

3. Corpus et Champs Disciplinaires

Plan

1. Introduction
 2. La question de Corpus
 - a. La notion de Corpus
 - b. Le Corpus en analyse du discours
 3. Du Corpus au texte et retour
 4. Conclusion
- Bibliographie

1. Introduction

L'objectif de ce bilan est d'apprécier l'interdépendance qui solidarise l'approche scientifique et l'objet qu'elle s'est donné.

En matière d'analyse linguistique, la réflexion sur la notion de Corpus fait appel à deux disciplines connexes : l'analyse du discours et la linguistique textuelle. L'analyse est ici conditionnée par la nature et les dimensions du terrain couvert, par les connaissances antérieures et les théories qui le préstructurent.

La réflexion sur la notion de Corpus émane d'une démarche réflexive qui interroge les implications d'une pratique de plus en plus répandue et diversifiée de la recherche linguistique.

2. La question de Corpus

a. La notion de Corpus

François Rastier rappelle que dans le sillage de L'École française d'Analyse du discours, texte et discours sont séparés, et une hiérarchie des objets est posée. L'analyse du discours se donne pour objet de connaissance le discours, et le discours est défini comme le texte mis en rapport avec ses conditions historiques de production, d'où le schéma le plus courant¹ :

DISCOURS = TEXTE + CONDITIONS DE PRODUCTION

TEXTE = DISCOURS – CONDITIONS DE PRODUCTION

Dans la logique de ce schéma, J.-M. Adam² signale que le texte est considéré comme l'objet *abstrait*, obtenu par soustraction, et le discours comme l'objet *concret*, « produit dans une situation déterminée sous l'effet d'un réseau complexe de déterminations extralinguistiques (sociales, idéologiques ...) »³. Pourtant d'autres options se défendent : chez G.-E. Sarfati, les pôles abstrait / concret sont inversés : *L'analyse du discours ne traite pas du texte, ni même de la textualité en soi, pas davantage n'a-t-elle vocation à rendre compte du discours – totalité aussi abstraite qu'idéale – mais bien d'une série de textes particuliers qu'il est permis, par la description, de rapporter à tel ou tel type de discours*⁴.

¹ Cf. les travaux de J.-M. Adam. On se reportera notamment aux *Éléments de linguistique textuelle*, Bruxelles – Liège : Mardaga, 1990, p. 23, ainsi qu'à *Linguistique textuelle. Des genres de discours aux textes*, Paris : Nathan, 1999, p. 39.

² *Idem*

³ C. Fuchs (éd.) (1985), *Aspects de l'ambiguïté et de la paraphrase dans les langues naturelles*, Berne : Peter Lang, citée par J.-M. Adam (1990).

⁴ G.-E. Sarfati (1997), *Éléments d'analyse du discours*, Paris : Nathan, pp. 13-14.

En soulignant le singulier, l'auteur découvre le point névralgique : l'opposition abstrait / concret vacille lorsqu'on passe du singulier au pluriel, et *discours* qui chez les uns renvoie à une production verbale occurrente, en somme du *discours-token*, renvoie chez les autres à du *discours-type*. La même polysémie affecte le mot *texte* : le texte objet matériel, empirique, qu'on peut appeler *énoncé*, et « le texte objet abstrait construit par définition et qui doit être pensé dans le cadre d'une théorie (explicative) de sa structure compositionnelle »⁵

C'est pourquoi, il faut s'entendre sur les référents. Si le Corpus est un objet construit, il rassemble des objets qui eux sont empiriques : à la suite de G.-E. Sarfati, je propose de considérer le texte comme « un ensemble suivi d'énoncés (cohésif et cohérent) » qui constitue « l'objet empirique de l'analyse du discours ». Dans les faits, texte et discours ne sont pas séparables, « ce n'est qu'à travers et à partir de l'analyse des textes que la théorisation de la notion de discours est possible ou, mieux encore, qu'un type de discours est connaissable »⁶.

Le lieu de leur articulation est le genre, lui-même associé à des pratiques sociales⁷. Toutefois la route est longue du générique au spécifique, ce qui a engagé différents plans d'analyse : un point de vue interne, celui du mode d'organisation et des facteurs d'unification du texte donnant matière à la branche appelée *linguistique textuelle*⁸.

Il faut cependant songer à relier les étapes : si le but est d'*interpréter le discours dont relève tel ou tel texte occurrent*, une troisième pratique a sa place, qui prolonge les précédentes, les fédère, et parachève le principe de ramification : il s'agit de la *stylistique*, d'une stylistique explicitement ancrée aujourd'hui dans ce que j'envisage comme un consortium des « sciences du discours ».

b. Le Corpus en analyse du discours

L'analyse du discours favorise deux démarches de constitution de Corpus. L'une, qu'on pourrait appeler « cadrative », part d'un cadre institutionnel et énonciatif et définit le Corpus comme partie prenante de ce cadre.

« *Le Corpus n'y est donc pas considéré pour lui-même, mais en ce qu'il est partie prenante dans une institution reconnue qui « définit pour une aire sociale, économique, géographique ou linguistique donnée les conditions d'exercice de la fonction énonciative »*⁹.

L'autre, qu'on appellerait « contextualisante », part de la matérialité linguistique et cherche à la situer par rapport à un genre ou par rapport à un cadre institutionnel et énonciatif. Cette démarche de contextualisation constitue un élément du processus d'interprétation des formes.

