délais de rédaction, nous avons finalement retenu neuf contributions qui nous semblent représentatives.

Autour de l'analyse de discours et des disciplines et sciences des textes, nous avons entrepris, depuis plus de dix ans, une réflexion épistémologique qui croise l'analyse textuelle et comparée des discours développée à Lausanne et l'analyse informatique des données textuelles du pôle *Archives, Bases, Corpus* de la Maison des Sciences de l'Homme de Franche-Comté. Nous avons en commun une réflexion qui porte sur les concepts de *texte*, de *discours*, d'*archive* et de *corpus* et, à la question qui est au centre de nos travaux de linguistique textuelle : « *Qu'est-ce qu'un texte ?* », le pôle 3 de la M.S.H. ajoute : « *Qu'est-ce qu'un texte à l'ère numérique ?* ». Nous avons également en commun la question que posait Jean-Marie Viprey dans le numéro 21 de *Semen* : « *Quelle place pour les sciences des textes dans l'analyse de discours ?* » (2006a : 167-182).

Si les conditions d'un dialogue de l'analyse textuelle des discours et de l'analyse des données textuelles sont réunies,

Corpus n°8 « Corpus de textes, textes en corpus »
(2009), 5-25

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 textes.

Tandis que nous, linguistes, mettons l'accent sur la définition des unités élémentaires, sur le traitement de la linéarité des textes, sur les enchaînements transphrastiques et sur la combinatoire d'unités de rangs de complexité supérieure à la phrase, les recherches informatisées insistent sur la structure non-séquentielle et réticulaire des textes. C'est ce que font deux articles de *Langages* n°163 (2006), consacré aux « Unité(s) du texte » et coordonné par Dominique Legallois du CRISCO de Caen. Dans ce numéro, Jean-Marie Viprey aborde l'organisation non-séquentielle des textes en portant son attention sur une composante lexicale de la cohésion textuelle indéniablement négligée par les linguistiques du texte, même quand elles intègrent le concept encore trop rudimentaire d'*isotopie*. Il montre que la cooccurrence ou la collocation est un facteur primordial de textualité : « La collocation modifie et constitue les vocables en tant que tels, non tellement comme individus que comme constituants d'un tout engrené, influant ainsi sur la cohésion par l'intermédiaire du palier lexical intégralement considéré (et non disséminé phrase par phrase) » (2006b : 73). Dominique Legallois montre que la répétition de trois lexèmes ou plus dans certaines phrases d'un texte est responsable d'une organisation réticulaire « en parfaite congruence avec l'étymologie du mot *texte*. C'est bien un *tissu* de phrases enchevêtrées, une *trame*, une *texture*, toute une constitution et une conception d'un objet complexe que met en évidence l'analyse de la répétition lexicale dans les discours » (2006 : 70). Ces travaux ne remettent pas en cause le caractère structurant de la linéarité textuelle, mais ils prouvent que la textualité doit résolument être pensée comme la combinaison de

parcours linéaires et réticulaires. Cette problématique est au cœur de l'article de MARIE-LAURE FLOREA dont nous reparlerons plus loin.

L'objectif interdisciplinaire que nous souhaitons voir représenté et problématisé l'a été dans quelques articles, en particulier par NATHALIE GARRIC et VALÉRIE CAPDEVIELLE-MOUGNIBAS qui montrent parfaitement comment l'analyse de discours peut guider un « cheminement vers un lieu de rencontre interdisciplinaire dans les Sciences Humaines et Sociales ». À partir de leur expérience de linguiste et de psychologue, elles discutent « l'intérêt de l'utilisation interdisciplinaire de la pratique de corpus et de l'analyse textuelle des discours ». Le fait que des littéraires, philologues ou généticiens, des linguistes, des psychologues et des informaticiens, des spécialistes des discours médiatiques ou des discours sociaux aient contribué à ce volume, souvent en co-écrivant un article, est la preuve du rôle fédérateur des questions que nous avons mis en débat.

Trois questions de notre appel ont reçu des réponses partielles :

- Dans quelle mesure le texte est-il amené à incorporer des éléments jusqu'ici désignés comme plus ou moins « extérieurs » à lui (péritexte, épitexte, intertexte, hypertexte, etc.), et comment les chercheurs gèrent-ils ce qu'il est convenu d'appeler en conséquence l'« ouvert » du texte ?
- Dans ce cadre, quelles solutions théoriques et pratiques se présentent pour gérer *la variation textuelle* ?
- Que faire de la « philologie numérique », peut-on et doit-on normaliser les procédures d'établissement des ressources textuelles, pourquoi et comment ?

Les travaux qu'on va lire se réfèrent à plusieurs de ces concepts que Jean-Marie Viprey qualifie fort justement de « régulateurs » ou « transactionnels [...] indispensables au jeu des disciplines humanistes entre elles, même si certains ont [...] une ou plusieurs acceptions dures dans ces mêmes disciplines » (2006a : 168). Parmi les concepts mis en circulation, nous croiserons ceux de *texte*, de *genres de discours*, d'*intertextualité*, de *co(n)texte*, d'*unités d'analyse* et de *construction des corpus*.

1. Texte et variations

Le concept de texte, attaché au participe passé du verbe *texere* (*textus* ce qui est tissé, tressé), possède une connotation de fixité et de fermeture structurale qui ne correspond ni à la réalité de l'écriture médiévale, ni à celle des œuvres de la Renaissance comme l'ont démontré magistralement aussi bien Bernard Cerquiglini que Michel Jeanneret.

