

François Rastier

Directeur de recherche
Ertim-INaLCO

La linguistique de corpus permet-elle d'échapper au principe de plaisir ?

Résumé. — Cet aide-mémoire rappelle le caractère expérimental de la linguistique de corpus, en soulignant les enjeux épistémologiques, puis en présente des exemples sur un corpus romanesque.

Mots-clé : Expérimentation, corpus, doxa, objectivation, instances, performances.

1. Renouveler les problématiques par l'analyse des corpus

À la classique dualité entre induction et déduction dans les disciplines d'observation, le renouvellement méthodologique favorisé par les corpus numériques engage à substituer le cycle suivant : (i) analyse de la tâche et production des hypothèses ; (ii) constitution d'une archive et sélection d'un corpus de référence ; (iii) élaboration des corpus de travail ; (iv) traitement instrumenté de ces corpus, en contrastant corpus de travail et corpus de référence ; (v) interprétation des résultats et retour aux sources textuelles pour valider l'interprétation. La puissance propre de ce dispositif heuristique permet de faire émerger de *nouveaux observables* inaccessibles autrement : par exemple, la phonostylistique, jadis condamnée à l'intuition, se voit à présent pourvue de moyens d'investigation par des statistiques sur corpus phonétisés. En outre, l'utilisation d'une instrumentation scientifique (analyseurs, étiqueteurs, etc.) participe du processus d'objectivation : les objets culturels ont beau dépendre de leurs conditions d'élaboration et d'interprétation, les valeurs qu'ils concrétisent peuvent cependant être objectivées comme des faits.

La linguistique de corpus pourvoit ainsi la linguistique d'un domaine où elle peut élaborer des instruments et définir une méthode expérimentale propre : elle ouvre aussi des champs d'application nouveaux et engage un mode spécifique d'articulation entre théorie et pratique. Alors que la linguistique théoricienne (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.

En 1999, Noam Chomsky, auteur d'une grammaire universelle, déclarait que la linguistique de corpus n'existait pas, alors même qu'elle était déjà en plein essor : il signalait ainsi qu'elle restait inconcevable pour la linguistique de fauteuil et qu'une rupture épistémologique était consommée. Cette rupture jouit d'une portée générale : en bref, la recherche part d'une diversité constatée, l'unifie dans le point de vue qui préside à la collection du corpus, éprouve enfin son objectivité par l'investigation instrumentée. Ordinairement, la régularité des observables sera portée au crédit du système, la diversité irréductible sera imputée à la contingence du corpus. Toutefois, l'opposition sommaire entre l'unité totalisante et l'irrégularité accidentelle peut sans doute être dépassée dans la description des normes, dont seules les plus générales, parmi l'ensemble des corpus étudiés, seront considérées comme propres à la langue.

Les corpus et l'espace des normes. — Sans prétendre tirer un bilan prématuré, il semble que la situation nouvelle de la linguistique impose une reconception de la dualité entre linguistique de la langue et linguistique de la parole, qu'il est de tradition d'opposer, tant chez Bally que chez Benveniste, tant en linguistique de l'énonciation qu'en pragmatique, alors que chez Saussure elles sont parfaitement complémentaires.

On a trop souvent réduit les langues à des dictionnaires et des grammaires, voire à des syntaxes. Il faut cependant tenir compte, outre du *système*, des *corpus* (corpus de travail et corpus de référence), de l'*archive* (de la langue historique), enfin des *pratiques* sociales où s'effectuent les activités linguistiques. Pour l'essentiel, une

langue repose sur la dualité entre un *système* (condition nécessaire mais non suffisante pour produire et interpréter des textes) et des *corpus* de textes écrits ou oraux¹.

Non contradictoire, la dualité dynamique entre corpus et système constitue la langue dans son histoire. Aussi ne saurait-on assimiler la *langue historique* à la *langue fonctionnelle* (celle qui fonctionne ici et maintenant) en négligeant que la langue historique détermine la langue fonctionnelle dans ses structures et ses contenus. Le corpus de référence sert de médiation entre la langue historique et la langue fonctionnelle, et les textes qui n'appartiennent plus qu'à la langue historique entrent dans l'*archive*. Soit :

<i>Système(s)</i>	<i>Corpus</i>
Langue fonctionnelle	Corpus de référence
Langue historique	Archive

Tableau 1 : Instances du système et types de corpus

En évoquant les corpus et non les signes, nous soulignons que la langue n'est pas un système de signes – comme le serait un code ; Saussure, à qui l'on prête cette définition, ne l'a d'ailleurs jamais formulée. Un signe au demeurant n'a pas de définition intrinsèque : il n'est qu'un *passage*, certes réduit, d'un ou plusieurs textes auxquels il renvoie.

