

CORPUS **Corpus**
10 | 2011
Varia

François RASTIER — *La mesure et le grain. Sémantique de corpus*. Paris : Champion, Collection *Lettres numériques*, 2011, 280 pages, 55 €

Damon Mayaffre



Édition électronique

URL : <http://journals.openedition.org/corpus/2146>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 1 novembre 2011

Pagination : 320-325

ISSN : 1638-9808

Référence électronique

Damon Mayaffre, « François RASTIER — *La mesure et le grain. Sémantique de corpus*. Paris : Champion, Collection *Lettres numériques*, 2011, 280 pages, 55 € », *Corpus* [En ligne], 10 | 2011, mis en ligne le 18 juin 2012, consulté le 19 avril 2019. URL : <http://journals.openedition.org/corpus/2146>

Comptes rendus

de suppression de postes, de diminution des moyens accordés à la formation des enseignants. Quant à la solution qui consisterait à employer massivement des locuteurs natifs, non linguistes, pour assurer une formation en langues à moindre coût au nom d'une supposée authenticité, on se doute de ce que les membres du GEPED et proches de Danielle Bailly peuvent en penser.

Nathalie SCHNITZER
Université Nice – Sophia Antipolis

François RASTIER — *La mesure et le grain. Sémantique de corpus*. Paris : Champion, Collection *Lettres numériques*, 2011, 280 pages, 55 €.

Décisif ouvrage que ce dernier opus de François Rastier. Après *Sémantique interprétative* (1987), et dans le fil conducteur de *Arts et sciences du texte* (2001), qui furent des références pour une génération de chercheurs, l'auteur nous propose une réflexion réinventée sur un objet renouvelé.

L'ambition du propos, telle qu'elle s'énonce dans le sous-titre de l'ouvrage, est de théoriser tout autant que pratiquer, une *sémantique de corpus* encore aujourd'hui dans les limbes. Ce projet, que Rastier décline en 3 parties, 9 chapitres et 272 pages, est l'occasion pour l'auteur de formuler sa vision de la linguistique.

Sans prétendre rendre compte de manière exhaustive de la richesse du propos, nous retiendrons seulement de l'ouvrage trois enjeux principaux ; ceux qui sont le plus directement à même d'intéresser les lecteurs de la revue *Corpus* : concevoir l'objet corpus ; instrumenter la sémantique (désormais pratiquée sur corpus) ; définir les tâches de la linguistique de corpus.

Concevoir l'objet corpus

Déjà pressentie dans ses articles au cours de la dernière décennie, la conception de l'objet corpus est présentée par l'auteur comme une question cardinale pour les pratiques linguistiques contemporaines, et, au-delà de la linguistique, pour ces *sciences de la culture* que Rastier appelle de ses vœux pour reconfigurer

Comptes rendus

le panorama épistémologique des Lettres et Sciences humaines et sociales. Trop évidente sans doute pour qu'on s'y attarde outre mesure, cette conception ne semble se comprendre, et s'impose aujourd'hui, que par la révolution numérique. Il y a chez Rastier – à la suite de nos pratiques quotidiennes – une mise en synonyme entre *corpus* et *corpus numériques*, tant et si bien que l'ouvrage apparaît aussi comme un des premiers livres qui affronte la question des *e-humanities* autour de la philologie numérique ou de l'herméneutique numérique. On notera cette mise en équation dès les premières lignes qui problématisent l'ensemble du propos :

La constitution et l'analyse de corpus sont en passe de modifier les pratiques voire les théories en lettres et sciences sociales. Toutes les disciplines ont maintenant affaire à des documents numériques et cela engage pour elles un nouveau rapport à l'empirique. En outre, la numérisation des textes scientifiques eux-mêmes permet un retour réflexif sur leur élaboration et leurs parcours d'interprétation. (p. 12)

Qu'est-ce donc qu'un corpus (numérique) ? Quelle est sa vocation dans nos pratiques (numériques) scientifiques comme sociales ? Telles sont deux des principales questions qui traversent le livre.

