

# Discours, texte et corpus

**Philippe Lacour**

Goethe Universität Frankfurt/Main - Centre Marc Bloch Berlin.  
Allemagne.

Mon propos sera essentiellement *logique* – au sens d'une réflexion sur la manière dont le langage naturel (logos) fait sens et s'articule. Dans un souci topique, je tenterai d'abord d'assigner une place au texte au sein d'une théorie du discours, avant de réfléchir sur la linguistique de corpus, et de montrer qu'elle est régie par les mêmes traits logiques que les autres sciences de la culture, que j'ai cherché à qualifier de : réflexif, herméneutique et clinique. J'essaierai en particulier de préciser la nature casuistique de la clinique, en montrant comment la méthode comparative et le souci descriptif se répondent, dans le cadre d'un nouveau rapport entre théorie et « données » empiriques – ce sera l'occasion de mettre à l'épreuve certaines opinions concernant la linguistique de corpus.

## 1. Une définition du discours

Du discours, je donnerai avec Ricoeur la définition suivante : *quelqu'un dit quelque chose à quelqu'un selon des règles (phonétiques, lexicales, syntaxiques et stylistiques)*, formule qu'on peut expliciter par un certain nombre de polarités : événement et signification, identification singulière et prédication générale, acte propositionnel et acte illocutionnaire, sens et référence, référence à la réalité et référence aux interlocuteurs<sup>1</sup>.

D'abord tout discours se produit comme événement, au sens où quelque chose se passe quand un discours est prononcé ; mais cette occurrence est

---

<sup>1</sup> Cf. « Discours et communication », in *La Communication. Actes du XV<sup>e</sup> Congrès de l'association des sociétés de philosophie de langue française. Université de Montréal, 1971*, Montréal 1973 (repris dans : *Paul Ricoeur*, L'Herne, 2004, S. 51-67) ; « La métaphore et le problème central de l'herméneutique », *Revue philosophique de Louvain*, 1972; et « La fonction herméneutique de la distanciation », in *Du texte à l'action* (Paris, Seuil, 1986). Cette conception est assez proche de celle de Coseriu (cf « L'homme et son langage », in *L'homme et son langage*, Louvain-Paris-Sterling-Virginia, 2001), la notion de « style », empruntée à Granger, ménageant une place fondamentale à la créativité du langage.

éphémère, apparaissante et disparaissante : c'est un événement « évanouissant ». En dépit de cette fugacité, la signification, demeure et peut être identifiée et réidentifiée comme la même ; par quoi le discours est compris.

Ensuite, la signification est portée par la structure spécifique qu'est la proposition (Strawson), qui combine une fonction d'identification singulière, par laquelle nous désignons une chose et une seule (cette homme, cette table) et une fonction de prédication générale, par laquelle nous donnons des caractères ou qualités à quelque chose (l'humanité comme classe, la clarté comme propriété).

En outre, tout discours implique une différence entre sens et référence, pour reprendre la distinction de Frege (Sinn/Bedeutung) entre ce qui est dit par la phrase prise comme un tout et par les mots en tant que parties de la phrase, et ce au sujet de quoi quelque chose est dit. Outre la référence à la réalité, le discours se réfère à son propre locuteur (pronoms personnels, temps verbaux, démonstratifs : tout ce que Hagège nomme le système de l'égophore, et qui organise les éléments du discours autour du sujet de l'énonciation<sup>2</sup>), et à son destinataire.

Enfin, le discours en tant qu'acte peut être considéré du point de vue du contenu de l'acte propositionnel ou de celui de la force de l'acte complet de discours, pour reprendre la distinction d'Austin entre l'acte *de* dire (locutionnaire) et l'acte que l'on fait *en* disant (illocutionnaire).

## 2. Extension textuelle de la notion de discours

Il faut tenir compte de *deux facteurs de complexification* de la discoursivité : la dimension transphrastique (a) et la distance temporelle (b), qui définissent la *textualité*.

