

**Zeitschrift:** Études de Lettres : revue de la Faculté des lettres de l'Université de Lausanne  
**Herausgeber:** Université de Lausanne, Faculté des lettres  
**Band:** - (2005)  
**Heft:** 1-2

**Artikel:** Philologie numérique et herméneutique intégrative  
**Autor:** Viprey, Jean-Marie  
**DOI:** <https://doi.org/10.5169/seals-870107>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

**Download PDF:** 13.03.2025

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

## PHILOLOGIE NUMÉRIQUE ET HERMÉNEUTIQUE INTÉGRATIVE

Les articles de Dominique Maingueneau et de Jean-Michel Adam oscillent, pour la *philologie*, entre une définition restreinte, illustrée par cette phrase de l'argument sur la base duquel se sont réunis les contributeurs du présent ouvrage: «Les domaines couverts vont de la constitution matérielle de l'objet d'analyse (philologie et traduction) à son interprétation [...]» et l'idée d'une «nouvelle philologie», qui frapperait à la porte des interdisciplines, et dont Dominique Maingueneau se demande si elle n'irait pas jusqu'à intégrer *l'Analyse de discours*. Entre les risques d'une refondation totalisante et ceux d'une réduction utilitariste, je voudrais, à partir de la position assez isolée et inconfortable qui est la mienne dans le champ des sciences discursives<sup>1</sup>, travailler sur un hiatus, un manque.

Une chose est d'indiquer ce manque, comme le fait Rastier (2001) sous une forme essentiellement programmatique. Autre chose est de l'incarner, d'être un de ceux-là qui, arrivés dans les sciences discursives au début des années 90, époque de la généralisation de l'informatique, ont projeté de donner sens à cette coïncidence. Il est peut-être trivial de poser que nous participons à la troisième révolution technique de l'histoire universelle du discours, après l'écriture et l'imprimerie. Ce le serait, en fait, si cette révolution était pensée du même pas qu'elle s'effectue.

Dans son article *philologie* de l'*Encyclopaedia Universalis*, Paul Zumthor écrit que l'informatique, qui «fascina les philologues dans

---

<sup>1</sup> Je serai donc amené à dire *je*, pour rendre compte de cette praxis, en alternance avec un *nous* qui désignera la communauté, problématique mais bien réelle, de l'analyse textuelle et comparée des discours, dont ce colloque est une étape.

les années 1970-1980, rendant notamment possible la “critique génétique”», et bien qu’«utile à la collecte des faits, [...] paraît stérile sur le plan notionnel» («Situation actuelle», §1). Peut-être faut-il s’entendre sur l’acception ici employée de «stérile». S’il s’agit de dire que la numérisation n’a pas été théorisée, c’est vrai, et ce n’est injuste qu’à l’égard de secteurs minoritaires des sciences discursives. Mais si l’on soutenait l’idée que la révolution digitale n’est pas porteuse des germes de bouleversements décisifs dans les domaines sémiotiques et herméneutiques, on se tromperait lourdement, se rendant de fait complice des usages (et non-usages) impensés, tous plus ou moins naïfs.

Sans tomber dans une illusion totalisatrice (afin de ne pas y tomber, pourrais-je dire aussi), nous avons impérieusement besoin, pour féconder notre rapport aux nouveaux outils, de croire dans le programme d’une *nouvelle philologie numérique*. Le problème de cette sorte d’utopie est tout entier dans le caractère de *mot-valise* de sa dénomination: à un bout, la *nouvelle philologie* (r)évoquée par Maingueneau; à l’autre, la *philologie numérique* appelée par Rastier. D’un côté, les sciences du discours; de l’autre, les arts du texte.

Il faut donc revenir, encore et toujours, sur les rapports de ces deux notions. Il est difficile de ne pas approuver Jean-Michel Adam lorsqu’il dénonce ici même trois déficits de l’ADF (Analyse de Discours «à la Française») dans son approche du texte: déficit de définition, d’établissement, d’interprétation. Je voudrais apporter quelques compléments à sa critique.

### 1. Texte et discours: définitions?

À propos de la notion du texte, notons tout d’abord que l’ambition d’une *analyse textuelle des discours* souligne bien, et définitivement, non seulement que *texte* et *discours* ne sont pas des objets du même plan, mais encore quelle est leur relation: le *texte* est un *mode* opératoire sur le *discours*. Ne pas parler d’*analyse de textes*, mais d’*analyse textuelle*, indique que le texte n’est pas un objet par soi, mais une *phase* vers l’objet fondamental des sciences humaines qu’est le *discours*. Le syntagme *analyse de textes* n’est peut-être qu’un résidu, au fond, de l’académisme scolaire, résidu qui a connu un paradoxal mais révélateur succès aussi bien dans le champ du structuralisme que dans celui des applications informatisées.

Dans le combat, certes légitime, contre la linguistique *classique* ou *structurale* arrêtée aux bornes de la phrase, ne faisons-nous pas encore quelques concessions quant à la forme même de ce combat?

Je développerai en dernier lieu la plus importante de ces concessions. Commençons par les plus évidentes.

Beaucoup de nos formulations semblent impliquer que *discours* et *textes* seraient «simplement» deux points de vue complémentaires. L'une des plus révélatrices est celle-ci, qu'approuve Jean-Michel Adam dans ce volume même:

En parlant de<sup>2</sup> discours, on articule l'énoncé sur une situation d'énonciation singulière; en parlant de texte, on met l'accent sur ce qui lui donne son unité, qui en fait une totalité et non une simple suite de phrases (Maingueneau 1996: 82).

Je critiquerai ensuite le sens de cette phrase, mais je note d'abord qu'elle présente bien, et fort classiquement, *discours* et *texte* comme deux points de vue opératoires sur un même objet qui serait l'*énoncé*. Un *énoncé* pourrait, selon le point de vue adopté, être alternativement regardé comme *un discours* ou comme *un texte*. Pour soutenir cela, ne faut-il pas donner à l'*énoncé* un statut exorbitant? Surtout, n'est-ce pas le point d'origine d'une désarticulation lourde de conséquences, celle-là même qui dichotomise deux sous-disciplines aussi infirmes l'une que l'autre: une science des textes sourde au discours, une analyse de discours qui «manque le texte en tant que tel» (Sarfati 2003: 432)?