Ces approches du Corpus ne sont pas antagoniques, elles se rejoignent d'ailleurs au niveau de la contextualisation par rapport à un cadre institutionnel et énonciatif. Nous verrons plus bas comment cette articulation s'est opérée dans notre étude. Il est à noter que, suite à la diffusion des travaux de Bakhtine, le cadre générique, ressenti comme plus opératoire, tend à se substituer au cadre institutionnel. En effet, le genre, produit d'une sphère sociale de l'activité humaine (Bakhtine 1984 : 265), constitue un cadre de production et d'interprétation des faits langagiers. C'est alors l'unité générique qui garantit l'unité du Corpus tout en permettant la comparaison (voir Corpus 1 ci-dessous).

⁵ J.-M. Adam (1992), *Les textes : types et prototypes. Récit, description, argumentation, explication et dialogue*, Paris : Nathan, p. 15.

⁶ *Ibid.*, p. 17.

⁷ F. Rastier souligne à juste titre ce lien (*Sens et textualité*, Paris : Hachette, 1989, p. 40).

⁸ La distinction et la corrélation de ces niveaux a été circonscrite et schématisée dans les travaux de J.-M. Adam (1990, 1992).

⁹ D. Maingueneau 1991 : 17, citation de Foucault 1969 : 153.

On insistera sur le fait que le Corpus ne tend pas à s'intégrer à un Corpus de référence permettant de décrire la langue fonctionnelle (Rastier 2007) mais revendique une contingence contextuelle en tant qu'échantillon d'une pratique langagière.

3. Du Corpus au texte et retour

Compte tenu de notre choix d'utiliser un Corpus en tant qu'« outil de compréhension », la constitution d'un « architexte » par la mise en série de plusieurs écrits de signalement et leur traitement lexicostatistique permet de mieux cerner les particularités de ces écrits. Ainsi conçu et traité, le Corpus devient un contexte interprétatif des phénomènes langagiers, tout en ouvrant sur d'autres lieux d'observations (genre, corpora contrastifs, etc.).

Pour notre projet, nous envisageons de faire appel au traitement informatique (logiciels d'analyses lexico- statistiques) comme outil de configuration de l'objet d'analyse.

Ces outils vont nous permettre, en tant que chercheur et analyste du discours d'explorer des Corpus volumineux, en donnant une vue en surplomb des faits lexicaux et textuels, grâce aux dictionnaires de fréquences, aux listes de segments répétés, aux fréquences relatives, aux cooccurrences et aux contextes élargis de formes présélectionnées.

De même, l'analyse lexico- statistique peut aider à une compréhension analytique du texte grâce à la mise en série. Pour analyser les rapports de signalement, nous numérisons des textes médiatiques, que nous souhaitons soumettre à une telle exploitation du texte. D'une part, l'analyse lexicométrique va permettre d'accélérer la reconnaissance et l'apprentissage d'un genre nouveau, en dégagant des régularités d'ensemble. D'autre part, elle va favoriser la prise en compte de l'intertexte, via l'étude des segments répétés.

Cependant l'on sait que l'outil configure l'objet : ainsi, l'utilisation de logiciels pour l'exploitation des différents textes numérisés induit un certain nombre de préconstruits d'analyse qui reformulent la question de la co(n)textualisation, en la posant non plus en termes de cadre interprétatif *a posteriori*, mais en termes de conditions interprétatives *a priori*. Les opérations interprétatives préalables (cf. Viprey 2006) ciblent des objets d'analyse prédéterminés et souvent constitués en listes closes. Car le balisage des Corpus, tout comme le choix des formes linguistiques dont on observe les cooccurrences, relèvent déjà d'une opération interprétative

Conclusion

Le Corpus contribue directement à la construction de l'objet linguistique. De fait, ce que l'on observe, ce sont des ajustements successifs entre l'objet et le lieu de son observation. On insistera aussi sur le fait que l'analyse lexicométrique, de par sa vision en surplomb, contribue à transformer le Corpus en contexte interprétatif, grâce aux régularités génériques qu'elle met en évidence et qu'il devient ensuite possible d'appliquer à l'analyse d'un texte afin d'en évaluer la cohérence et de dégager les principes sur lesquels cette dernière prend appui.

Cependant, si le Corpus peut se constituer en outil d'interprétation d'un texte, il n'épuise pas pour autant le contexte, qui reste une donnée ouverte. De même, le fait que le discours construise son propre hors-discours à travers une série de formes linguistiques garantit un garde-fou assez fiable contre une contextualisation aléatoire du Corpus.

Bibliographie

- J.-M. Adam. N(1990), *Éléments de linguistique textuelle*, Bruxelles – Liège : Mardaga.
- J.-M. Adam (1992), *Les textes : types et prototypes. Récit, description, argumentation, explication et dialogue*, Paris : Nathan.
- J.-M. Adam. N(1999), *Linguistique textuelle. Des genres de discours aux textes*, Paris : Nathan.
- C. Fuchs (éd.) (1985), *Aspects de l'ambiguïté et de la paraphrase dans les langues naturelles*, Berne : Peter Lang, citée par J.-M. Adam (1990).
- F. Rastier (1989), *Sens et textualité*, Paris : Hachette.
- G.-E. Sarfati (1997), *Éléments d'analyse du discours*, Paris : Nathan.