(a) D'abord, on ne construit pas le sens phrase par phrase, comme le voudrait la grammaire textuelle (encore moins mot par mot), mais par paquets de phrases. Une dimension holistique apparaît, qui fait obstacle à l'idée d'une compositionnalité du sens, soutenue, de façon plus ou moins explicite et nuancée, par la plupart des philosophes analytiques (Davidson<sup>3</sup>, Descombes<sup>4</sup>) : il y a des effets des phrases les unes sur les autres,

---

<sup>2</sup> Claude Hagège, *L'homme de paroles*, Paris, Gallimard, 1985, p. 328.

<sup>3</sup> Donald Davidson, *Enquêtes sur la vérité et l'interprétation*, Nîmes, Jacqueline Chambon, 1993. Cf également Pascal Engel, *Davidson et la philosophie du langage*, Paris, PUF, 1994, p. 8.

<sup>4</sup> Vincent Descombes, *Le complément de sujet*, Paris, Gallimard, 2004. Descombes combine l'analyse logique d'A. Kenny (sur la polyadicité variable des verbes d'action) et l'approche structurale de la syntaxe de L. Tesnières (sur la valence des verbes) – pp. 67-89. A aucun moment n'est posée la question du transphrastique.

notamment des effets de « cohésion » mis en valeur par Halliday et Hasan, et la signification se construit comme une totalité qui est plus que la somme de ses parties. Corrélativement, ce qui définit d'abord le *texte*, c'est d'être une unité sémantique de rang supérieur à la phrase, donc un terrain spécifique de signification, que Harris appelle « transphrastique »<sup>5</sup>. Le discours commence donc bien avec la phrase (Benveniste), mais celle-ci n'en marque que la limite inférieure.

(b) Ensuite, la dialectique de l'événement et de la signification ouvre au sens, qui survit à la disparition de son support éphémère, une carrière singulière, en l'éloignant de la situation initiale d'énonciation. Le texte est le support fondamental de cette *distanciation* dans la communication, qui éloigne radicalement le discours du dialogue. Aussi convient-il de dissocier, comme le fait Ricoeur, l'écriture et le texte, lequel est d'abord synonyme d'inscription et de médiation par la distance<sup>6</sup>. *La notion de texte est donc une catégorie du discours et non de l'écriture* ; c'est pourquoi elle est liée à d'autres notions qui débordent l'écriture comme telle, comme la notion d'œuvre et celle de lecture. On peut ainsi, par extension, souscrire à la définition du texte par Rastier comme « suite empirique attestée, produite dans une pratique sociale déterminée et fixée sur un support quelconque », où la notion de support relativise l'opposition de l'oral et de l'écrit<sup>7</sup>.

Dire que le texte est une catégorie du discours, c'est immédiatement souligner que le discours ne se réduit pas à la *communication immédiate*, même s'il commence avec elle<sup>8</sup> – celle-ci ne saurait donc constituer la norme de la discursivité (K.O. Apel et J. Habermas). En ce sens, le *texte* constitue un *modèle*, « le paradigme de la distanciation dans la communication »<sup>9</sup>. Le texte introduit un écart, qui modifie le régime de la signification, dans la mesure où le discours se détache de ses trois pôles initiaux: l'intention de l'auteur, la référence ostensive, le destinataire primitif. Ceux-ci ne disparaissent pas, mais se trouvent complexifiés dans ce qu'on peut appeler un « devenir-texte du discours »<sup>10</sup>. A la polysémie déployée par le texte répond – réflexivement – un exercice

---

<sup>5</sup> Pour une généalogie de l'idée, cf J.-M. Adam, *La linguistique textuelle. Introduction à l'analyse textuelle des discours*, Paris, Armand Colin, 2005, avant-propos.

<sup>6</sup> Cf. Ricoeur, « La fonction herméneutique de la distanciation », *art. cit.*, p.114.

<sup>7</sup> Rastier, *Arts et sciences du texte*, Paris, PUF, 2001, p. 21. L'auteur se donne un cadre pluri-sémiotique (le texte étant une « production sémiotique complexe»). On suivra toutefois dans cette étude le fil directeur du langage.

<sup>8</sup> La linguistique doit donc nécessairement prendre en considération les situations de communication différées, comme le souligne Rastier. Cf. *Arts et sciences du texte, op. cit.*, p. 107.

<sup>9</sup> Cf « La fonction herméneutique de la distanciation », *art. cit.*, p. 114.