En première approximation, une langue est faite d'un corpus de textes oraux ou écrits et d'un système. Le système reconstitué par les linguistes est une hypothèse rationnelle formulée à partir des régularités observées dans le corpus. Entre le corpus et le système, les normes assurent un rôle de médiation : ancrées dans les pratiques sociales, les normes de discours, de genre et de style témoignent de l'incidence des pratiques sociales sur les textes qui en relèvent². Pour éviter la fausse antinomie entre la langue en tant que système de formes et la langue comme produit d'une culture – qui se traduit dans l'enseignement par la distinction entre « cours de grammaire » et « cours de civilisation » – il paraît préférable de considérer que le système comprend des *règles* et des *normes* diversement impératives. Par exemple, les règles de la ballade française diffèrent de celles de la ballade anglaise et relèvent du système des normes de la langue française.

Les règles et les normes ne diffèrent sans doute que par leur régime d'évolution diachronique. On sait que les mots (lexies, puis morphèmes) sont issus du figement et de l'érosion de syntagmes ; ce qui vaut pour ces unités linguistiques vaut sans doute pour les règles qui norment leurs relations et les constituent ainsi en unités : les règles sont vraisemblablement des normes discursives invétérées.

En synchronie, toute règle voisine avec des normes qui accompagnent voire conditionnent son application : sans elles, par exemple, on ne pourrait arrêter des enchâssements indéfiniment récursifs mais grammaticalement corrects. On ne peut donc juger de la grammaticalité d'une phrase que si l'on connaît le discours, le genre et le texte où elle est prélevée — outre évidemment la datation et le lieu d'origine de ce texte. Bien qu'élémentaire, cette observation frappe d'inanité les discussions sur l'agrammaticalité et l'asémantisme qui surgissent d'elles-mêmes dès que l'on accepte de discuter de phrases non attestées ou hors contexte.

Ainsi, à la différence de celui d'un langage formel, le système d'une langue est-il en fait pluriel et se décline en régimes structurels différents selon les niveaux et paliers d'analyse. Ses domaines d'organisation locaux ou régionaux ne sont pas unifiés dans une hiérarchie attestant l'existence d'un système unique et homogène, comme en témoigne au demeurant l'évolution continue des langues qui trouvent dans leur hétérogénéité systémique le moteur interne de leur changement perpétuel par perturbations et ajustements.

Non moins plurielles que les instances, les performances se spécifient *a minima* par la distinction entre corpus (de travail et de référence) et archive³. À la grande diversité des pratiques sociales correspond celle des corpus produits en leur sein. Soit, schématiquement :

¹ Dans le corpus d'une langue, les *œuvres* tiennent une place particulière parce qu'elles sont hautement valorisées : par exemple, l'italien est la langue de Dante au sens où son œuvre demeure le parangon historique qui a présidé à la formation de la langue italienne en tant que langue de culture.

² Un texte en effet ne peut pas être produit par un système, comme l'a montré l'échec de la grammaire générative appliquée à des systèmes de génération automatique de phrases et *a fortiori* de textes.

³ Le *corpus de travail* du linguiste n'est qu'une partie du corpus de référence défini par l'ensemble des textes accessibles dans l'empan spatio-temporel considéré. L'ensemble des performances linguistiques non recueillies sur support constitue le *corpus virtuel* de la langue : il garde une incidence, car toute performance modifie peu ou prou les instances normatives qui lui sont associées (règles et/ou normes).