Surtout, comment soutenir que c'est le statut de *texte* qui, spécifiquement, soulignerait les aspects d'*unité* et de *totalité*? Le statut de *discours* n'y contribue-t-il pas tout aussi décisivement, sinon plus encore? Sans sa dimension de discours, c'est-à-dire sans les enjeux d'une situation, un texte n'est-il pas, précisément, une *simple suite de phrases* — je dirais même: *de mots*? N'est-ce pas la variété des mises en discours qui fait varier les structures mêmes de cette *simple suite*, au point de lui conférer des sens, locaux et/ou globaux, parfois diamétralement opposés?

Plus fondamentalement encore, revenons sur la première proposition: un/le *texte* n'est-il pas lui-même articulé sur une situation d'énonciation? Il serait peut-être préférable de noter que ce sont des situations d'énonciation particulières qui instituent un/le texte.

---

<sup>2</sup> *Parler de* doit être bien compris. Il me semble qu'ici, cette locution signifie employer [pour désigner qqch] le terme *de*. Le gérondif pourrait être paraphrasé: *en lui donnant la dénomination de discours, on articule [...]*. De même est-il important de s'accorder sur la référence des pronoms *lui* et *en*: il me semble qu'ils renvoient eux aussi (comme la sorte d'ellipse qui marque le gérondif initial) à l'*énoncé*.

Pour y réfléchir, relisons par exemple l'ouvrage, insuffisamment signalé aujourd'hui, de Renée Balibar (1985). Relisons-le comme une des plus poussées et brillantes entreprises de mise en correspondance, de don réciproque de sens, entre l'analyse de discours «à la française» et la grande tradition philologique dont la linguistique restreinte apparaît bien ici comme un avatar affaibli. Relisons d'ailleurs, tout d'abord, son plus important et utile signalement actuel, à savoir l'article *Colinguisme* du *Dictionnaire d'analyse du discours*, rédigé par Sonia Branca-Rosoff (2002: 102). Il me semble très révélateur que Sonia Branca-Rosoff n'emploie pas une fois, dans ces deux pages, le terme de *texte*. Alors que la démonstration par Balibar du *colinguisme* et de la *grammatisation* repose presque entièrement sur l'histoire de quelques textes co-institués avec les langues qu'ils vectorisent, en des situations d'exception.

Si, plutôt que de chercher un critère unique, ou une grille homogène, de distinction des termes *texte* et *discours*, nous œuvrons à les faire servir ensemble, ne devrions-nous pas dire qu'un texte est, à maint égard, l'institution d'un discours (et que nombre de discours ne donnent ainsi pas lieu à un texte)? Ainsi surmonterions-nous la triviale opposition oral/écrit sans perdre l'intuition fondée et féconde qui la sous-tend.

Un *discours* est un acte socio-affectif, *le discours* au plan générique étant cette activité incessante, ininterrompue, qui engrène tout discours singulier. Le *texte* serait alors la teneur verbale d'un discours en tant qu'il s'institue à un moment et en un lieu cruciaux.

Le *texte* est un artefact au sens plein du terme. Non seulement un *construit*, mais le construit d'une *tekhné*, une mise en ordre valorisante, toujours tierce (l'établissement du texte est le fait d'un tiers, même si ce tiers partage l'identité de l'énonciateur; en d'autres termes, rédiger, grammatiser la teneur d'un discours est une action *supplémentaire*, d'ordre nécessairement technique, que le rédacteur soit ou non celui qui a pris en charge le discours). C'est à ce prix, au prix de formulations aussi précises et plus encore, que la dénonciation du *déficit philologique* pourra déboucher sur des propositions positives et riches de sens.

Mais en conséquence et en retour, *textualiser*, qui devient en propre un travail singulier et normé, réagit sur le statut du discours concerné. Lorsque Nithard rédige et consigne les *Serments de Strasbourg*, il institue les discours de Charles le Chauve et de Louis le Germanique, actes d'abord circonscrits (février 842), en monuments de deux langues nationales encore à naître. Il est assez facile de s'entendre là-dessus. Mais lorsque Jean-Michel Adam critique la tradition invérifiée

et approximative du discours d'Alger (*Je vous ai compris*) ou de celui de Montréal, et (r)établit une version de leurs textes (plus) conforme à l'enregistrement sonore, il ne fait pas que ce qu'il avait d'abord l'intention de faire. Il leste aussi ces discours d'une charge historique supplémentaire, d'une «plus-value de plus».

Responsabilité du *philologue*: non seulement il ne faut pas (comme on le fait encore trop souvent en *lexicométrie*) prendre n'importe quel «texte» comme échantillon de n'importe quoi, mixer *du texte* en des corpus de bric et de broc, mais il faut songer que, *nolens volens*, on ajoute une valeur supplémentaire au discours correspondant, chaque fois qu'on en «réveille» le texte. Réponse, ou élément de réponse, à la question légitime de la *valeur* souvent négligée, et quelquefois cyniquement, dans la pratique des documents.

## 2. Philologie numérique

Si le texte est bien la *mise en ordre valorisante* d'un discours, on perçoit mieux les enjeux de la philologie numérique, qui se présente légitimement comme *nouvelle philologie*. Nouvelle, non pas au sens seulement où elle rénoverait l'ancienne, mais au sens plus large où elle renouvelle les pratiques du texte après une phase de relatif mépris pour ce dernier, phase peut-être inévitable dans le développement de l'analyse de discours, mais aujourd'hui révolue. Seule une philologie renouvelée au contact et au défi du numérique peut espérer combler le *déficit philologique*.

J'envisagerai trois aspects de cette question : le balisage instancier, l'étiquetage lexico-flexionnel et l'annotation «savante».

Le balisage instancier est, primitivement, l'ensemble des moyens par lesquels parmi un corpus de textes, on peut instancier un texte, un segment de texte, *i.e.* l'attribuer, pour des recherches contrastives ultérieures (sans parler d'impératifs documentaires, qui ne sont pas ici notre objet), à une variable énonciative (auteur, titre de presse, tranche diachronique, et plus largement tout critère exogène codifié par un manipulateur du corpus: genre, critères de contenu, style, etc.). En voici un exemple élémentaire, tiré d'un pré-traitement du corpus du *Monde diplomatique 1980-2000*:

```
<TITRE_Le déclin industriel de l'Europe est-il irréversible?>
<AUTEUR_MESSINE PHILIPPE;DEVILLE GEORGES>
<DATE_198312>
```

L'EUROPE est en crise: constatation banale, dictée par l'observation quotidienne de l'actualité. Crise économique, avec sa cohorte de chômage, de sous-investissement et de rigueur budgétaire.