<sup>10</sup> Ricoeur, « De l'interprétation », in *Du texte à l'action, op. cit.*, p. 35.

d'herméneutique, que Rastier nomme un « parcours interprétatif », dans la lecture (l'interprétation du sens dans la conversation dialoguée n'est qu'une pré-herméneutique<sup>11</sup>). Nier la difficulté corrélatrice de la complexité de la construction de la signification dans la réception à distance serait être victime de l'illusion de ce que Rastier appelle les herméneutiques de la clarté<sup>12</sup>.

La notion de « parcours interprétatif » revêt une double importance. D'abord, elle inverse le sens de la sémiose, au sens où le signe ne préexiste pas comme unité au texte, mais où c'est au contraire le parcours sur un texte qui « discrétise » les unités de sens. Ce renversement herméneutique de la sémiose est beaucoup plus radical et probant que l'audace apparente de Tognini-Bonelli<sup>13</sup>, qui plaide pour un dépassement du lexical (le mot) dans le lexico-grammatical (le « pattern », qui fait intervenir des aspects sémantiques relatifs à l'environnement du mot, comme la valence des verbes par exemple). Par ailleurs, ce renversement, qui va à l'encontre de l'oubli de l'herméneutique par la linguistique, milite en faveur d'une relativisation des catégories de l'agrammatical et de l'inacceptable<sup>14</sup>.

La théorie de Rastier paraît cependant souffrir d'une légère ambiguïté, dans son opposition au paradigme communicationnel, d'une part, et au paradigme de la cognition, d'autre part.

(i) Tout d'abord sa théorie du texte semble manquer d'une théorie du discours, de même que, symétriquement, le paradigme « communicationnel » est une théorie du discours qui manque d'une théorie du texte. Certes, c'est à bon droit que Rastier stigmatise le « déficit herméneutique » du modèle communicationnel, lequel ne tient pas compte « des différences de situation historique et culturelle qui peuvent séparer la production de l'interprétation »<sup>15</sup>. On comprend que, sans s'y opposer, il ambitionne d'inclure ce paradigme dans un ensemble plus vaste, incluant « les situations de communication codifiées, différées, et non nécessairement interpersonnelles »<sup>16</sup>. C'est à cette fin qu'il redéfinit la

---

<sup>11</sup> Cf. Ricoeur « Rhétorique, poétique, herméneutique », *art. cit.*, p. 489-490. Ajoutons que l'herméneutique n'est pas pour Ricoeur (en tout cas après son livre sur Freud de 1965) un exercice de restitution d'un sens caché ou latent, mais une opération de réduction de la polysémie déployée par le niveau transphrastique du discours, sous la forme d'un texte.

<sup>12</sup> Rastier entend par là les « herméneutiques positives qui considèrent la compréhension « normale » comme naturelle et inconditionnée. Cf *Arts et sciences du texte, op. cit.*, pp. 114-5.

<sup>13</sup> E. Tognini-Bonelli, *Corpus Linguistics at Work*, Amsterdam, John Benjamin's Publishing, 2001, ch. 2 et 6.

<sup>14</sup> Sur ce point, les remarques de François Rastier (*Arts et sciences du texte, op. cit.*, p. 105) et de Claude Hagège (*L'homme de paroles, op. cit.*, 1985, p. 342 et 283-4) paraissent convergentes.

<sup>15</sup> F. Rastier, *Arts et sciences du texte, op. cit.*, p. 107.

<sup>16</sup> *Ibid.*, p. 107.

communication comme un cas particulier du triangle de l'échange, moyennant un dédoublement de la dimension de l'allocution. L'adresse est en effet distinguée de la destination : un bénéficiaire absent (il) venant toujours s'ajouter à un destinataire présent (toi, à qui « je » parle) dans la situation de communication interpersonnelle. Toutefois, la démonstration de Rastier n'est pas pleinement convaincante, car il ne donne pour illustrer son propos que l'exemple d'un « écrit », qui échappe par définition à la situation de communication interpersonnelle<sup>17</sup>. Les seuls exemples de destination à l'oral sont en réalité des cas limites (l'acteur de théâtre qui parle à la cantonade, l'enfant qui parle à l'objet transitionnel), dont il paraît difficile de tirer une règle<sup>18</sup>.