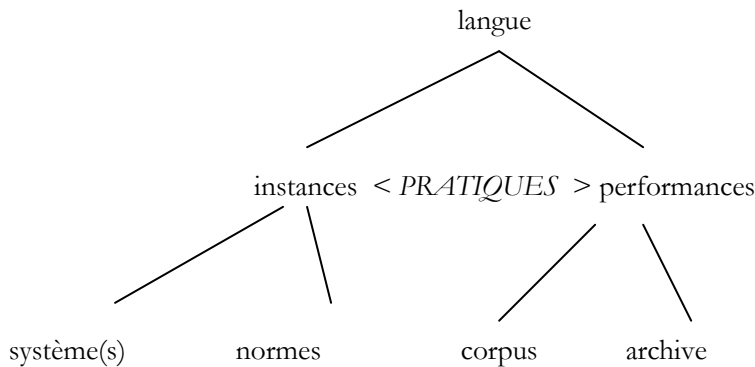


Figure : Instances et performances

Mise à part la mention de la *langue*, le schéma ci-dessus jouit d'une grande généralité et peut être transposé à des sémiotiques non verbales complexes, comme l'iconologie, par exemple. Nous l'avons d'ailleurs utilisé pour un système d'aide à l'indexation de photographies.

La généralité de ce modèle pourrait lui conférer une portée épistémologique et méthodologique. La dualité entre *langue* et *parole* chez Saussure est un cas particulier du rapport entre instances et performances. Au plan méthodologique, la flèche qui va des performances aux instances symbolise l'extraction de régularités ; et la flèche inverse symbolise la caractérisation de singularités, les deux processus restant interdépendants.

La linguistique au demeurant n'a aucune exclusivité épistémologique dans la réflexion sur les corpus : l'ensemble des sciences sociales et des disciplines littéraires se doivent d'élaborer à leur propos une réflexion coordonnée en gardant leurs objectifs spécifiques. Elles gagnent à des échanges d'expériences, loin d'une interdisciplinarité fusionnelle d'ailleurs illusoire.

Beaucoup cependant reste à faire pour convaincre de la nécessité de travailler sur corpus. La technicité, le détour instrumental, la notion même de méthode expérimentale, inquiètent certains ; l'attachement à la recherche sans sanctions empiriques, parfois même dans des disciplines littéraires la répugnance à l'égard de toute objectivation censée porter atteinte à la subjectivité souveraine des auteurs et des lecteurs, tout cela conduit certains à considérer l'étude des corpus comme un leurre⁴.

Ils formulent une objection récurrente : on ne trouve jamais que ce que l'on cherche. Soit ils regrettent par là que l'on vérifie l'intuition sans songer qu'il est parfois difficile de prouver des évidences, ni que cela fait partie de l'ingrate mission des sciences ; soit encore ils estiment qu'on trouve toujours quelque chose : c'est faux, car des résultats bruités peuvent inviter au silence.

De fait, on ne trouve pas toujours ce que l'on cherche, mais souvent autre chose que l'on ne cherchait pas : de nouveaux observables. Certes, on ne trouve trop souvent que ce que l'on sait voir et l'on reste dépendant d'un état de l'art et des problématiques routinières de la « science normale » ; une démarche critique permet cependant de les dépasser ensemble.

2. Applications

La linguistique de corpus peut mettre en évidence de nouveaux observables imprévus. Ainsi, des différences dans la catégorie du *nombre* peuvent opposer des discours. Par exemple, *femme* au singulier est commun en littérature, mais rare dans le discours politique où *femmes* prédomine. Pour en avoir le cœur net, nous avons comparé la répartition de ces deux formes dans les discours des premiers secrétaires aux Congrès du PCF et dans les œuvres complètes de divers écrivains. Le mot *femme* n'apparaît guère au singulier dans les discours

⁴ « Dans les programmes nationaux pour les SHS [sciences humaines et sociales], l'accent est souvent mis sur les « terrains », les « corpus » et autres « archives ». Certes, on connaît leur importance pour les SHS, mais on peut douter qu'il s'agisse de priorités scientifiques ou sociétales. On n'imagine pas la chimie des matériaux construire un programme de recherche sur les meilleurs gisements de matières premières ou sur les fournisseurs les plus efficaces en poudres ou autres produits de base ». Fontanille, J. (2006) Supplément d'âme ?, *Vie de la recherche scientifique*, 365, pp. 16-17, ici p. 17. Soit, mais les corpus, nécessairement élaborés par leur constitution même, ne sont pas des matières premières ; la chimie des matériaux et la besogneuse alchimie des corpus n'ont d'ailleurs pas le même lustre.

communistes, alors que quelques sondages en littérature laissent apparaître une prééminence numérique du singulier sur le pluriel (Balzac : 7989 occurrences de *femme* contre 2746 de *femmes* ; Baudelaire 93/65 ; Éluard 542/168, Gracq : 197/100⁵), les textes politiques inversent la proportion (24/142).