Mais au-delà de ces fonctions primitives, au demeurant fort utiles, le balisage instancier se révèle très prometteur en vue de l'intégration à la documentation du texte, à son fichier même, de ses propres variantes et ramifications. On prendra cette fois l'exemple des *Fleurs du mal*, sur lesquels j'ai publié un premier essai de *mise en ordre valorisante* (Viprey 2002). Dans cet ouvrage, qui reste encore largement en attente des soins d'une philologie numérique raisonnée et outillée, je discute du mode d'exposition des variantes. Contraint, à cette date, de choisir une édition de référence, je me rallie à la bonne pratique de la dernière version du vivant de l'auteur, en l'occurrence celle de 1861. Les variantes sont stockées dans des fichiers à part, stigmatisées en quelque sorte. Les problèmes «techniques», même pour rendre compte d'une histoire éditoriale très simplifiée (1857, 1861, 1868), notamment de la place des pièces condamnées en 1857, de l'impact de leur retrait, restent indomptés. L'ordre linéaire des pièces du corpus est intangible, alors que les éditions successives les rangent différemment; qui niera pourtant que l'emplacement d'un poème fait partie de sa textualité?

Aujourd'hui, nous pouvons donner accès à l'une ou l'autre des éditions, voire à des configurations entièrement inédites mais susceptibles d'intéresser le lecteur<sup>3</sup>, grâce à une mise en série parallèle des variantes que seul le moteur de recherche aura à gérer. Voir par exemple la Fig. 1<sup>4</sup>.

```
<poème> <PO_TITRE> Un Voyage à Cythère </PO_TITRE>
<PO_VAR> MS,DP,RDM,57,61,68</PO_VAR>
<vers>Mon cœur<VAR_MS>, comme un oiseau, s'envolait tout</VAR_MS>
<VAR_DP>comme un oiseau s'envolait tout</VAR_DP> <VAR_RDM,57>se
balançait comme un ange</VAR_DP> <VAR_61,68>, comme un oiseau,
voltageait tout</VAR_61,68> joyeux <VAR_RDM>,</VAR_RDM></vers>
<vers>Et planait librement à l'entour des cordages;</vers>
<vers>Le navire roulait sous un ciel sans nuages,<.vers>
<vers>Comme<VAR_MS,DP>un oiseau qu'enivre un</VAR_MS,DP>
<VAR_RDM,57,61> >un ange enivré d'un</VAR_RDM,57,61> <VAR_68>
>un ange enivré du</VAR_68> <VAR_MS>Soleil</VAR_MS> <VAR_DP,RDM,57,61,68>soleil</VAR_DP,RDM,57,61,68>radieux.</vers>
```

Fig. 1: Exemple de balisage possible des variantes du 1<sup>er</sup> quatrain d'«Un Voyage à Cythère».

<sup>3</sup> On admettra définitivement ici une équivalence *lecteur/chercheur*.

<sup>4</sup> Ici le premier quatrain d'«Un Voyage à Cythère». Cet exemple est simplifié, puisque les annotations diverses ont été estompées (notamment l'étiquetage morpho-flexionnel, voir *infra*).

Ainsi devient-il matériellement beaucoup plus tangible (qualitativement plus tangible), devient-il aussi opératoire, que le texte est la configuration de toutes ses variantes, attestées ou potentielles.

Sur l'étiquetage lexico-flexionnel, on ne s'étendra pas en longueurs inutiles ici. Notons avant tout que l'identification lexicale et grammaticale n'est pas d'une nature ou d'un ressort fondamentalement différents de la graphémisation. Il s'agit dans les deux cas d'un artefact technique valorisant et instituant: l'incorporation au texte de signalements codifiés, normalisés, favorisant la lecture et l'interprétation cohésives. Certes, en l'état actuel, l'étiquetage n'est jamais le fait d'un scripteur «premier», mais rien n'interdit de penser que la saisie des textes puisse évoluer vers ce type de prise en charge.

Quant à l'annotation «savante», c'est le volet stratégiquement le plus important de la nouvelle perspective philologique. L'annotation numérique permet d'envisager la réconciliation des tenants du texte «nu» et des éditeurs scientifiques, à condition toutefois que les premiers adoptent un point de vue critique de nature à problématiser la «nudité» en question, et que les seconds acceptent d'offrir une option d'affichage «lisse», sans aucun appel de note ou de variante, auquel le lecteur aspire en effet en alternance avec l'affichage de l'apparat critique. Ainsi deviendra-t-il plus acceptable de se voir, d'une manière plus potentielle donc, car plus optionnelle, expliquer ce que c'est qu'un *houka*, à quelle acception de l'adjectif *inquiet* La Fontaine renvoie au vers 80 des «Deux Pigeons»... voire à quelle maîtresse de la biographie de Baudelaire, d'après le commentateur, renvoie *naturellement* telle figure poétique... et même peut-être de se faire dire, par Antoine Adam, à propos de ces vers de «Chant d'Automne»<sup>5</sup> (*Les Fleurs du mal*)<sup>6</sup>:

Bientôt nous plongerons dans les froides ténèbres<sup>2</sup>;  
 Adieu, vive clarté de nos étés trop courts!  
 J'entends déjà tomber avec des chocs funèbres  
 Le bois retentissant sur le pavé des cours<sup>3</sup>.

Successivement en note 2: «Ce vers doit être pris au pied de la lettre. L'hiver approche. Nous sommes au mois d'octobre 1859», et en

<sup>5</sup> Poème dont, pour illustrer le point immédiatement précédent, l'excellent Claude Pichois nous dit tout de même (Pléiade, p.934): *Poème inspiré par Marie Daubrun, dont on reconnaît (sic) les yeux verts.*

<sup>6</sup> Les appels de notes inclus dans la citation sont ceux de l'édition d'Antoine Adam (1994, Classiques Garnier). C'est à eux que correspondent les notes évoquées dans le § suivant.



note 3: «Dans la vie parisienne d'autrefois, le mois d'octobre se passait à faire la provision de bois pour l'hiver tout proche...»<sup>7</sup>.

Il n'y a aucune malveillance gratuite dans la mobilisation et la mise en vedette de ces exemples qui paraîtront extrêmes. Toute intrusion encyclopédique, quelles que soient sa pertinence et/ou sa naïveté, toute insertion philologique, pourtant si nécessaire dans les perspectives de plus en plus profondément diachroniques que comporte la lecture intégrale du texte littéraire français, interviennent comme des déchirures dans le discours-lecture. L'ergonomisation, l'optionnalisation, la hiérarchisation des annotations ne sont pas un luxe, mais un élément essentiel du retour au texte, du confort intellectuel, du contact des œuvres. Elles permettent de spéculer sur une annotation infinie, multi-critère et processuelle, dont les signets hypertexte sont une préfiguration partielle.