En fait, Rastier est conscient de l'existence de deux dimensions de la communication (l'interlocution et le texte), et de la nécessité de les distinguer *et* de les articuler. L'ambiguïté de sa position est donc avant tout lexicale : d'un côté, il souligne que le paradigme herméneutique « englobe » le paradigme communicationnel, de l'autre, il refuse de parler de communication (surtout au sens transcendantal). Cette hésitation entre rapport d'inclusion et d'opposition pourrait être facilement levée, moyennant la distinction d'un sens étroit (immédiat) et large (distant) de la communication. En réalité, on peut parfaitement soutenir que le texte relève de la discursivité, donc de la communication, mais à condition de préciser, comme Ricoeur, que la communication ne se réduit pas à la communication immédiate (dialogue). Cette précision faite, les positions des deux auteurs convergent de façon remarquable : d'une part, les analyses de Rastier sur la destination, la textualité, la zone anthropique distale (de l'absence) et la transmission (culturelle) conservent toute leur pertinence<sup>19</sup> et rejoignent celles du philosophe<sup>20</sup> ; d'autre part, on fait droit à l'insistance de Ricoeur sur l'appartenance du texte au régime de discursivité, donc de la communication (au sens large). Cette appartenance n'exclut pas la spécificité : le texte relève en effet d'une communication très particulière, dont les termes ne sont plus : locuteur/récepteur/référent, mais auteur/lecteur/monde du texte.

(ii) Pour Rastier, le paradigme herméneutique échappe tout autant au paradigme de la cognition qu'à celui de la communication. Autant cette dernière donne la priorité à la pragmatique, autant la cognition insiste sur la

---

<sup>17</sup> *Ibid.*, p. 129.

<sup>18</sup> *Ibid.*, p. 131.

<sup>19</sup> *Ibid.*, p. 131-2. En particulier, Rastier a raison de voir dans la communication à distance le propre du langage humain par rapport aux systèmes de communication animaux.

<sup>20</sup> Sur la culture comme horizon de la discursivité, cf Ricoeur, *Sur la traduction*, Paris, Bayard, 2004, p. 55-6.

sémantique, en lui donnant un sens exclusivement représentationnel<sup>21</sup>. Loin d'être incompatible, ces deux accentuations référentielle (vériconditionnelle) et inférentielle sont parfaitement complémentaires, le risque étant précisément que sémantique logique et pragmatique cognitive se partagent sans reste la question du sens<sup>22</sup>. C'est pour éviter une telle forclusion de l'herméneutique que Rastier cherche à unifier les approches de la signification (référentielle, inférentielle et *différentielle*) par la sémantique, à laquelle il confère une dimension *interprétative*. Reste qu'il glisse peut-être insensiblement d'une désontologisation méthodologique<sup>23</sup> à un anti-référentialisme de fait. L'ambiguïté apparaît clairement lorsqu'on suit Rastier dans sa distinction du sémiotique (dans l'ordre du réel) et du symbolique (dans l'ordre de la connaissance). De fait, la « dé-ontologie » que prône Rastier ne vaut qu'au niveau de la médiation symbolique qui, par définition, ne se soucie pas de la référence. Mais l'insistance pour ajouter, dans l'axe de la représentation, le niveau sémiotique comme intermédiaire entre les niveaux physique et représentationnel, correspond à une forme d'engagement ontologique, comme l'atteste le vocabulaire employé (« couche de l'Être » ou encore « ontologie du sémiotique »<sup>24</sup>).

Chercher à résoudre cette ambiguïté en substituant la praxéologie à l'ontologie ne saurait être pleinement satisfaisant. En donnant parfois l'impression d'exclure la référence de la sémantique, Rastier paraît commettre une erreur symétrique de celle qui consiste à réduire la sémantique à l'ontologie. Sans doute s'agit là d'une ambiguïté similaire à celle qui affectait les rapports du paradigme herméneutique et du paradigme communicationnel, entre intégration et opposition. Là encore, il n'y a pas à opposer radicalement sémantique et référence, mais plutôt à ménager une place logique à la référence *et* à l'interprétation au sein d'une sémantique élargie<sup>25</sup>, qui ménage la spécificité de l'herméneutique. Il est vrai que cela implique de redéfinir l'ontologie. On peut pour cela suivre certains penseurs qui refusent de réduire le réel à l'actuel (l'effectif), en y intégrant également une dimension virtuelle<sup>26</sup>. La textualisation est un mouvement de

---

<sup>21</sup> F. Rastier, « Communication, interprétation, transmission » *art cit.*, p. 121.