Dans un chapitre de son *Essai de poétique médiévale* (1972), Paul Zumthor mettait en cause les concepts d'*auteur* et de *texte unique* auxquels il proposait de préférer ceux d'*anonymat* et de *mouvance*. Bernard Cerquiglini a prolongé cette analyse en soutenant que le texte médiéval n'existe que sous forme de variations manuscrites de copistes, qui donnent parfois lieu à des continuations. L'écriture médiévale « raboute, tisse à nouveau et perpétuellement des œuvres, *œuvre* sans cesse » (Cerquiglini 1989 : 59).

L'écriture de la Renaissance se caractérise par un recyclage intertextuel incessant et elle a tendance à contester et à « ruser avec les effets stabilisateurs de l'imprimerie et [à] préserver, dans une large mesure, la mobilité de l'ère du manuscrit » (Jeanneret 1994 : 25). Michel Jeanneret a montré que de Montaigne et Ronsard à Erasme, certes « la mobilité existe au sein des versions imprimées ; si les premières étapes de la genèse sont perdues, celle-ci continue au fur et à mesure des éditions. Loin d'immobiliser le texte, l'impression le propulse d'un état à l'autre, elle épouse l'histoire de ses métamorphoses » (1994 : 25). Ce que résume cette réponse d'Erasme à un correspondant qui se plaignait de devoir acheter des éditions en variation perpétuelle et se demandait pourquoi publier un ouvrage avant qu'il ne soit réellement achevé : « De même que, tant que nous vivons, nous ne cessons de travailler à nous rendre meilleurs, de même nous ne cessons de corriger et de compléter les œuvres de notre esprit qu'en cessant aussi de vivre » (Lettre d'Erasme à John Botzheim, 1552 ; citée par Jeanneret 1994 : 35).

Au XVII^e siècle, à propos des éditions 1623, 1626 et 1633 du *Francion* de Charles Sorel et des trois éditions successives de l'*Adonis* de La Fontaine (manuscrit d'apparat

offert à Fouquet en 1658 et éditions de 1669 et 1671), Patrick Dandrey aboutit à cette conclusion : chacune des éditions antérieures peut être rétrospectivement considérée comme une sorte de « brouillon » avant-textuel de la suivante. À propos de ce même « grand siècle », l'article que nous avons demandé à DELPHINE DENIS est exemplaire. Tout en retraçant l'histoire éditoriale particulièrement complexe de *L'Astrée* d'Honoré d'Urfé, cet article est l'occasion d'une mise au point importante qui, sans « récuser l'approche philologique au profit de lectures "hypertextuelles" », parvient magistralement à repenser la philologie moderne dans un contexte épistémologique qui invite aux « décloisonnements fructueux » que nous appelions de nos vœux en prenant l'initiative de ce volume.

Les variations éditoriales portent sur des changements socio-historiques des conditions de circulation et de lecture des textes. Leur analyse nous permet de comprendre qu'un texte n'est pas seulement un objet spatial ayant la stabilité de ses trois dimensions (gauche-droite de la ligne, haut-bas de la page, et épaisseur du volume d'édition), mais qu'il possède une quatrième dimension qui fait de lui un objet temporellement changeant dans le temps de son écriture (du vivant de l'auteur) et un objet qui parcourt éditorialement le temps et l'espace socioculturel avant de parvenir à ses lecteurs. La stabilité et l'abstraction de l'objet-texte sont mises à mal dans le processus médiateur de l'édition, comme Jacques Neefs le synthétise fort clairement :

Quand les œuvres, et leur texte, sont livrés à l'imprimé, surtout quand cela est fait avec « privilège », autorisation, « imprimatur », bon à tirer, il semble qu'elles accèdent à une existence stable. Pourtant, les éditions vont varier, souvent considérablement, dans le temps – et pas seulement par modifications voulues par les auteurs ou par variation de la qualité typographique. [...] Si l'appartenance des œuvres à un moment singulier (la manière dont elles se profilent dans une culture faite de tout le passé qu'elle se donnent) est le mode même de leur surgissement, leur appartenance tout aussi forte aux temps indéfinis et illimités que ces

œuvres traversent (la manière dont elles sont activées, reformées par les lectures, agissantes dans les nouvelles littératures qui sont leur avenir) est le mode de leur existence en tant qu'œuvres d'art ou de pensée, c'est-à-dire irréductibles à ce qu'elles ont pu dire un temps, à ce qu'elles peuvent sembler dire dans une époque. (1990 : 7-8).

Dans une perspective génétique, Almuth Grésillon (2008 : 302-303) reprend deux remarques de Michel Jeanneret : « [...] l'œuvre ne coïncide pas avec un texte mais plusieurs, elle tient à ses variantes autant qu'à ses constantes, elle se donne à saisir comme un objet muable, dont l'histoire fait partie intégrante. [...] Une œuvre ne se réduit pas à l'une de ses manifestations ; elle réside aussi dans la somme – et la différence – de ses divers états ; la production signifie autant que le produit fini » (1994 : 27). Ce changement de paradigme s'applique aussi bien à l'œuvre poétique de Francis Ponge qu'aux contes de Perrault et à *L'Astrée*, aux variations journalistiques des énoncés cités et des articles recyclés dans divers titres qu'aux états médiatiques des discours politiques (enregistrements oraux, publications écrites, présence sur sites de partis).