On est même en droit d'espérer que dans un avenir point trop lointain, les travaux critiques pourront — devront — être publiés en lien matériel avec leur corpus d'étude, de manière à rendre, au moins en partie, ressaisissables les démarches qui les fondent. C'est d'ores et déjà possible pour certains travaux sur des corpus libres de droits. La sphère d'application s'élargira au fur et à mesure que seront remaniés, sous l'impact des nouveaux supports, les cadres juridiques en matière de droits d'auteur et d'éditeur, de propriété intellectuelle, commerciale, industrielle.

On peut conclure cette seconde partie en remarquant que toutes ces composantes entrent technologiquement sous une bannière unique, celle de l'enrichissement par balisage. Quels que soient les choix logiciels adoptés par tel ou tel système ou environnement, tel laboratoire, telle école, tel individu isolé, pourvu qu'ils soient réglés dès le départ par les soucis de la compatibilité et de la réversibilité optimales, que toute l'annotation (instancielle, linguistique, encyclopédique) soit stockée dans un seul et même fichier avec les données textuelles, ou fasse au contraire l'objet d'un référencement externe, dès lors que des liens permanents et explicites (les hyperliens, de façon générique) sont marqués, apparaît clairement une plus complète définition du texte, qui n'est donc pas indépendante de la question du *déficit* ou du *comblement* philologiques: loin d'être réductible à l'énoncé du

---

<sup>7</sup> André Gazier, Professeur en Sorbonne, éditeur d'un recueil sélectif des *Fables* en 1925 (Armand Colin), précise de même dans le paratexte didactique de «La Cigale et la Fourmi»: *Les cigales ne chantent pas: le bruit aigu qu'elles font entendre est produit par le frottement de deux membranes [...]*.

discours, le texte qui en est l'institution, est un ensemble multicouches et multipolaire, extensible et réductible selon les besoins de toute remise ultérieure en jeu, en discours. Ce n'est donc pas seulement parce qu'il comporte des dimensions tabulaires et réticulaires (qu'il partage avec l'énoncé), que le texte est non-linéaire. C'est parce qu'il est, de par sa nature d'artefact et dès l'origine, hypertexte.

C'est pourquoi on peut s'étonner que Paveau & Sarfaty (2003: 186), dans une récapitulation par ailleurs importante qui accorde une place d'honneur aux linguistiques discursives, citent le schéma texte/discours de Jean-Michel Adam, tel qu'il fut formulé en 1990 (Adam, 1990: 23):

DISCOURS = Texte + Conditions de production  
 TEXTE = Discours – Conditions de production

sans mentionner l'importante rectification opérée par l'auteur lui-même (Adam, 1999: 23), en particulier pour retirer la notion de *soustraction du contexte*, qui était des plus problématique.

Les rapports du texte et du discours semblent décidément plus complexes. Passons sur la connotation économiste du terme *conditions de production*, encore souvent privilégié. On songe à *formation discursive*, mais la formule devient alors tautologique. Le véritable problème est bien l'intervention exclusive d'une soustraction, comme si le texte était simplement une réduction du discours. Or le texte est aussi *plus* que le discours (son institutionnalisation en langue(s), comme le montre Renée Balibar), il y *ajoute* donc en quelque sorte des aspects qu'on pourrait dire *productifs* eux aussi. Comment, par exemple, Chateaubriand s'y prend-il pour «produire» les *Mémoires d'outre tombe*? Il semble bien qu'il passe directement au niveau du texte, cousant à des segments inédits les textes remaniés de discours antérieurs, ajoutant son propre appareil de scénographie et de «navigation». L'effort de texte réintègre ici tout ce qui peut l'être des *conditions de production* accumulées à travers plusieurs décennies historiques, en fait le ciment même du monument à l'œuvre, et règle (comme c'est majoritairement le cas dans tout le discours «littéraire», mais c'est ici particulièrement emblématique, car entièrement réflexif) sa remise en discours. Ainsi l'effort de texte est-il bien, convenons-en, un désancrage *énonciatif*, ce qui est notablement distinct de la question de la *production*. L'incorporation de la scénographie discursive (dont relèvent pour l'essentiel aussi les variantes, celles du moins qui sont attestables, parce qu'elles ont été — sciemment ou non — conservées), l'incorporation de l'interdiscours, les remaniements macro et micro, justifieraient même de réserver le terme de *production* au

texte, et peut-être de substituer à *Conditions de production* une formule référant à l'*énonciation* ou à la *mise en circulation* (comme on parle de *mise sur le marché*).

Quant à savoir, d'autre part, si dans la perspective critique-herméneutique il suffit d'extraire la composante verbale d'un discours pour accéder au texte, il me semble que tout plaide pour un *non* ferme et catégorique à cette question, et ce tout spécialement dans les questionnements de ce colloque même, et dans sa dénonciation des déficits philologique et herméneutique.

Il se trouve même qu'une partie au moins de la question du déficit herméneutique, que je vais tenter de traiter en troisième lieu, semble conduire aux mêmes déterminations.

### 3. *Texture, réseaux, dictionnaires dynamiques: propositions pour développer une herméneutique intégrative*

On ne peut qu'admettre le postulat fondateur des linguistiques textuelles, à savoir que les faits de langue ne sont pas circonscrits à la phrase grammaticale, et sa conséquence, à savoir que «la compétence des sujets leur permet de distinguer une collection de phrases sans liens d'un tout unifié» (Adam 1999: 9, commentant Halliday & Hazan). Il était sans doute nécessaire et inévitable que, dans le cours de son développement, dans l'affirmation de sa place légitime, dans son dialogue avec l'analyse de discours et la sémantique des textes, «notre» linguistique textuelle en vienne à privilégier grandement les principes et les aspects de la *cohésion*, à affirmer même l'inséparabilité de celle-ci d'avec la *progression thématique* (Adam, in Charaudeau & Maingueneau 2002: 99), bref à mettre au premier plan de l'analyse les dimensions macro-séquentielles. Il est temps néanmoins de nous mettre en garde contre les risques réductionnistes qu'emporterait la poursuite non maîtrisée de cette priorité.