<sup>22</sup> Cf. P. Lacour, « L'oubli de la sémantique dans le programme cognitiviste. Réflexions sur l'œuvre de François Rastier », Revue *Texto !*, déc. 2004, [http://www.revue-texto.net/Inedits/Lacour\\_LOubli.html](http://www.revue-texto.net/Inedits/Lacour_LOubli.html).

<sup>23</sup> Sur le refus de réduire le sens d'un texte à l'un des trois « pôles extrinsèques de la signification », en l'occurrence la référence, cf. Rastier, *Arts et sciences du texte*, *op. cit.*, p. 15.

<sup>24</sup> F. Rastier, « Sémantique et recherche cognitive », Paris, PUF, 1991, p. 242.

<sup>25</sup> Hagège donne certaines indications en ce sens, Cf. *L'homme de paroles*, *op. cit.*, pp. 284-295.

<sup>26</sup> Cf. Gilles Deleuze, *Différence et répétition*, Paris, PUF, 1968, et Gilles-Gaston Granger, *Le possible, le probable et le virtuel. Essai sur le non-actuel dans la pensée scientifique*, Paris, Odile Jacob, 1995.

virtualisation du sens discursif, tandis que la lecture correspond au mouvement inverse d'actualisation de la signification. Alors que, à l'oral, le référent est la situation du discours, et tend à se confondre avec une désignation ostensive, avec l'écrit, « le mouvement de la référence vers la monstration se trouve intercepté »<sup>27</sup>. C'est alors, dans ce « suspens où la référence est différée » que le texte, placé en quelque sorte « en l'air » rentre « dans un rapport intertextuel qui constitue la littérature »<sup>28</sup>. Mais cette interception ne vaut pas suppression, puisque la lecture effectue la référence<sup>29</sup>.

On le voit, il n'y a pas à choisir entre ontologie et sémantique interprétative, pas plus d'ailleurs qu'entre interprétation et communication. Une conception *intégrale* de la discursivité se doit d'intégrer les trois paradigmes que Rastier distingue et oppose, tout en ménageant la spécificité logique irréductible de la textualité. C'est d'ailleurs à cette aune d'une discursivité exhaustive qu'il faut interpréter le récent rapprochement de la linguistique du texte de l'analyse du discours<sup>30</sup>.

### 3. Qu'est-ce qu'un corpus ?

La définition du terme fait problème. Contrairement à ce que disent E. Tognini-Bonelli<sup>31</sup> et B. Habert<sup>32</sup>, il n'est pas sûr qu'on puisse caractériser le corpus par sa taille (grand ensemble de mots), puisqu'il peut exister de petits corpus ; et contre Tognini-Bonelli et G. Williams<sup>33</sup> je ne crois pas que le support électronique constitue un critère pertinent – l'ordinateur n'a pas inventé le corpus) ; il ne fait qu'automatiser des opérations intellectuelles que l'homme a réussi à extérioriser. On ne saurait donc totalement souscrire à l'idée de Tognini-Bonelli d'un changement qualitatif intervenu en linguistique sous pression des modifications quantitatives : cette version de l'histoire (la « redécouverte » du corpus) ne vaut que pour ceux qui,

---

<sup>27</sup> Ricoeur, « Qu'est-ce qu'un texte », in *Du texte à l'action*, *op. cit.*, p. 157.

<sup>28</sup> *Ibid.*, p. 157.

<sup>29</sup> *Ibid.*, p. 171-2. Ricoeur parle d'actualisation ou d'effectuation du texte par la lecture, opérations par laquelle le texte se rattache à la théorie du discours.

<sup>30</sup> Jean-Michel Adam, « Autour du concept de *texte*. Pour un dialogue des disciplines de l'analyse de données textuelles », in *JADT 2006* ; et *Linguistique textuelle*, Paris, Armand Colin, 2005.