Retenons que le discours politique privilégiera les femmes (ou la femme en général), le discours littéraire la femme exemplaire ou singulière. On pourrait, non sans malice, trouver une image de cette disparité chez un poète communiste comme Éluard : le pluriel l'emporte sur le singulier dans les périodes militantes (en 1937 la distance entre les deux formes s'établit à -1. 8/+4.9), cette proportion s'inversant avec la période de coexistence pacifique (1951 : 4.9/-3.5), sans doute plus propre au lyrisme.

Étudions à présent quelques phénomènes de figement qui relèvent eux aussi de la doxa.

Phénomènes de canonicité. — L'incidence des normes de la doxa se manifeste très simplement par les phénomènes de canonicité. On peut dire que le système numéral du français appartient au système de la langue, mais de fait les chiffres canoniques, comme dix ou trente-six, sont beaucoup plus fréquents que neuf ou trente-sept : cela dépend de normes qui restent à décrire. Par exemple, dans le corpus roman 1830-1970 de la banque Frantext, qui compte environ 350 œuvres, on trouve seulement 5 sortes de fractions de seconde, et, sur 241 occurrences du nom *seconde*, 18 nombres de secondes seulement (cf. Deza, 1999, p. 227, et annexe 1 à la fin de ce chapitre).

On pourrait objecter que la canonicité des nombres n'appartient pas à la doxa, et semble fort loin de la conception marxiste de l'idéologie. Soit, mais en tant que système de normes, l'idéologie ne se limite pas à refléter des rapports humains qu'il serait « politiquement correct » de résumer à des relations de domination. Par exemple, l'idéologie trifonctionnelle indo-européenne articule des relations entre castes, mais également une division du cosmos en trois mondes, sans que l'on puisse dire que la division en castes est la cause de la division en mondes.

Prenons cependant un autre exemple, celui des âges des personnages, qui correspond mieux à la notion ordinaire de la doxa comme système de préjugés, et illustre également le phénomène de la canonicité. En lisant cette phrase de *La Cousine Bette*, « à quarante-sept ans passés la baronne Hulot pouvait encore être préférée à sa fille par les amateurs de couchers de soleil », j'avais formulé l'hypothèse que certains âges n'apparaissent pas dans les romans. Elle a été vérifiée par Deza (1999) dans le corpus Roman de Frantext (1830-1970). En utilisant des patrons d'extraction automatique, elle a relevé 4488 mentions d'âge. On en compte 2650 pour les personnages masculins (59 %) et 1838 (41%) pour les personnages féminins⁶. Certains âges n'apparaissent pas : 41 ans pour les femmes (en revanche 40 ans est un âge canonique)⁷, 49 ans pour les hommes (en revanche 50 ans est un âge canonique), 71 ans ou encore 92 ans ; d'autres sont surreprésentés, par exemple 15, 18 et 20 ans pour les deux sexes ; 16 ans pour les personnages féminins (Deza, 1999, p. 235). Dans le roman français, on a presque toujours vingt ans... Nathalie Deza a pu ainsi, ce n'était pas si simple, vérifier ce qui peut paraître évident, comme le caractère canonique des nombres de secondes ou des âges des personnages romanesques. Sous une apparente banalité, les résultats transforment des intuitions en faits établis qu'une sémantique de la doxa se doit d'interpréter.

Antonymies. — Comme la doxa est affaire de valorisations et que les valeurs s'opposent en premier lieu par paires, étudions quelques antonymes.

Amour et argent. — Depuis le Romantisme, le roman a massivement opposé l'amour et l'argent. Dans l'ensemble des romans de Balzac, nous trouvons une corrélation négative de - 0,42 entre les deux termes (ou plus exactement formes : *amour* et *argent*). En somme, plus un roman parle d'amour, par exemple *Le lys dans la vallée*, la *Duchesse de Langeais* et les *Mémoires de deux jeunes mariées*, moins il parle d'argent⁸.