En conséquence, comme le disent aussi bien Grésillon que Cerquiglini : « L'analyse doit être comparative, et non pas archéologique. Et ceci d'autant plus que l'archéologie réduit à l'unique ce qui prend sens de sa différence » (Cerquiglini 1989 : 68)¹. De nombreux articles du présent volume s'interrogent sur la constitution de corpora contrastifs. GEORGETA CISLARU et FRÉDÉRIQUE SITRI soulignent très clairement que « les spécificités d'un genre discursif ne sont réellement accessibles que par la comparaison des genres ». Les procédures de contextualisation décrites par VINCENT CAPT, JÉRÔME JACQUIN et RAPHAËL MICHELI visent à contraster des ensembles de texte au sein d'un corpus défini.

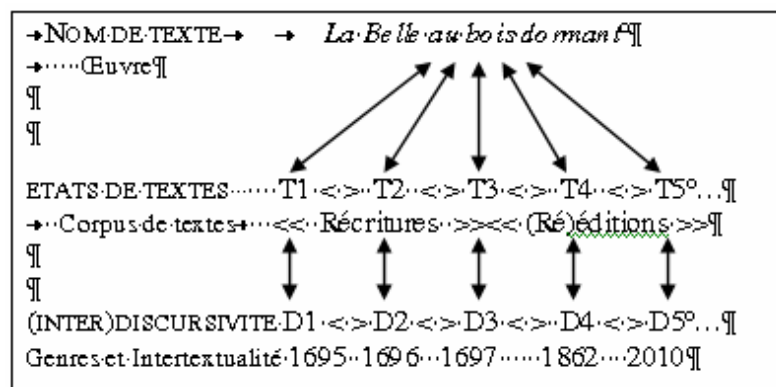
¹ Sur la méthodologie de la démarche comparative, voir Ute Heidmann 2005a & 2005b.

Corpus de textes, textes en corpus

Par ailleurs, admettre que l'écriture ne produit pas des variantes, mais qu'elle *est variance* remet en cause le « rôle d'appendice » (Grésillon 1994 : 17) que la critique littéraire fait jouer aux « variantes » rejetées dans les appareils de notes. Une attention comparative aux états stabilisés de textes entiers ou de segments textuels identifiables comme unités textuelles est tout le contraire de la réduction des différents états textuels à l'unicité d'un texte possédant des variantes considérées comme autant de segments isolés. Cela vaut certes pour les textes littéraires mais tout autant pour les discours politiques. À la différence de la philologie classique, l'objectif du chercheur n'est pas d'établir un état stable et supposé vrai d'un texte, avec pour visée ultime son exégèse. Il s'agit bien de rompre autant avec le positivisme de la philologie classique qu'avec l'idéalisme subjectiviste de l'herméneutique.

Pour prendre un exemple plus simple que *L'Astrée*, et complémentaire, les réécritures auctoriales et les éditions successives des contes de Perrault aboutissent à des états textuels auxquels nous avons matériellement accès. En d'autres termes, un texte dénommé « *La Belle au bois dormant* » ou « *L'Astrée* » a connu des variations auctoriales et éditoriales liées à un certain état de la circulation des textes dans des formations discursives aux caractéristiques socio-historiques changeantes :

Schéma 1



Au concept de texte conçu comme unité close sur elle-même et achevée, il s'agit de substituer un concept de *texte en variation*, soumis à des mises en texte dont résultent des états successifs, identifiables et regroupables en corpus. Pour *La Belle au bois dormant*, paraître dans le *Mercure galant* de février 1696 (T2), c'est apparaître dans un co(n)texte socio-discursif (D2) très différent de *La Belle au bois dormant*-T1, premier conte du manuscrit d'apparat des *Contes de la mère Loye* de 1695, envoyé à la nièce du roi de France ; contexte (D1) lui-même très différent de celui de *La Belle au bois dormant*-T3, premier conte des *Histoires ou Contes du temps passé*, paru chez Barbin en 1697 (D3). Dans cette livraison du *Mercure galant*, *La Belle au bois dormant* est publié en lieu et place d'un récit que les lectrices et lecteurs avaient l'habitude de trouver sous le titre « Histoire ». Maurice Lever définit ainsi le genre de ces « Histoires » auxquelles se substitue le conte : « Chacune d'elles trouve son point de départ dans un événement mondain, souvent mineur, que le narrateur développe ensuite au gré de sa fantaisie. [...] Sous leur apparente frivolité, ces historiettes visent souvent à illustrer une petite leçon de morale. [...] Le code de la galanterie est mis en apologues, chacun d'eux proposant un petit problème sentimental, avec des exemples à suivre ou à éviter, des recettes, des conseils. Cela tient à la fois du courrier du cœur et du divertissement de salon » (1996 : 186-187). Cette détermination co(n)textuelle explique certains choix d'écriture qui distinguent fortement T2 par rapport à T1 et T3 (analyse détaillée dans Adam 2009).

L'attention aux récritures auctoriales successives est ici même illustré par l'article de THOMAS LEBARBÉ à propos des manuscrits de Stendhal. Cette étude porte sur des avant-textes qui ne présentent pas encore une nature textuelle établie, c'est-à-dire auxquels le critère de *jugement de complétude*, dont parlent VINCENT CAPT, JERÔME JACQUIN et RAPHAËL MICHELI, s'applique difficilement. Le schéma 1 ne prend pas en compte des états avant-textuels, mais bien des textes satisfaisant au jugement de complétude pour des raisons péritextuelles et de matérialisation manuscrite allographe ou éditoriale. Il a pour but de mettre en avant, non pas l'existence de variantes d'un texte

unique, mais de variations textuelles d'une œuvre donnée. Le passage d'un état de texte T1 à des états T2 et T3 ne transforme pas T2 ou T3 en variantes de T1, mais en d'*autres* textes. T1 ou T2 ne sont pas inférieurs à T3 dans une vision téléologique méliorative aboutissant à l'œuvre parfaite. T4 et T5 ne sont pas inférieurs à T3 dans un sens de dégradation du texte publié initial. Ces états textuels permettent de traiter le sens comme un phénomène différentiel :