<sup>31</sup> E. Tognini-Bonelli, *Corpus Linguistics at Work*, Amsterdam, John Benjamin's Publishing, 2001.

<sup>32</sup> Benoît Habert, Adeline Nazarenko, André Salem : *Les linguistiques de corpus*, Armand Colin, 1997.

<sup>33</sup> Geoffrey Williams, dans *La linguistique de corpus*, Rennes, PUR, 2005, soutient également que la linguistique de corpus est constubstantiellement liée à l'informatique.

oublieux de la tradition herméneutique et philologique, avaient négligé la dimension interprétative (et concrète, effective<sup>34</sup>) de leur discipline.

La définition de Sinclair<sup>35</sup> du corpus comme « collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage » offre l'avantage de la souplesse puisqu'elle n'est pas directement indexée sur les textes, comme le remarque Habert. Il me semble toutefois que celle de Rastier est plus pertinente, parce qu'elle s'insère dans un dégradé de différents ensembles. Pour Rastier, le corpus est en effet un « regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications »<sup>36</sup>. Or cette approche présente de multiples. Le premier, c'est celui de partir des productions discursives effectives, qui ne sont pas des mots, ni des phrases, mais des discours – en l'occurrence fixés sur des supports, donc des textes. Ces productions sont effectives au sens où elles correspondent au *dictum* dans son intégralité ; le texte étant la véritable unité de sens.

Le second avantage, c'est qu'une telle approche permet de conserver la souplesse de la définition de Sinclair, mais en l'intégrant dans un dégradé convaincant parce que nuancé : l'archive désigne l'ensemble des documents accessibles, le stock, et ce n'est pas un corpus parce qu'elle ne correspond pas à une recherche déterminée ; le corpus de référence, est l'ensemble des textes sur lequel on va contraster le corpus d'étude ; le corpus d'étude est constitué en fonction d'une problématique, d'une orientation de l'enquête ; et enfin le sous-corpus varie selon les phases de l'étude, et ne contient que les passages pertinents des textes étudiés. Dans un tel dégradé, on peut parfaitement inclure les études de lexiques (les « sacs de mots ») mentionnées par Sinclair<sup>37</sup>, Habert, Tognini-Belli et Mayaffre<sup>38</sup>, ainsi que celle des unités grammaticales, ou grammatico-sémantiques (les « patterns »

---

<sup>34</sup> Pour de nombreux auteurs, la priorité donnée aux productions linguistiques authentiques (tangibles, publiquement observables), qui consacre la supériorité épistémique de la *performance* sur la *compétence*, est définitoire de la démarche de la linguistique de corpus.

<sup>35</sup> J.M. Sinclair, *Preliminary Recommendation on Corpus Typology*, 1996.

<sup>36</sup> François Rastier, « Enjeux épistémologiques de la linguistique de corpus », in Habert et alii (éds.), *Les linguistiques de corpus, op. cit.* Reproduit en ligne : <http://www.revue-texto.net/Corpus/Corpus.html>

<sup>37</sup> J.M. Sinclair, *Corpus Concordance Collocation*, Oxford, Oxford University Press, 1991.

<sup>38</sup> « Rôle et place des corpus en linguistique : réflexions introductives ». *Texto!*, décembre 2005, vol. X, n°4. Disponible sur : [http://www.revue-texto.net/Reperes/Themes/Mayaffre\\_Corpus.html](http://www.revue-texto.net/Reperes/Themes/Mayaffre_Corpus.html).



de Hunston et Francis<sup>39</sup>), mais aussi l'analyse d'autres combinaisons, peut-être plus inattendues : sous-corpus prosodiques, phonologiques, graphiques (notamment pour les manuscrits), axés sur la ponctuation, infra-lexicaux (de syllabes, de lettres) ou supralexicaux (expressions, paragraphes, périodes...). On conserve donc la souplesse de la définition de Sinclair, mais en le situant sur une échelle où le texte tient la place essentielle d'unité de sens linguistiquement produit. Notamment, la taille peut varier : petite (les poèmes présocratiques) ou grande (le théâtre comique dans la littérature occidentale), mais cela ne concerne pas le même niveau (le corpus d'étude dans un cas, le genre dans l'autre).