Le concept de *texture*, notamment, mériterait d'être clairement et à nouveaux frais distingué de celui de *textualité*. Ce dernier terme conviendrait mieux que *texture*, retenu par Paveau & Sarfaty (2003: 188) pour désigner l'ensemble des faits d'organisation du texte, assurant *sa continuité sémantique*. Dès lors, le terme de *texture* pourrait être réservé, par-delà toute dimension macro-séquentielle, à la/aux dimension(s) micro-distributionnelle(s)<sup>8</sup>, dont on admettra que pas

---

<sup>8</sup> Ce que fait Jean-Michel Adam (1999: 40): «*texture* pour désigner les faits micro-linguistiques».

plus que les précédentes, elles ne peuvent faire défaut: problématique et dialectique du global et du local, qui ne s'enferment pas dans la seule question aporétique du *style*, mais qui concernent plénièrement la consistance sémantique d'ensemble. Celle-ci (plutôt que la *cohérence*, qui est une autre dimension), dépend bien en effet non seulement de l'occurrence et de la répétition des items lexicaux, mais de leur signification contextuelle, livrée avant tout par les faits de *cooccurrence* (de collocation).

Or le principe même de l'observation de la *texture* ainsi entendue, c'est la délinéarisation du texte, l'occultation des bases mêmes de la séquentialité. L'informatique, la numérisation des ressources, et plus nettement encore les opérations de philologie numérique telles que décrites *supra*, ouvrent grande la voie de retour au «micro-local». Et celui-ci n'a rien à voir (et c'est peut-être là que s'installe un malentendu avec Meschonnic sur nos propositions en matière de statistique *textuelle*<sup>9</sup>) avec la linguistique réductrice, *intra-phrastique*: est tout autant *transphrastique* que les concepts de *plan*, *périodes*, *phrases*, *propositions*, celui de l'organisation *micro-distributionnelle* du texte, faite de répétition et d'altération locales, internes, endogènes, sans laquelle on ne saurait en droit évoquer la *texture*.

J'ai tenté de montrer (Viprey 2000) que sous certains régimes textuels en tout cas même le phonème devait être considéré comme un macro-organisateur (et qu'il s'agit même de l'authentique interprétation explicative de l'*allitération*). Que ne pourraient dès lors les vocables, dès lors qu'on les observe en réseau?

Une première manière, éprouvée, d'observer le vocabulaire reconduit directement à la séquentialité: c'est l'étude de la répartition des occurrences au fil du texte (que l'on nomme parfois, au risque d'une confusion, leur *distribution*, et que je nomme plus volontiers leur *macro-distribution* ou leur *distribution séquentielle*, justement). On peut voir ainsi se dessiner des séquences thématiques plus ou moins marquées, dont les profils s'ajusteront d'ailleurs plus ou moins aux repérages séquentiels effectués selon d'autres critères (marqueurs spécifiques, grammaticaux ou typographiques, notamment).

Une seconde voie est l'analyse, qui nous a été inspirée de la lecture de Harris (1969), mais aussi de Spitzer (1970) et de Riffaterre (1971): analyse contrastive, appuyée sur les méthodes multi-dimensionnelles

---

<sup>9</sup> Dont l'expression, le développement opératoire doit être pris comme une pauvre *maquette*, et dont il faut avant tout (re)lire le *programme* avec la sollicitude qui sied au scientifique.

d'affiliation benzecriste<sup>10</sup> des profils collocatifs des vocables *dans la globalité* du texte. Quels que soient les hauts et les bas de l'emploi d'un vocable *au fil du*<sup>11</sup> texte, sa contribution globale (sa signification) ne saurait être pleinement envisagée sans donner son plein statut à la place qu'il va ainsi occuper dans le *système* des collocations.

Développons, très sommairement, un exemple.

Soit la collection du *Monde Diplomatique 1980-2000*, éditée par le journal lui-même en CD-Rom. S'agit-il bien d'un texte? Sans doute, si l'on admet que cette publication, effectuée à l'occasion du nouveau siècle, est une (re)mise en discours sciemment concertée par la rédaction de l'époque. Ce texte pose certes, d'une manière différente mais tout aussi insistante que le faisait un recueil poétique (*Les Fleurs du mal*, voir Viprey 1997: 296 ssqq), des problèmes encore irrésolus de cadre méthodologique quant à la séquentialité. Quelle est la séquentialité d'un texte aussi colossal? Tient-elle à des marques «pleines», explicites? Non. S'exprime-t-elle dans des catégories aléatoires et accidentelles comme les années, les lustres? Guère.

Considérons donc plutôt ce qu'il peut en être d'une *progression thématique*. Qu'est-ce qui peut nous y donner prise? Observons l'évolution diachronique de l'emploi de deux cents items thématiques<sup>12</sup>, en constituant un tableau à double entrée au fil du texte: en lignes, les items lexicaux, en colonnes, les années; à l'intersection, le nombre d'occurrences de l'item dans les douze numéros de l'année, paratexte, tableaux et cartes exclus. La fig. 2 présente un échantillon d'un tel tableau.

	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
accord	196	177	202	294	243	231	254	255	214	253	164	201	227	323	258	247
action	300	250	254	241	266	218	211	230	184	161	211	166	193	191	184	174
administration	150	237	231	195	238	217	145	142	138	149	87	146	116	150	117	137
affaires	288	283	223	257	253	206	191	222	212	255	256	278	292	286	291	266
afrique	298	361	308	225	359	332	450	192	273	196	244	256	246	282	304	165
agit	189	187	188	184	182	204	173	177	160	179	165	144	173	156	130	173
aide	394	333	282	318	331	260	375	297	249	276	245	282	302	317	276	202
allemagne	112	107	75	145	109	128	72	109	138	125	321	232	238	206	205	120
américain	226	222	244	252	281	242	291	312	210	175	184	241	198	201	187	177
américaine	293	264	295	336	348	293	304	262	230	205	191	273	239	213	194	207
américains	283	229	288	325	300	291	298	322	262	227	232	299	206	201	181	204

Fig. 2: Distribution diachronique de 200 items lexicaux fréquents dans 21 années du «Monde diplomatique» (échantillon du tableau).

<sup>10</sup> On verra avant tout pour ce domaine Salem & Lebart (1994).

<sup>11</sup> On notera, en italiques, deux expressions explicitement antonymiques.