C'est « là où ça varie », là « où ça se différencie » que se construit le sens (variable et variant) à inter-échanger ; [...] ce n'est pas une analyse du « cohésif/cohérent » qui permet de fonder une sémantique. Mais peut-être, plutôt, la prise en compte du « différentiel » en variance, où s'enracinerait une « sémiotique de l'altération », une « sémiotique différentielle ».
(Peytard 1993 : 147)

Définissant sa « sémiotique de l'altération » comme une attention au fait de dire autrement, Jean Peytard ajoutait :

Il n'est de version définitive d'un texte que celle sur laquelle on s'arrête, comme *ne varietur*, pour en montrer la constante stabilité acquise certes après quelques variations. N'est-ce pas dans le cours de cette mouvance, de ces variations, qu'il convient de se situer pour apercevoir le sens, qui n'est jamais, semble-t-il, une fois pour toutes donné, mais constamment effet et source d'altérations ? (1993 : 176)

C'est cet intérêt commun pour le fonctionnement variationnel du texte qui, comme le dit DELPHINE DENIS, contribue au rapprochement entre linguistique textuelle ainsi redéfinie, analyse de discours et philologie, « à rebours d'une conception étroite de cette dernière comme entreprise de réduction *des* textes empiriques à *un* texte idéal, que les aléas de leur transmission auraient opacifié ».

2. Discours, contextualisations et généricité

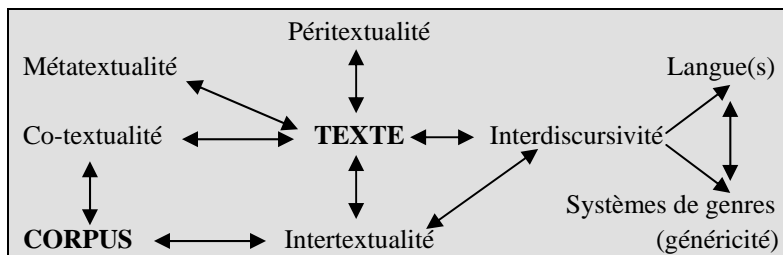
Le schéma 1 signale qu'en D1 ou D3, accompagnant les textualisations T1-1695 ou T3-1697, dans la mémoire des

lecteurs du XVII^e siècle, circulent, entre autres, des (inter)textes latins, des contes de La Fontaine et de Fénelon, des nouvelles de Basile, etc. Ces (inter)textes, qui ne font plus partie de notre mémoire interdiscursive, étaient déjà ignorés par ceux du XIX^e siècle (D4) et ceux du XX^e siècle (D5). En revanche, les contes des Grimm faisaient partie de D4 en 1862 et leurs versions Walt Disney de D5 en 2010. Si tout texte écrit se prête à des lectures renouvelées, superficielles ou profondes, de la part de lecteurs qu'auteur et éditeur n'ont pu que conjecturer, c'est parce que les *textualisations* successives sont soumises à des *re(con)textualisations* infinies (dans le sens de CAPT, JACQUIN et MICHELI).

Pour saisir la socialité de ces *re(con)textualisations* infinies, NATHALIE GARRIC et VALÉRIE CAPDEVIELLE-MOUGNIBAS, renvoient au Michel Foucault de *L'archéologie du savoir* et elles ancrent clairement leur propos dans le champ de l'analyse de discours. Illustrant ce que nous venons de dire, elles prennent prioritairement en compte les « régularités qui délimitent des familles d'énoncés caractéristiques d'une pratique discursive », ces régularités propres à des formations discursives fonctionnent comme des normes qui règlent les discours, normes qui s'incarnent dans des genres.

On comprend ainsi que la *textualité* est plus complexe que le *texte*, au sens étroit du terme. Le schéma 2 rend compte de cette complexité et de cette ouverture :

Schéma 2



Pour nous, la textualité est un système ouvert de relations entre des faits de co-textualité, de péri-textualité et de métatextualité inséparables de faits d'intertextualité et de généricité. La textualité est définissable comme la relation d'un

texte aux énoncés *péritextuels* qui en délimitent les frontières (DELPHINE DENIS examine en ce sens l'adjonction, en 1610, d'une « Épître au Roi » absente de l'édition originale de *L'Astrée*). La composition d'un recueil de poèmes ou de nouvelles, comme d'une hyperstructure journalistique affecte la *co-textualité*. Nous entendons par là l'ensemble des relations qu'un texte entretient avec les textes co-présents dans un même volume ou dans la même aire scripturale d'une page ou d'une double page de quotidien ou de magazine. Le *système de genres* et la *langue* d'une époque donnée sont les deux composantes de base de l'interdiscursivité. Ni le système de genre ni la langue ne sont stables et ce sont leurs changements qui font varier les états de textes successifs et conditionnent leur réception.