⁵ Ces effectifs se réduiraient de façon inquiétante si l'on écartait des *Œuvres complètes* de Gracq la traduction de la *Penthésilée* de Kleist : les femmes abondent évidemment dans cette histoire d'amazones. Dès qu'on tient compte des genres au sein des recueils d'œuvres complètes, des disparités apparaissent. Hors des romans et des poèmes, les femmes manquent au singulier comme au pluriel ; par exemple dans les *Paradis artificiels* de Baudelaire, ou dans les lettres du Harrar, alors qu'elles abondent dans les *Poésies* de Rimbaud.

⁶ La disparité entre le nombre de mentions d'âge pour les hommes et pour les femmes peut être due au fait que les personnages masculins sont plus nombreux, ce qui reste à établir, ou plutôt à ce que les mentions d'âge, par courtoisie peut-être, restent moins nombreuses pour les femmes. Toutefois, quatre âges tendres, sur les quelque 110 représentés, comptent plus d'occurrences féminines (15, 16, 18 et 19 ans), on peut en conclure qu'après vingt ans l'âge des femmes intéresse moins les narrateurs et leurs lecteurs implicites.

⁷ Il y a là, comme souvent en perception sémantique, un phénomène d'*imbibition latérale* : les femmes de 41 ans sont « absorbées » par les femmes de 40 ans, les hommes de 49 ans par ceux de 50 ans.

⁸ Les contrastes d'effectifs sont tels que l'on peut négliger les polysémies. Bien entendu, *argent* est polysémique, comme le montrent les contextes où l'argent est tantôt une couleur, un métal, du numéraire. Dans le roman, la dernière acception domine, à l'inverse de la poésie.

Si nous cherchons à présent les mêmes corrélations dans l'œuvre de Rimbaud, nous obtenons la plus grande opposition entre les *Poésies*, de prime jeunesse, où *amour* est relativement très fréquent (coefficient 8,7) et *argent* rare (-1,7), et d'autre part les dernières lettres, où la corrélation est inverse (-10,9 pour *amour* et +5,4 pour *argent*), car les trafiquants d'armes souffrent sans doute d'un déficit d'*amour*.

Cependant, cette opposition n'a rien de propre au discours littéraire. Dans le discours religieux, par exemple, on la trouve chez Montalembert : « Mais ce n'est pas par des présents ni avec de l'argent que la jeune princesse pouvait satisfaire son amour pour les pauvres du Christ » ; de même, chez Monod, « une heure où l'éclat de trente pièces d'argent effaça aux yeux de l'apôtre celui des paroles, des œuvres, de la sainteté et de l'amour de Jésus ».

Dans le discours historique, on la retrouve chez R. de Vertot : « le petit peuple reprit cœur, et quoiqu'il n'y eût à Rome ni hommes, ni armes, ni argent, on trouva tout cela dans cet amour pour la république qui faisoit le véritable caractère d'un romain » ; voire en philosophie chez l'abbé de Mably : « La soif de l'argent qui nous dévore a étouffé l'amour de la patrie. »

Cette opposition reste vivace dans le discours politique contemporain. Prenons par exemple pour corpus le texte des discours des premiers secrétaires aux congrès du Parti communiste français. Au VIII^e congrès, chez Thorez, l'amour (+3,9) l'emporte sur l'argent (+0,2), alors qu'au XXI^e congrès, qui prélude à l'Union de la Gauche, l'argent (+4,8) l'emporte sur l'amour (-4). Est-ce à dire qu'à l'époque du Front Populaire ait prévalu un révolutionnarisme du cœur, ou que le XXI^e congrès soit le Harrar du PCF ? Ce serait persifler ; il ne s'agit d'ailleurs sans doute point du même amour que dans les romans, si l'argent reste peut-être le même partout.