La définition de Rastier semble également permettre de sortir de l'aporie de la « représentativité » du corpus (*i. e.* de la généralisation des traits linguistiques qu'on y observe à un corpus plus large). Cette difficulté, imputable à la définition du corpus comme « échantillon » du langage (Sinclair), est à peine affronté par Tognini-Belli, qui l'évite plus qu'elle ne le résolve, en considérant la représentativité comme un postulat (révisable en fonction d'indications de fréquence)<sup>40</sup>. Rastier a raison d'y voir un faux problème, dans la mesure où aucun corpus ne peut être dit « représenter » la langue dans son fonctionnement usuel ou historique. Un corpus est toujours constitué réflexivement pour accomplir une certaine action. En ce sens, un corpus est *adéquat* ou non à une tâche par rapport à laquelle on peut déterminer les critères de sa représentativité et de son homogénéité. Cette difficulté rappelle beaucoup le débat qui a agité la communauté des historiens lors de la confrontation avec la « microhistoire ». Revel et Lepetit ont bien diagnostiqué le paralogisme : l'oxymore de « l'exceptionnel normal » (critère improbable de représentativité) importe moins que le principe technique : les jeux d'échelles<sup>41</sup>. Le produit de ces variations de focale est moins caractérisé par sa représentativité (toujours floue, lacunaire) que par sa « validité », dans cet ensemble ainsi découpé, et relativement à une certaine tâche.

La définition de Rastier aide enfin à sortir d'une hésitation épistémologique concernant la *logique* de constitution du corpus. Tognini-Belli oppose terme à terme linguistique *sur* corpus et linguistique *de* corpus.

<sup>39</sup> S. Hunston and G. Francis, *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*, Amsterdam and Philadelphia, Benjamins, 2000.

<sup>40</sup> E. Tognini-Bonelli, *Corpus Linguistics at Work*, *op. cit.*, p. 58 et 185. A son insu, Tognini-Bonelli est très proche de la notion de corpus « réflexif », développée par Damon. Mayaffre ; cf « Les corpus *réflexifs* : entre architextualité et hypertextualité », *Corpus*, 1, nov. 2002, <http://corpus.revues.org/document111.html> ; et « Rôle et place des corpus en linguistique : réflexions introductives », *art. cit.*

<sup>41</sup> Cf Jacques Revel (éd.), *Jeux d'échelles. La micro-analyse à l'expérience*, Paris, Gallimard/Seuil, 1996.

La première serait marquée par un rapport de déduction des exemples à partir des règles de la théorie : le vecteur de la connaissance serait alors orienté vers l'abstrait, dont les réalisations concrètes ne seraient que des instances, des illustrations, des échantillons finalement substituables. La seconde serait marquée par un rapport d'induction de la théorie à partir des exemples effectifs ; le vecteur de la connaissance restant alors orienté vers le concret. Pourtant, sa propre formulation de cette opposition souffre de certaines ambiguïtés : elle dit notamment que l'induction ne saurait être pure (ni mécanique) dans le cas d'une étude « corpus-driven »<sup>42</sup>. Ailleurs, elle parle de processus tantôt linéaire (observation, hypothèse, théorie<sup>43</sup>), tantôt circulaire (aller et retour<sup>44</sup>). Cette hésitation trahit le fait que la logique de la linguistique de corpus n'est ni déductive (pour les raisons bien exposées par Tognini-Bonelli) ni non plus inductive, dans la mesure où l'induction a vocation à dépasser les exemples individuels pour s'élever à une hypothèse générale dont la pertinence n'est plus relative au contexte de ces différentes illustrations – elle prend alors une valeur prédictive (et non simplement rétrodictive), d'où son lien avec le (faux-)problème de la représentativité. On peut qualifier la vocation « concrète » de la linguistique de corpus de souci *clinique*, ce terme désignant la connaissance de l'individuel, le désir de saisie du « ceci »<sup>45</sup>. Mais sa logique est celle de la pensée par cas, où les généralisations, toujours possibles, sont nécessairement « tenues » par le contexte des cas dont elles sont issues<sup>46</sup> : ce contexte « collant », pour lequel Passeron utilise la métaphore de la « force de rappel d'un ressort », c'est celui de la casuistique, par laquelle Passeron et Revel ont bien caractérisé le régime épistémologique spécifique des sciences de la culture<sup>47</sup>.