<sup>12</sup> De manière très grossière, nous définirons comme *items thématiques* des unités graphiques (le texte n'est pas lemmatisé), ne renvoyant pas à des catégories

L'Analyse Factorielle des Correspondances (AFC), appliquée à (l'ensemble de) ce tableau, offre un compte rendu synthétique de la macro-distribution lexicale, représenté par le graphe de la Fig. 3. *Grosso modo*, dans ce nuage de points indexés: les items lexicaux sont répartis sur le graphe, par projection sur chacun des deux axes orthonormés, en fonction de leurs parentés (proximité) et oppositions (éloignement) de profils macro-distributionnels (ici, donc, diachroniques), et ce d'autant plus significativement qu'ils (leurs points) sont excentrés; de même pour les items *années*, dont les proximités signent la parenté de profil lexical (les mêmes excédents et déficits en occurrences d'items lexicaux); la superposition des deux nuages (lexique et années) signale les principales correspondances positives.

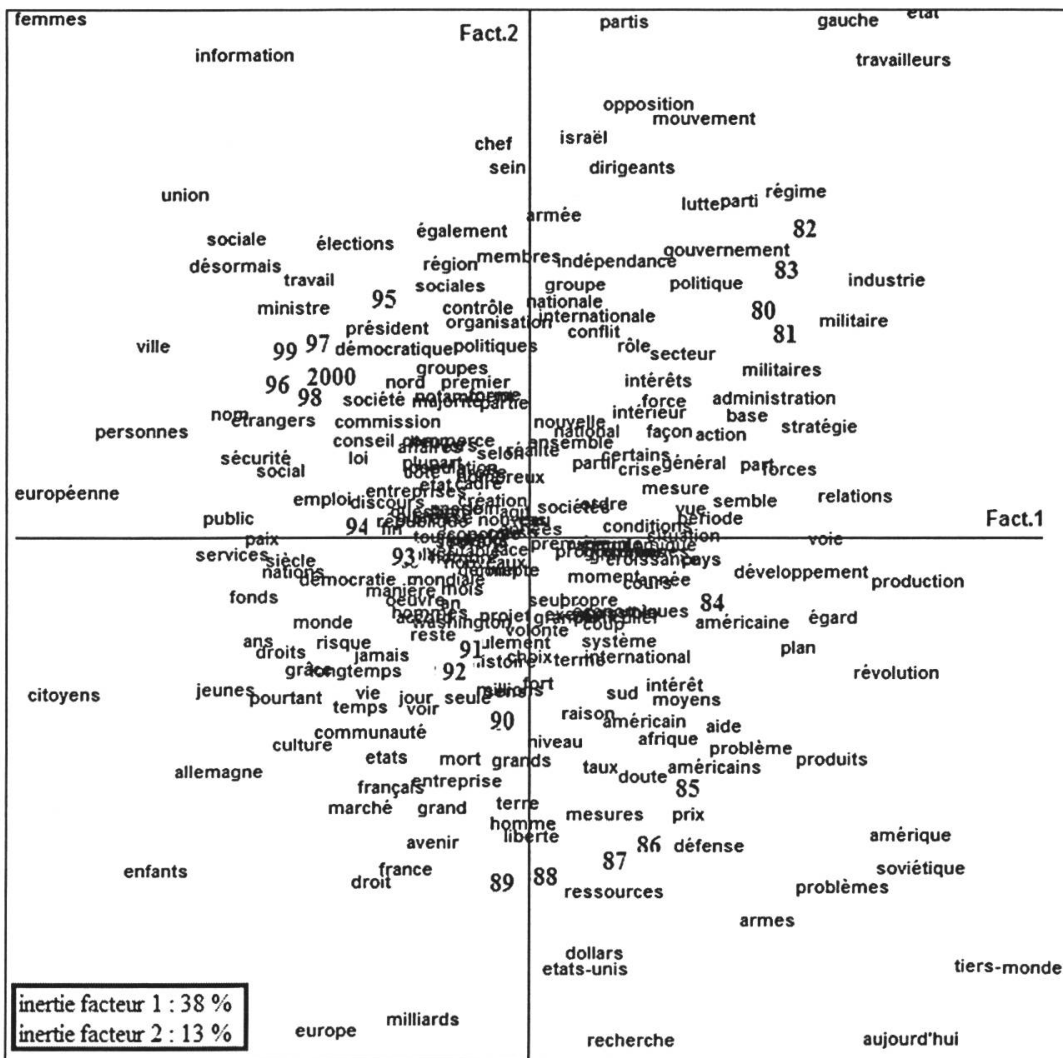


Fig. 3: AFC de la distribution diachronique de 200 items lexicaux fréquents dans 21 années du «Monde diplomatique» 2 premiers facteurs.

La Fig. 3 permet de structurer une hypothèse diachronique plus précise et pertinente que celle des années, quant à la *progression thématique* particulière d'un tel type de texte: on observe d'abord, bien sûr, un clair ordonnancement de 1980 à 2000 et on peut repérer les groupements d'items lexicaux responsables, par leur changement (proportionnel) d'emploi, de cet ordonnancement; mais on peut aussi repérer des ébauches de «séquences»: 1980-83, 1984, 1985-89, 1990-94, 1995-2000. «Étrangement», à l'exception de 1984, isolée, on retrouve des «lustres» ou «quasi-lustres», comme si une périodisation symbolique s'était imposée au discours. En tout état de cause, aucune interprétation hâtive ne doit être tirée. Outre que beaucoup de raffinements doivent être envisagés, seul un examen approfondi comportant une vaste alternance de retours au texte linéaire et tabulaire permettrait d'échafauder cette hypothèse ou toute autre.

Envisageons maintenant ce qu'il en est, toute séquentialité mise à part, neutralisée, de la structure micro-distributionnelle de ce texte, en partant de la même liste d'environ 200 items thématiques. Nous constituons à cet effet un autre tableau à double entrée, dont les lignes sont toujours les items lexicaux, mais dont les colonnes sont aussi ces mêmes items lexicaux. À l'intersection, le nombre de cooccurrences des deux items concernés dans les limites d'une phrase<sup>13</sup>, et d'un empan maximum de 15 mots. La Fig. 4 présente un échantillon de ce tableau, qui est symétrique autour de sa diagonale.

	accord	action	administration	affaires	afrique	agit	aide
accord	91	42	53	73	65	38	77
action	42	78	26	33	54	44	75
administration	53	26	48	65	48	16	72
affaires	73	33	65	100	86	47	66
afrique	65	54	48	86	119	28	121
agit	38	44	16	47	28	54	33
aide	77	75	72	66	116	33	117

Fig. 4: Micro-distribution (cooccurrences) de 200 items lexicaux fréquents dans 21 années du «Monde diplomatique» (échantillon du tableau).