L'antonymie entre *amour* et *argent* dans le discours communiste de l'entre-deux-guerres traduit peut-être ce que Staline nommait le *romantisme révolutionnaire* ; mais nous ne prétendons pas qu'il dérive directement du romantisme littéraire. On peut parler ici d'une opposition transdiscursive. Certes, dans les différents discours, les formes *amour* et *argent*, reçoivent différentes acceptions, et manifestent donc différents sémèmes. Leurs traits communs se situent donc à un niveau inférieur, microsémantique, avec une opposition comme /expansion/vs/restriction/. Leur ubiquité en fait des catégories prégnantes, historiquement très stables. Déjà, dans le *Banquet* de Platon (§ 203), Éros est le fils de Pénia (la racine est la même que *pénurie*) ; jusqu'à nos jours, jusqu'à Amélie Poulain, on imagine que les jouvenceaux vivent d'amour et d'eau fraîche – sans même recourir aux cartes de paiement.

Des maris et des amants. — Dans les *Confessions*, Rousseau formule cette réflexion éclairante sur la « possession » d'une dame : « Je ne puis pas ajouter : *Auctius atque Di melius fecere* ; mais n'importe, il ne m'en fallait pas davantage ; il ne m'en fallait pas même la propriété : c'était assez pour moi de la jouissance ; et il y a longtemps que j'ai dit et senti que le propriétaire et le possesseur sont souvent deux personnes très différentes, même en laissant à part les maris et les amants » (livre V, p. 72). Or, voici que Proudhon, dans *Qu'est-ce que la propriété ?*, essai politique qui relève évidemment d'un autre discours et d'un autre genre, écrit : « Si j'ose me servir de cette comparaison, un amant est possesseur, un mari est propriétaire » (1840, p. 157)⁹. Nous obtenons ainsi une proportion à quatre termes qui semble assez générale. Poursuivant la recherche d'homologies attestées dans divers genres et discours (politique, philosophique, littéraire, etc., dont nous donnons une liste dans l'annexe 2)¹⁰, nous sommes parvenus à un tableau d'oppositions, en notant simplement les oppositions associées, par exemple dans *Les mémoires de deux jeunes mariées*, Louise de Chaulieu écrit à Renée de Maucombe : « L'homme qui nous parle est l'amant, l'homme qui ne nous parle plus est le mari. » (p. 230). Soit :

⁹ Tout cela est bien bourgeois, et il fut un temps, sous Henri II, à l'époque où se déroule *La Princesse de Clèves*, où les maris pouvaient être des amants ; Madame de La Fayette écrivait du moins de M. de Clèves : « pour estre son mari, il ne laissa pas d'estre son amant, parce qu'il avoit toujours quelque chose à souhaiter au delà de sa possession » (1678, p. 33). Non seulement la Princesse avait un amant, mais c'était son mari et le ressort principal du romanesque ultérieur est ainsi mis en pièces avec toutes les lectures romantiques du chef-d'œuvre de Mme de La Fayette.

¹⁰ Nous avons simplement sélectionné dans la banque Frantext les phrases contenant les formes *amant* et *mari*.

MARI	AMANT
propriétaire	possesseur
refusé	comblé
corps	âme
public	privé
argent	gloire
tranquillité	bonheur
puissant	soumis
présent	absent
bon sens	[folie]
trompé	jaloux
prosaïque	romanesque
qui ne parle plus	qui parle
vivoter	vivre
liards	pierreries
province	Paris
abaïsser	élever
affaires	passion
popote	restaurant
objet de complaisances	objet de vivacités
droite	gauche

Tableau 2 : Des maris et des amants

De faible poids statistique, ces résultats ne doivent rien à la lexicométrie et nous en restons ici à l'analyse qualitative. Ce qui semble ici caractéristique de la doxa, c'est la *corrélation entre isotopies* : entre les diverses isotopies, économique (*propriétaire* vs *possesseur*, *liard* vs *pierreries*, etc.), géographique (*province* vs *Paris*), culinaire (*popote* vs *restaurant*), etc., de multiples correspondances peuvent s'établir : le mari rabaisse le corps, l'amant élève l'âme, etc. Ces homologations semblent caractéristiques du mythe — dont dérive sans doute le discours littéraire. En revanche, le discours scientifique n'admet pas la pluralité des isotopies, puisqu'il institue un champ d'objectivité en domaine sémantique. Il admet encore moins leur corrélation et c'est pourquoi l'on discute tant du statut de la métaphore au sein du discours scientifique (voir naguère « l'affaire Sokal ») : la corrélation entre domaines sémantiques qu'elle établit révélerait une « dérive » mythifiante.