Par là, Rastier, dans une démarche différentielle ou contrastive (qu'il préconise d'ailleurs pour la sémantique ; *cf. infra*) distingue *corpus* et *archives*, *corpus* et *document*, *corpus* et *œuvre*, *corpus* et *web*, ou encore de manière problématisée *corpus* et *texte*. Si les textes restent pour Rastier depuis *Arts et Sciences du texte* « les seuls objets empiriques » (p. 24) d'une linguistique adulte, l'auteur affirme désormais sans détour que « tout texte doit être rapporté à un corpus pour être interprété » (p. 25), et plus loin :

...le morphème est l'unité élémentaire, le texte est pour une linguistique évoluée l'unité minimale et le corpus l'ensemble dans lequel cette unité prend son sens. (p. 33)

Comptes rendus

Ou encore :

Un texte ne peut se lire que dans un corpus, qu'il soit implicite, comme en général dans les études littéraires, ou explicite, comme en linguistique de corpus. (p. 64)

Au final, Rastier refuse de « réifier le corpus » (p. 236) et milite pour une « conception critique des corpus » (p. 80). Ceux-ci apparaissent avant tout, de manière dynamique, comme la condition de l'interprétation. Si dans le cadre de la sémantique interprétative, le contexte et la contextualisation sont nécessaires à l'émergence d'un sens non pas donné mais construit dans des parcours de lecture, Rastier semble avoir trouvé dans les corpus le lieu empirique ou la définition pratique du contexte élargi ; contexte ou corpus essentiels parce que non pas seulement « déterminants » mais « constituants » des signes (p. 242)⁹. En ce sens, cet ouvrage nous semble fournir une clef de voûte à une pensée qui depuis longtemps, en référence à l'herméneutique, pose que le global (désormais le corpus) détermine le local (le texte, le *passage*, le mot) et que le contexte (désormais le corpus) informe le signe.

Evidemment, cette importance accordée aux corpus réclame comme rançon des précautions importantes pour leur constitution. Et l'on trouvera notamment plusieurs fois répétée l'exigence d'homogénéité générique des corpus. De manière pratique, Rastier s'applique aussi à distinguer *corpus d'étude* et *corpus de référence*. Le *distinguo* n'est pas anecdotique et dit beaucoup de la méthode différentielle ou de la « sémantique différentielle » (p. 221) qu'il présente à plusieurs reprises : « Le sens est fait de différences, non de références » (p. 64, cf. aussi p. 24, 29, 51). A la suite d'Etienne Brunet, à qui il rend hommage, Rastier développe l'idée simple et désormais possible grâce aux disponibilités numériques, de constituer de vastes corpus de référence (toute la littérature française par exemple telle qu'elle est consignée dans Frantext) pour y faire contraster

⁹ « ... le contexte est constituant et non simplement déterminant : il ne modifie pas des signes déjà donnés, il permet de les instituer comme signes » (p. 242).

un corpus d'étude (tel auteur) afin d'en déterminer les caractéristiques linguistiques et les particularités sémantiques.

Instrumenter les corpus, la sémantique, la linguistique

Réputé habituellement pour sa dimension théorique, le propos de Rastier est dans cet ouvrage aussi et avant tout méthodologique ; quitte ensuite à théoriser (mais aussi à illustrer) les méthodes préconisées. Derrière le titre de l'ouvrage *La mesure et le grain*, l'auteur entend réfléchir à l'articulation entre quantitatif et qualitatif.

C'est d'abord le *grain* ou la granularité des textes que Rastier questionne en s'interrogeant sur la qualité ou la pertinence des « données textuelles ». Loin des seules données prétendument naturelles (comme le mot), Rastier nous sensibilise sur la diversité et la complexité des unités textuelles à traiter ; composantes textuelles multi-variées, multi-niveaux, hétérarchiques (p. 59), pas forcément discrètes, volontiers discontinues, souvent instables. Dans ce cadre, l'apport le plus important est la définition du *passage* qui se caractérise par un *extrait* sur le plan de l'expression et un *fragment* sur le plan du contenu, et représente localement le « lieu privilégié de la sémiosis » textuelle (p. 62). Pourtant l'ambition suprême est ailleurs pour devenir un leitmotiv de l'ouvrage : au-delà de la lexie ou de telle étiquette morpho-syntaxique, les « données » ou composantes textuelles de demain restent à découvrir, et l'objectif de la linguistique de corpus devient précisément de faire émerger de « nouveaux observables » linguistiques « inaccessibles autrement » (p. 13, 19, 20, 50, 211).

C'est ensuite à la *mesure* ou un « déchiffrement » du texte que Rastier se consacre. Ici, on ne peut qu'être frappé par la référence appuyée qu'il fait à la lexicométrie originelle ou à la textométrie d'aujourd'hui. Bien sûr, Rastier balise leur usage afin de ne jamais verser dans le positivisme ou le réductionnisme ; il prend soin de ne pas céder à la fascination des chiffres et conclut par exemple : « bien entendu la qualité l'emporte sur la quantité, et le « grain » sur la mesure » (p. 52).