Ici encore l'exemple de *L'Astrée* est utile et on comprend que, dès 1712, l'avertissement de *La Nouvelle Astrée* puisse proclamer que « Cent ans dans une Langue vivante, mettent tout hors de mode ». Mais à ces changements de la « langue vivante », s'ajoute l'ensemble complexe des savoirs encyclopédiques du lectorat qui, comme le montre DELPHINE DENIS, conditionnent les réécritures éditoriales. Nous définissons l'interdiscursivité comme un ensemble de formes socio-discursives et historiques disponibles tant à la production qu'à l'interprétation dans une communauté d'auteurs, d'éditeurs et de lecteurs. Nous distinguons l'*interdiscursivité* comme système de genres et répertoire de formes langagières partagés et l'*intertextualité* comme ensemble de (fragments de) textes disponibles dans les mémoires d'une communauté discursive et d'individus donnés. L'intertextualité est un partage de textes ou de segments de textes précis, identifiables et citables, alors que l'interdiscours est un partage de formes et de normes langagières, constituées d'énoncés et même de textes emblématiques de ces normes et formes. L'intertextualité est, en ce sens, un sous-ensemble de l'interdiscursivité.

En tant que textes identifiables, les co-textes, les intertextes et les (re)textualisations éditoriales successives peuvent être regroupés au sein d'un *corpus de travail*. Ce point méthodologique essentiel est également illustré par l'article de DELPHINE DENIS. La contextualisation de *L'Astrée* est réalisée au moyen de la constitution d'un corpus raisonné de documents

de nature hétérogène qui constituent autant de ressources hypertextuelles susceptibles d'être enrichies au gré des recherches et des découvertes.

3. Texte, discours, langue

Comme le souligne ici-même PATRICK CHARAUDEAU, les concepts de texte (T) et de discours (D) gagnent à être soigneusement distingués en gardant bien en tête qu'« un même texte est porteur de divers discours » et qu'« un même discours peut irriguer des textes différents ». Ainsi, le discours galant peut irriguer aussi bien *La Belle au bois dormant* de Perrault que *La Princesse de Montpensier* de Madame de La Fayette, le genre des Histoires tragiques peut traverser aussi bien *La Barbe bleue* que les nouvelles écrites par Jean-Pierre Camus et François de Rosset quelques années auparavant. En ce sens, CHARAUDEAU a raison de dire « Le discours n'est pas le texte mais il est porté par des textes. Le discours est un parcours de signification qui se trouve inscrit dans un texte, et qui dépend de ses conditions de production et des locuteurs qui le produisent et l'interprètent ». Dès lors, constituer un corpus en vue d'une démarche d'analyse de discours peut aussi bien se faire à partir d'un texte central (ce qu'illustre DELPHINE DENIS avec le cas de *L'Astrée*) que sur la base d'un genre réunissant des textes qui l'actualisent pleinement ou partiellement ; corpus défini dans des balises temporelles étroites (définition d'un genre à un moment historique de son évolution) ou larges (évolution d'un genre).

Définissant, à la suite de François Rastier (2005 : 31), le texte comme l'unité minimale et le corpus comme l'ensemble dans lequel cette unité fait sens, VINCENT CAPT, JÉRÔME JACQUIN et RAPHAËL MICHELI proposent de croiser trois « sphères de contextualisation » : générique, auctoriale et thématique. Leur approche clarifie la question du *contexte* que posent également GEORGETA CISLARU et FRÉDÉRIQUE SITRI, mais de façon plus classique, en parlant de données « extra-textuelles » et d'« extérieurs » aux textes. CAPT, JACQUIN et MICHELI font de la contextualisation un processus commun aux participants à un échange, aux lecteurs et aux analystes-

chercheurs. Faisant du genre le facteur principal de contextualisation, ils considèrent les contextualisations auctoriale et thématique relativement à la contextualisation générique. Dans leur pratique, CISLARU et SITRI recourent également au critère générique, en notant que sous l'influence des travaux du Cercle de Bakhtine « le cadre générique, ressenti comme plus opératoire, tend à se substituer au cadre institutionnel », et au critère thématique lorsqu'elles procèdent à une analyse contrastive des indices interprétatifs émotionnels dans leur corpus d'écrits de signalement et un corpus de presse écrite traitant de risques et dangers divers : « Dans cette optique, les écrits de signalement peuvent s'intégrer à une série de genres "thématiques" assimilant discours sécuritaires, discours protectionnistes, discours de défense, etc. ; le thème correspondant dans ce cadre à un "discours sur" ». Ce même critère thématique intervient explicitement dans les « paramètres externes » considérés par PATRICK CHARAUDEAU, à côté du « dispositif énonciatif ». Il découpe la « thématique » en capital et hiérarchisation thématiques. De son côté, la perspective choisie par THOMAS LEBARBÉ est d'abord orientée vers une contextualisation auctoriale (articulée autour du nom de Stendhal), complétée dans l'établissement des sous-corpus par des considérations génériques (genre du Journal, Textes sur l'art, etc.). On peut également considérer que le corpus travaillé par NATHALIE GARRIC et VALÉRIE CAPDEVIELLE-MOUGNIBAS est réglé par une contextualisation générique (bilans de savoir et entretiens semi-directifs), les contextualisation auctoriale et thématique étant respectivement redéfinies à travers des catégories comme : « apprenti / lycéen / féminin / masculin » et « échec / réussite / maintien / rupture ».