---

autres que verbes (à l'exception des auxiliaires les plus usités), substantifs, adjectifs (à l'exception des déterminatifs), et adverbes de temps, de lieu, ou dérivés d'adjectifs.

<sup>13</sup> Définie sommairement par des critères de ponctuation.

L'AFC, appliquée à ce tableau, offre maintenant une synthèse de la micro-distribution lexicale, représentée par le graphe de la Fig. 5. On ne garde qu'un nuage de points (celui des colonnes étant exactement identique); dans ce nuage, les proximités signent cette fois les parentés de profils collocatifs, avec toujours cette réserve: d'autant plus significativement que les points sont excentrés. En d'autres termes, les items proches sur le graphe ont des contextes massivement similaires du point de vue de leur constitution lexicale, ils ont les mêmes cooccurrents.

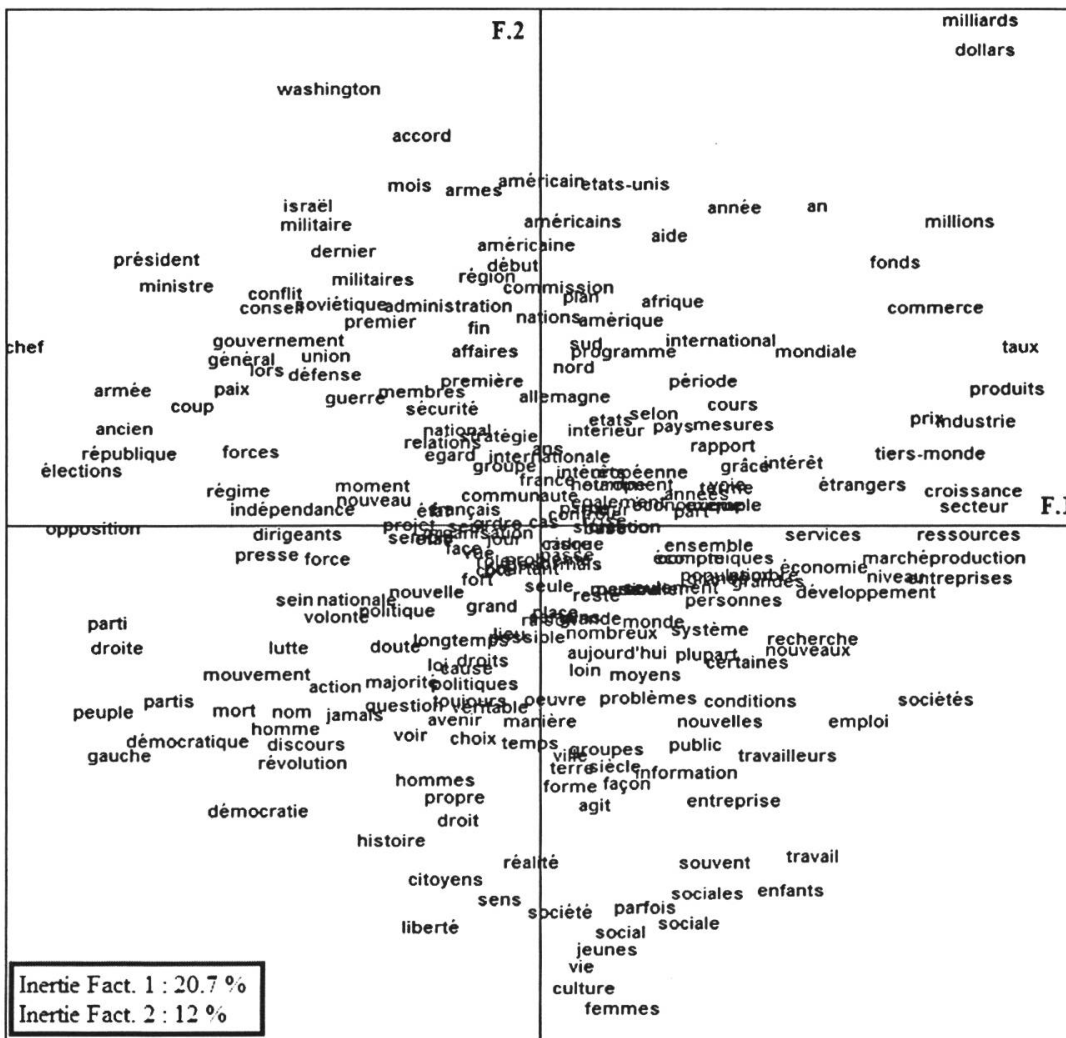


Fig. 5: AFC de la micro-distribution de 200 items lexicaux fréquents dans 21 années du «Monde diplomatique» 2 premiers facteurs.

J'ai proposé de compléter et d'éclairer, sinon de remplacer, la notion-clé en analyse textuelle d'*isotopie* par celle d'*isotropie* (*trepein*: pencher vers), qui correspondrait justement aux tendances collocatives,



et surtout à la structuration lexicale systématiquement singulière que ces tendances déterminent, telle que la manifeste ce genre de projection graphique. Le fait, par exemple, que ces termes qui occupent le bas du graphe, d'une part «aillent ensemble» le plus communément, d'autre part montent du texte même, par une opération de simple explicitation, comme l'une de ses saillances majeures, est à la fois typique et singulier. De même bien sûr pour tous les groupements, plastiques, non-dichotomiques, que l'observation minutieuse, exhaustive du graphe, alternante avec de substantielles phases de retour au texte linéaire et tabulaire, mettra en valeur.

La méthodologie de l'AFC, appliquée à la macro-distribution, est d'un usage bien établi, connu. Son application aux micro-structures lexicales est plus récente. Si je l'évoque ici, ainsi que les travaux qu'elle engendre, c'est pour en proposer un usage radicalement nouveau, en lien explicatif et illustratif direct avec le sujet qui nous préoccupe: la philologie numérique et les redéfinitions du texte. Typiquement, de telles représentations devraient être désormais, non seulement considérées comme des «états de texte», ce qu'elles sont parfois, au moins implicitement (Salem & Lebart 1994), mais encore technologiquement intégrées au dispositif textuel à part entière, au titre de sa cartographie. Contrairement à ce que l'on a pu d'abord croire, ce ne sont pas des résultats, des états terminatifs. Reçus comme tels, elles sont trompeuses ou au mieux opaques. En revanche, si les items du graphe sont cliquables, dans le cadre d'un dispositif d'exploration (que je souhaite nommer *hypertexte expert*), et que ce click ouvre sur l'affichage des contextes, ou sur le retour au texte, ou sur d'autres ressources que l'on aura fait monter, elles aussi, du texte, alors le graphe devient une véritable carte au service de la navigation dans le texte, d'une *autre* lecture (non nécessairement linéaire). Carte, mais aussi dictionnaire dynamique, en liste ou en graphe, puisque débouchant sur des concordances, complètes ou sélectives, et sur des extracts statistiques de ces concordances (cooccurents privilégiés de l'item cliqué, notamment).