Dans cette logique « intermédiaire » (ni inductive, ni déductive) description et la comparaison jouent un rôle complémentaire. Au sens strict, les productions linguistiques sont des *individus*, des « ceci ». Or, dans le monde du ceci, rien n'est identique, tout est différent, donc similaire *sous certains aspects* (à déterminer). Parce que ceci n'est pas cela, on peut

---

<sup>42</sup> Tognini-Bonelli, *Corpus Linguistics at Work, op. cit.*, p. 85: l'auteure s'inspire de la position « libérale » de son maître J.R. Firth : "Each scholar makes his own selection and grouping of facts" (*ibid.*, citation).

<sup>43</sup> *Ibid.*, p. 85.

<sup>44</sup> *ibid.*, p. 58 et 61.

<sup>45</sup> Cf G.-G. Granger, *Pensée formelle et sciences de l'homme*, Paris, Aubier, p. 188.

<sup>46</sup> C'est d'ailleurs exactement ce que remarque Tognini-Bonelli, *ibid.*, p. 84.

<sup>47</sup> Même si Passeron et Revel ne disent rien de la linguistique (par méfiance), on rapprochera leur « libéralisme » de celui de Firth: « (...) les raisonnements menés sur des cas ou à partir d'eux (...) ont dû inventer les chemins de leurs généralisations propres » (*Penser par cas*, Paris, EHESS, p. 17).

toujours comparer ceci à cela : non pas assimiler les deux termes, ni refuser tout rapprochement, mais les connecter en précisant toujours le « rapport » sous lequel la comparaison est construite afin, notamment, que les ressemblances ne fassent pas oublier les différences<sup>48</sup>. La singularité des cas « indexe » toute théorie sur des particularités indépassables, de sorte qu'un lien logique profond lie comparaison et description (comparer plus, c'est décrire mieux). La logique de l'analogie est ainsi fondamentalement solidaire de la mise en évidence de nouveaux « observables ». Voilà qui pourrait bien ouvrir à la linguistique et à la littérature comparées de nouveaux possibles.

### Références

- Davidson D.** 1993. *Enquêtes sur la vérité et l'interprétation*. Nîmes : Jacqueline Chambon.
- Deleuze G.** 1968. *Différence et répétition*. Paris : PUF.
- Descombes V.** 2004. *Le complément de sujet*. Paris : Gallimard.
- Granger G.-G.** 1995. *Le possible, le probable et le virtuel. Essai sur le non-actuel dans la pensée scientifique*. Paris : Odile Jacob.
- Habert B., Nazarenko A., Salem, A.** (dir.) 1997. *Les linguistiques de corpus*. Paris : Armand Colin.
- Hagège C.** 1985. *L'homme de paroles*. Paris : Gallimard.
- Hunston S. & Francis G.** 2005. *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam and Philadelphia : Benjamins.
- Rastier F.** 2001. *Arts et sciences du texte*. Paris : PUF.
- Revel J.** 1996. *Jeux d'échelles. La micro-analyse à l'expérience*. Paris : Gallimard/Seuil.
- Ricoeur P.** 1965. *De l'interprétation. Essai sur Freud*. Paris : Seuil.
- Mayaffre D.** 2005. « Rôle et place des corpus en linguistique : réflexions introductives ». *Texto!*, vol. X, n°4. [http://www.revue-texto.net/Reperes/Themes/Mayaffre\\_Corpus.html](http://www.revue-texto.net/Reperes/Themes/Mayaffre_Corpus.html)
- Sinclair J.M.** 1996. *Preliminary Recommendation on Corpus Typology*. <http://www.ilc.cnr.it/EAGLES96/typology/typology.html>
1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.

---

<sup>48</sup> Cf. Ute Heidmann, « Comparatisme et analyse de discours. La comparaison différentielle comme méthode », in J.-M. Adam et U. Heidmann, *Sciences du texte et analyse de discours*, Genève, Slatkine, pp. 99-116.

Philippe Lacour

**Tognini-Bonelli E.** 2001. *Corpus Linguistics at Work*. Amsterdam : John Benjamin's Publishing.

**Williams G.** 2005. *La linguistique de corpus*. Rennes : PUR.