Comme le pose également l'article de PATRICK CHARAUDEAU, l'établissement d'un corpus en vue d'une analyse de la langue (linguistique de la langue) et celui d'un « corpus de discours » doivent être soigneusement distingués. L'article que nous avons demandé à MICHÈLE MONTE² montre

² A la suite de plusieurs de ses articles et de l'important volume collectif dirigé par Sylvie Mellet : *Concession et dialogisme. Les connecteurs concessifs à l'épreuve des corpus* (Berne, Peter Lang, 2008).

parfaitement que : « La prise en compte du texte comme ensemble cohérent [...] éclaire le fonctionnement de la langue ». Pour préciser le noyau sémantique des connecteurs concessifs *pourtant, pour autant, néanmoins, toutefois* et *cependant* et le fonctionnement de l'apostrophe, elle croise des données de corpus diachroniquement variés, idiolectalement et génériquement différents. De cette diversité d'emploi et de la co(n)textualisation des occurrences des marqueurs considérés surgissent des valeurs que les études classiques ne pouvaient que manquer. Elles prenaient pour « fait de langue » des faits de discours restreints, en procédant à des jugements d'acceptabilité opérant sur des opérations de permutation effectuées sur des exemples réduits et souvent fabriqués.

L'intérêt méthodologique de la démarche proposée part de deux hypothèses fortes. D'abord la constitution d'un corpus aussi varié de formes attestées nous confronte autant à la diversité des données textuelles qu'à la variabilité de la langue. Par ailleurs, l'hypothèse monosémique d'une valeur sémantique propre présente dans tous les emplois d'une forme linguistique donnée « nous oblige à ne pas nous en tenir à un classement superficiel de cette apparente diversité et à chercher la "forme schématique abstraite, stable et plastique à la fois" qui puisse être fédératrice de tous ces emplois ». Les variations pragmatico-discursives observées s'originent dans une valeur commune qu'on peut dire « linguistique » dans la mesure où elle tend à dégager un système de marqueurs concessifs ayant, dans un état de langue donné, des valeurs relatives. Ainsi menée, « l'approche sur corpus à la fois relativise la notion de singularité textuelle en valorisant la famille de textes plutôt que telle réalisation particulière, et revalorise le texte comme totalité multidimensionnelle où les facteurs de cohésion et de cohérence peuvent être étudiés indépendamment mais ne prennent sens que dans leur mise en relation ».

4. « Ouvert » du texte, variation, spatialité

Les réflexions liés à la *variation textuelle*, éclairées sous l'angle particulier de la *mise en corpus*, ont le mérite de déboucher sur un problème d'ordres progressivement logistique,

méthodologique, théorique, et finalement épistémologique, problème qui concentre une bonne partie du questionnement de cette livraison : quelle part joue la spatialité dans la définition différentielle du *texte* (*vs discours, document, énoncé* notamment), telle que travaillée ici dans plusieurs contributions ?

La contribution de MARIE-LAURE FLOREA est à cet égard tout à fait centrale, tant par les avancées qu'elle synthétise, que par les questions, objections et réponses qu'en retour elle ne peut que susciter. La distinction typologique *texte linéaire / texte tabulaire* remet bel et bien en cause les conceptions les plus datées du texte comme linéarité, que dénoncent efficacement les définitions citées³. La *modularité* a été intégrée à la description textuelle dès la fin des années 80, et c'est aujourd'hui un problème terminologique que de distinguer les deux acceptions de *module* qui s'affrontent implicitement chez M.-L. FLOREA. De même que nous allons pouvoir approfondir, dans le programme de la linguistique textuelle (Viprey 2006 ; Legallois 2006 ; Adam 2008, en particulier pages 179-182), la distinction des *régimes de textualité linéaire* et *tabulaire*. Car la démonstration de l'auteure conduit bien à se demander si ces deux *régimes* s'actualisent dans une typologie dichotomique (*textes linéaires vs textes tabulaires*) susceptible à la fois de naturaliser outrancièrement le texte, et surtout de faire perdre de vue la *tabularité* de tout texte (ce n'est certes pas ce que fait M.-L. FLOREA).

Derrière cette discussion, tant pour la servir en propre que pour l'élargir, nous apercevons bien la nécessité d'en revenir toujours aux distinctions conceptuelles (et secondairement terminologiques) déjà évoquées ci-dessus et reprise dans la question : *à quelle échelle parler de texte ?* Ce sont ces questions de spatialité qui nous amènent tout spécialement à invoquer les débats si fructueux relayés par Adam-Heidmann 2005. Revenons un instant sur le terme *module* : lorsque M.-L. FLOREA nomme ainsi les éléments

³ Datées en effet, car même ceux des auteurs cités aujourd'hui disparus ont eu le temps d'y revenir à la lumière des avancées collectives (on pense notamment à Denis Slatka).

d'une double page de manuel ou d'une hyperstructure nécrologique, on peut se demander si ce n'est pas ce que vise Viprey 2009 lorsqu'il propose de considérer le texte comme couture d'énoncés (et de faire passer par ce dernier terme / concept la plurisémiotité, si inconfortable finalement pour le texte proprement dit).

Les suggestions de l'auteure pour remédier à une *conception du corpus, qui ne se conçoit que de façon linéaire* rencontrent déjà un début de réalisation dans mainte entreprise de numérisation, notamment de bases de presse et plus largement bien sûr dans l'intense activité de la *Text Encoding Initiative* (TEI), où les informations sur la localisation (dans le volume, la page, la colonne, la ligne) sont recueillies et normalisées dans le format XML s'imposant à toute archive, et rendues ainsi disponibles pour des études « quantitatives » aussi bien que « qualitatives » (au sens où le formule CHARAUDEAU) : allant de simples comptages contrastifs (le vocabulaire de telle ou telle zone spatiale de la Une des journaux), à l'examen (lui aussi statistique) des mises en relation (d'ordres innombrables) susceptible d'intéresser une recherche comme celle ici décrite⁴.