Les dictionnaires de corpus, ouverts sur les concordances et sur les graphes de micro- et de macro-distribution, sont en rupture avec une conception du *signe* comme *objet de l'interprétation* (Rastier 2001: 103), en ce qu'ils ne *retracent pas indéfiniment des ontologies* ni ne *stabilis[ent] les significations de manière normative*. En contrepartie, séparés du corpus qui les fait monter et les justifie, ils n'existent tout simplement plus<sup>14</sup>.

---

<sup>14</sup> C'est ce qui est ébauché dans une première tentative, très artisanale et empi-

Que faut-il enfin, pour conclure de manière ouverte, à un *corpus* pour être un *texte*? Il y faut et il y suffit qu'une intention critique pertinente et homogène soit à l'œuvre (voir *supra*, fin du 1). L'ambition d'une *herméneutique matérielle*<sup>15</sup> qui réuni[rait] l'*herméneutique* et la *philologie*, me semble précisément d'autoriser, voire de contraindre la critique de textes, la communauté de l'*analyse textuelle des discours* à assumer une fonction de (re)mise en ordre et de (re)mise en jeu publiante, instituante pour ces derniers, renouvelant à la fois l'analyse et l'édition scientifiques.

Jean-Marie VIPREY  
*Université de Franche-Comté*

---

rique, autour des *Fleurs du mal*, où le texte est déposé sur un CD-ROM (Viprey, 2002) en l'état de multiples fichiers: un pour le texte «brut» de 1861, un pour chaque strate de variantes, un pour les étiquettes lexicales, un pour la transcription métrique-phonétique, plusieurs pour les dictionnaires (listes de formes et de vocables), plusieurs pour des graphes d'AFC macro- et micro-distributionnelle, le tout régi par un moteur de recherche visant à permettre au lecteur d'aller sans hiatus d'un état à l'autre, de prendre alternativement des vues d'ensemble par niveaux et des vues au ras de l'énoncé. C'est ce que tente de mieux modéliser le développement ®Astartex, dans le cadre des programmes de la Maison des Sciences de l'Homme de Franche-Comté, pour des textes plus étendus ou pour des corpus spécialisés.

<sup>15</sup> Je reprends le terme bien qu'il me paraisse mériter une certaine mise en discussion. La locution *herméneutique matérielle* semble être une invention de Jean Bollack dans la préface de la traduction par Mayotte Bollack du livre de Peter Szondi, où elle n'est pas employée en dehors de la table des matières, qui est de la main de Jean Bollack (on trouve en revanche *herméneutique pratique*, p. 9 de la traduction) et je ne trouve pas qu'elle figure non plus dans les travaux connus de Schleiermacher, où le thème contrastif *formell/matériel* est très discontinu et assez peu explicite. C'est donc la préface (et son titre) et la table des matières qui semblent créer la notion et la mettre sous la plume de Szondi. Or le principal développement à ce jour de la notion, qui est le chapitre 4 de Rastier (2001), après avoir entériné (p. 100) le coup de force supposé de Bollack: *Peter Szondi a formé jadis le projet d'une herméneutique matérielle. L'expression herméneutique matérielle, reprise de Schleiermacher, [...], lui apporte une certaine instabilité: il est concurremment question, passim, d'herméneutique philologique (p. 127), d'herméneutique critique (p. 99), voire de sémantique interprétative (p. 126). Ne s'agit-il pas de référer nos disciplines discursives à un horizon herméneutique plus net et dégagé des nébulosités ontologiques? En d'autres termes, pour ceux qu'occupe et préoccupe la révolution numérique dans les arts et les sciences des textes, d'insuffler un sens non apriorique, endogène, à nos praxis? C'est en ce sens que je formule l'herméneutique intégrative (ou pratique, pourquoi pas), qui appelle bien, pour sa part, une épistémologie, une méthodologie, une déontologie (Rastier, 2001: 101).*

## RÉFÉRENCE BIBLIOGRAPHIQUES

- ADAM J.-M. 1999: *Linguistique textuelle: des genres de discours aux textes*, Paris, Nathan.
- AMOSSY R. & MAINGUENEAU D. 2003: *L'Analyse du discours dans les études littéraires*, Paris, Presses Universitaires du Mirail.
- BALIBAR R. 1985: *L'Institution du français: essai sur le colinguisme, des Carolingiens à la République*, Paris, PUF.
- CHARAUDEAU P. & MAINGUENEAU D. 2002: *Dictionnaire d'analyse du discours*, Paris, Seuil.
- HARRIS Z. S. 1969: «L'Analyse de discours», *Langages n°13*, Paris, Larousse.
- PAVEAU M.-A. & SARFATI G.-E. 2003: *Les Grandes théories de la linguistique*, Paris, Colin.
- RASTIER F. 2001: *Arts et sciences du texte*, Paris, PUF.
- RIFFATERRE M. 1971: *Essais de stylistique structurale*, Paris, Flammarion.
- SALEM A. & LEBART L. 1994: *Statistique textuelle*, Paris, Dunod.
- SPITZER L. 1970 (1929): *Études de style*, Paris, Gallimard
- SZONDI P. 1989 (1975): *L'Herméneutique littéraire* (traduction de Mayotte Bollack, préface de Jean Bollack), Paris, Cerf.
- VIPREY J.-M. 1997: *Dynamique du vocabulaire des Fleurs du mal*, Paris, Champion.
- 2000: «Pour un traitement textuel de l'allitération», *Semen n°12*, Paris, Les Belles-Lettres.
- 2002: *Analyses textuelles et hypertextuelles des Fleurs du mal*, Paris, Champion.
- ZUMTHOR P. 1994: Article «philologie» de *l'Encyclopaedia Universalis multimedia Version 10*.