Cependant de très nombreux exemples textométriques, parfois empruntés à ses collègues (E. Brunet, C. Poudat, E. Bourion, M. Valette, etc.), parfois extraits de ses propres

travaux, viennent montrer tout l'intérêt de la caractérisation quantitative, et ceci sur des corpus riches et variés (littérature, sites web racistes, discours théoriques ou philosophiques). En déclarant que la linguistique de corpus a deux origines, computationnelle et lexicométrique, et en rejetant la première, Rastier place centralement la textométrie au cœur de son dispositif. « Sans préconception du langage » (p. 224), la textométrie donne accès à des descriptions qui à ce stade ne veulent rien dire mais sont autant de stations dans un parcours interprétatif qu'il reste à définir. Et Rastier de mettre en garde contre une représentation arborée ou une analyse factorielle des correspondances : « cela suppose une herméneutique des sorties logicielles » (p. 44).

Définir les tâches scientifiques de la linguistique de corpus

Reste enfin à préciser la vocation épistémologique des pratiques linguistiques sur ces corpus dûment constitués, dûment instrumentés. Reprenant la littérature en linguistique de corpus et quelques grands débats inépuisés, Rastier entend confronter le système à l'usage, l'instance à la performance, et la théorie à l'empirie.

Alors que la linguistique théoriciste (sans corpus) portait, en extrapolant quelques observations sur des exemples souvent forgés, des jugements universels sur le langage, la linguistique de corpus, sans renoncer à l'élaboration théorique, en limite la portée aux corpus étudiés, et sans se satisfaire de la seule démarche déductive, procède par essais et erreurs. (p. 13)

On remarquera en effet la conversion de Rastier à l'expérimental, au pratique, à l'empirique aussi loin que possible des constructions théoriques *a priori*. Car cet ouvrage apparaît avant tout comme un « plaidoyer pour une sémantique applicable » (p. 29), comme un engagement « pour une sémantique instrumentée » (p. 51) ou comme la définition « d'une approche empirique » de l'objet linguistique (p. 144). L'auteur imagine ainsi une ère nouvelle où certains grands débats pourront être enfin tranchés empiriquement. Il en va par exemple du traitement de la doxa (ou idéologie) et des valeurs communes (chapitre 4) qui se stéréotypent dans le discours et que la sémantique doit

Comptes rendus

traiter en premier chef : on « dispose à présent, avec le traitement de corpus numériques, des moyens pratiques pour traiter empiriquement cette question » (p. 246) ou « les méthodes de la linguistique de corpus peuvent transformer le problème des idéologies en question empirique » (p. 122). Il en va aussi, par exemple, de l'épineux problème de la polysémie (et de la néosémie) que Rastier traite dans un chapitre 5 qu'il conclut ainsi :

La question étrangement lancinante de la polysémie peut aussi quitter la sphère spéculative pour devenir un problème empirique de la linguistique de corpus. (p. 159)

Il en va encore de la question de la caractérisation linguistique des discours posée dans les années 1980 et qui peut devenir, elle aussi, aujourd'hui « une question empirique » (p. 163). Il en va enfin de la question des genres littéraires, toujours contestés par les littéraires, et pourtant démontrés empiriquement par la « lexicométrie, textométrie et linguistique de corpus » (p. 238).

Trop modestement nous semble-t-il, à propos de la typologie des textes et de l'incidence du global sur le local (chapitre 3), Rastier écrit :

Ces deux questions seront abordées brièvement de manière théorique, puis les hypothèses formulées seront tranchées de manière empirique. La linguistique de corpus permet en effet, et c'est là son enjeu épistémologique le plus précieux, de valider et d'infirmer des hypothèses. (p. 71)

Heureusement, l'ensemble de l'ouvrage montre qu'à cette vertu expérimentale et probatoire, s'ajoute une puissance heuristique plus importante encore. La linguistique de corpus doit permettre de nourrir l'interprétation et la théorisation ; elle doit susciter de nouvelles hypothèses de travail et proposer de nouveaux parcours de lecture en mettant à jour, grâce à de nouveaux outils, de nouveaux observables linguistiques.

Damon MAYAFFRE
CNRS, Université Nice – Sophia-Antipolis