Car on ne saurait en effet en rester sinon à une *conception*, du moins à une exploitation linéaire des corpus, qui a bel et bien prévalu dans la période que l'on peut qualifier de « tâtonnante » dont sort progressivement la linguistique de corpus. Grâce aux *métadonnées* enregistrées dans la documentation de chaque pièce, l'accès à ces pièces (aussi bien « manuel », par interrogation simple dans un logiciel de lecture, qu'automatisé dans la perspective d'explorations assistées statistiquement) se fait désormais, dans les logiciels de la

4 Cf. par exemple le *Petit Comtois* http://laseldi.univ-fcomte.fr/petit_comtois . La thèse de doctorat de Virginie Lethier, responsable de ce projet et de sa mise en ligne, porte à la fois sur les questions de philologie numérique parmi lesquelles cette prise en charge des coordonnées d'un énoncé dans l'espace, et sur certains aspects de leur exploitation en vue ici d'une mise en œuvre critique des notions de rubricage et d'aire scripturale.

génération actuelle⁵, sur des paramètres variés et multivariés qui délinéarisent le corpus et répondent – de manière non accessoire – à d’autres questions de ce volume (articulation quantitatif / qualitatif notamment) toutes reliées, parfois de manière très profonde donc partiellement masquée, à ce qu’il faut concevoir comme *spatialité textuelle*.

La contribution de NADINE LUCAS explore donc les marges conceptuelles de notre appel à contribution, tout en lui apportant, parfois obliquement, des éclairages très stimulants. C’est notamment le cas lorsqu’elle aborde la spatialité, évoquée ci-dessus, sous la rection ferme du point de vue d’une linguistique informatique soucieuse de textualité. Sa proposition, qui méritera d’être reliée à la teneur entière du n°6 de la présente revue⁶, ainsi qu’aux travaux en cours d’achèvement dans le volet conceptuel du projet *Textométrie* financé par l’ANR⁷, privilégie le focus « recherche » (délimitation et qualification d’un *espace de recherche*, en relation avec les *segments typo-dispositionnels*, paramétrable selon des critères que l’on peut relier à la modularité – sans que le concept soit mentionné). Elle se situe sans équivoque dans une perspective linéaire sur le texte (*début, fin...*), où s’inscrit également la notion de *fenêtre d’observation* (qui se subordonne fortement les éléments usuellement repérés comme constitutifs du modulaire, par exemple ici les parties d’un article de presse), pour déboucher néanmoins sur un ensemble notionnel, *granularité* et *résolution*, qui nous replace au cœur de la discussion animée par M.-L. FLOREA. En effet, NADINE LUCAS fait d’abord se succéder des éléments vraiment comparables par *grains* (articles de longueurs croissantes), pour passer ensuite à un dispositif analogue, *mutatis mutandis*, à ce qu’était tout à l’heure la page de journal : le dossier de presse. C’est alors qu’elle introduit la discussion décisive, grâce à cette citation de Lévy & Lussault 2003 : « [...] autre chose que la

5 Ce qui manque le plus, ce ne sont pas les logiciels d’exploitation (textométriques), mais les environnements raisonnés d’établissement des données (philologiques), et surtout leur mise en commun.

6 *Corpus* n°6, *Interprétation, contextes, codage*, coord. Bénédicte Pincemin.

7 <http://textometrie.ens-lsh.fr/> coord. Serge Heiden.

taille change, quand change la taille ». Elle réintroduit symptomatiquement les illustrations initialement évacuées – à titre expérimental – et nous montre comment la prise en charge de ce dispositif met en péril le respect des propriétés du modèle éprouvé (ici *thème / rhème*), en obligeant de fait à le reconsidérer dans la perspective (évoquée – et mise de côté au début de l'article) de l'articulation *micro / macrostructure*. Reste à savoir si la résolution optimale, dont NADINE LUCAS indique qu'elle est sans doute à la convergence d'une recherche (a) *par modèle* et d'une recherche (b) plus *ad hoc* et empirique, peut intégrer – et à quelles conditions, surtout dans l'optique (b) –, *granularité* et réglages d'échelle.

5. Qu'est-ce qui a « changé » ?

Notre appel se concluait sur cette question un peu naïve et programmatiquement risquée. On constatera qu'elle n'a été explicitement prise en charge que de façon marginale. Cela ne nous dispense pas d'esquisser de modestes éléments de réponse, en guise de synthèse.

Si cette livraison fait bien état de multiples impacts du travail en corpus sur l'appareil conceptuel et ses agencements internes, on ne peut pas en dire autant des conséquences sur les modalités interprétatives et, plus largement, sur l'ensemble des aspects « déontiques » des praxis évoquées. La suspension de l'activité interprétative dans le cours du travail sur corpus informatisé, conçue comme condition d'une relance herméneutique mieux contrôlée et plus conforme aux vœux d'un Jean Bollack ou d'un François Rastier, reste à décrire et à conceptualiser. L'approfondissement de la question même de la construction de l'interprétation, traitée *de facto* et *per factum* par les deux « binômes » d'auteurs (GARRIC / CAPDEVIELLE-MOUGNIBAS et CISLARU / SITRI), attend son heure avec une véritable généralisation des principes de l'Analyse *Textuelle* des Discours.

Si nous estimons ce volume assez représentatif, c'est justement parce que les quatre grandes rubriques de cette courte présentation résument en effet « là où ça change ». Un progrès décisif concerne donc la prise en charge de la pluralité / plasticité

de l'objet texte sous le geste philologique renouvelé, qu'il intègre la *variation stricto sensu*, la contextualisation *lato sensu*, l'ouverture générique, ou encore la réflexivité de toute mise en corpus à des fins intelligibles. Cette avancée entretient des liens réciproques de conditionnement avec les percées technologiques certes, mais surtout avec leur traduction dans les domaines théoriques « amont » (normes informatiques, XML-TEI, etc.) et « aval » (reviviscence des recherches interdisciplinaires sur corpus en sciences humaines), et la manière dont ces liens se construisent, ici attestée, rendent prudemment optimiste. Ils permettent enfin, en effet, de croire fermement que les nouvelles spatialisations textuelles, qui n'attendent nullement leur théorisation complète pour envahir le futur immédiat des pratiques culturelles humaines, font et feront l'objet de l'attention des spécialistes les mieux placés pour cela : les chercheurs confrontés aux matérialités multiformes du texte dans leurs quêtes discursives.

6. Plan du volume.

Nous avons divisé le volume en deux parties. La première regroupe les synthèses critiques et épistémologiques de PATRICK CHARAUDEAU sur la problématique générale de la construction des corpus en analyse de discours et de DELPHINE DENIS sur la philologie moderne et la question de l'édition d'un texte de l'Ancien Régime aussi complexe que *L'Astrée* d'Honoré D'Urfé. Suivent deux articles auxquels l'écriture en binome confère l'éclairage critique de deux angles disciplinaires convergents, l'un à propos d'écrits de travail (les signalements d'enfants en danger) (GEORGETA CISLARU et FRÉDÉRIQUE SITRI), l'autre sur des énoncés suscités (bilans de savoir rédigés par des lycéens et des apprentis) (NATHALIE GARRIC et VALÉRIE CAPDEVIELLE-MOUGNIBAS).

La seconde partie regroupe des réflexions méthodologiques appliquées à des corpus. C'est le cas de l'article sur les « Sphères de contextualisation » de VINCENT CAPT, JÉRÔME JACQUIN et RAPHAËL MICHELI, de MICHÈLE MONTE sur le traitement de connecteurs concessifs et de l'apostrophe, de MARIE-LAURE FLOREA sur la tabularité, de

NADINE LUCAS sur les problèmes d'échelle des unités de traitement et enfin de THOMAS LEBARBÉ sur le traitement de brouillons et manuscrits de Stendhal.

Références bibliographiques

- Adam Jean-Michel (2005). « Les sciences de l'établissement des textes et la question de la variation », in J.-M. Adam & U. Heidmann (éds), *Sciences du texte et analyse de discours. Enjeux d'une interdisciplinarité*. Genève : Slatkine, pp. 69-96.
- Adam Jean-Michel (2008). *La linguistique textuelle. Introduction à l'analyse textuelle des discours*. Paris : Armand Colin.
- Adam Jean-Michel (2009). « Réécritures et variation : pour une génétique linguistique et textuelle », *Modèles linguistiques* Tome XXX-1, volume 59 : 25-52.
- Adam Jean-Michel et Heidmann Ute (2009). *Le texte littéraire. Pour une approche interdisciplinaire*. Louvain-la-Neuve : Bruylant.
- Cerquiglini Bernard (1989). *Eloge de la variante. Histoire critique de la philologie*. Paris : Seuil.
- Gresillon Almuth (1994). *Eléments de critique génétique*. Paris : PUF.
- Heidmann Ute (2005a). « Comparatisme et analyse de discours. La comparaison différentielle comme méthode », in J.-M. Adam & U. Heidmann (éds), *Sciences du texte et analyse de discours. Enjeux d'une interdisciplinarité*. Genève : Slatkine, pp. 99-118.
- Heidmann Ute (2005b). « Epistémologie et pratique de la comparaison différentielle », in M. Burger & C. Calame (éds), *Comparer les comparatismes*. Lausanne : Etudes de Lettres, pp. 141-159.
- Jeanneret Michel (1994). « Chantiers de la Renaissance. Les variations de l'imprimé au XVIe siècle ». *Genesis* 6 : 25-45.

Corpus de textes, textes en corpus

- Legallois Dominique (2006a). « Le texte et le problème de son et ses unités : propositions pour une déclinaison », *Langages* 163 : 3-9.
- Legallois Dominique (2006b). « Des phrases entre elles à l'unité réticulaire du texte », *Langages* 163 : 56-70.
- Lever Maurice (1996). *Romanciers du grand siècle*. Paris : Fayard.
- Lévy Jacques et Lussault Michel, dir. (2003). *Dictionnaire de la géographie et de l'espace des sociétés*. Paris : Belin.
- Mayaffre Damon (2002). « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus* 1 : 51-69.
- Neefs Jacques (1990). « Manuscrits littéraires : comparaisons et histoire littéraire », in D. J. Bevan & P. M. Wetherill (éds), *Sur la génétique textuelle*. Amsterdam-Atlanta : Rodopi, pp. 7-18.
- Peytard Jean (1993). « D'une sémiotique de l'altération », *Semen* 8 :145-177.
- Rastier François (2005). « Enjeux épistémologiques de la linguistique de corpus », in G. Williams (éd.), *La linguistique de corpus*, Rennes : PUR, pp. 31-45.
- Viprey Jean-Marie (2006a). « Quelle place pour les sciences des textes dans l'Analyse de Discours ? », *Semen* 21 : 167-182.
- Viprey Jean-Marie (2006b). « Structure non-séquentielle des textes », *Langages* 163 : 71-85.
- Viprey Jean-Marie (2009) « Language / Discourse : textual analysis, computer and statistics », *Belgian Journal of Linguistics* 23 (1) : 191-206.
- Zumthor Paul (1972). *Essai de poétique médiévale*. Paris : Seuil.