Le travail et l'œuvre, le mariage et l'adultère. — Pour éclairer ces phénomènes d'homologation, poursuivons donc notre enquête en prenant l'exemple d'une corrélation entre paires d'antonymes. L'opposition entre le travail et l'œuvre remonte sans doute à l'opposition platonicienne entre *prattein* et *poïen*, la *praxis* et la *poiesis*. À l'époque moderne, toutes les théories romantiques de la création opposeront le poète au fabricant ou faiseur. Dans *Condition de l'homme moderne*, Hannah Arendt porte cette opposition à un plan métaphysique en opposant le travail (*labour*) et l'œuvre (*work*). Cela rappelle fort la polarité romantique, d'abord hégélienne puis marxiste, entre le travail aliéné et la possession ; voire, sans persifler, l'opposition romanesque (bovaryenne) entre le mariage et l'adultère. Voici du moins, en littérature française, quelques indices qui justifient cette analogie.

Le travail s'oppose à l'adultère. — Dans *Pierrot mon ami*, Queneau écrit : « allons au travail, dit Léonie en retirant brusquement sa main de l'étreinte adultère. » (1942, p. 45).

L'œuvre s'oppose au mariage. — Montherlant, dans *Pitié pour les femmes*, formule cette hypothèse panique : « Supposé que ce mariage se fit, il était fatal qu'un jour vînt où il lui faudrait divorcer, à la fois pour sauver son œuvre, et sauver son âme » (1936, p. 1130). Il y revient dans *Le démon du bien* : « Après quelque temps, il voit, d'évidence, que le mariage nuit à ce qu'il appelle son œuvre » (1937, p. 1272). Jean Dutourd renchérit sur un autre ton : « Il faut dire ce qui est, si irritant que cela soit pour un célibataire et un philosophe du célibat : Boulard, s'il a manqué son œuvre, a réussi son mariage. » (*Pluche ou l'amour de l'art*, 1967, p. 141).

Le travail et le mariage se complètent. — Dans la première conversation de la troisième partie de *La Prétieuse*, l'abbé de Pure remarquait : « comme le travail sied bien à un nécessaire et décrie un homme aisé, de même aussi le mariage est honnête à un homme sensuel » (1658, p. 237). Cette heureuse convenance se retrouve dans les contes de fées : « Je ne vous ai pas dit que, dès le premier mois, j'avais obtenu son consentement au mariage de

ses six garçons avec six jeunes filles pauvres, mais belles, sages, pieuses et pleines d'amour pour le travail, qui en étaient adorées. » (Nodier, *La fée aux miettes*, 1831, p. 171). On s'en doute, les romans réalistes sont des formes rassises du conte de fées : « Sandoz expliqua ses idées sur le mariage, qu'il considérait bourgeoisement comme la condition même du bon travail, de la besogne réglée et solide, pour les producteurs modernes. » (Zola, *L'œuvre*, 1886, p. 174).

Le lien entre le mariage et le travail est évidemment l'argent et la propriété : le code civil français, dit *code Napoléon*, commence son chapitre *Des moyens d'acquérir la propriété* par une section intitulée *Le mariage*. On en voit l'incidence dans des exemples comme ceux-ci : « Son père, ingénieur intelligent, inventif et débrouillard, ouvert aux idées nouvelles, avait fait sa fortune grâce à son travail, ses relations politiques et son mariage. » (Romain Rolland, *Jean-Christophe*, 1910, p. 1103)¹¹.

L'adultère fait son œuvre. — Paul Bourget en atteste par trois fois : « Ce venin de l'adultère, dont il avait infecté cette créature, accomplirait son œuvre de destruction. » (*Un crime d'amour*, 1886, p. 271 ; cf. aussi p. 286 ; et *Cruelle énigme*, 1886, p. 108). Le lien établi entre l'adultère et l'œuvre (ou du moins entre ces mots en français) tient à un stéréotype verbal issu de la traduction autorisée du Décalogue : « l'œuvre de la chair ne désireras qu'en mariage seulement. » (cité entre autres par Maupassant, *La confession de Théodule Sabot*, *Contes et nouvelles*, 1883, p. 43 ; Roger-Victor Pilhes, *La rhubarbe*, 1965, p. 225)¹².

Avec la laïcisation progressive de la société, l'œuvre de chair s'est identifiée avec l'adultère, car le mot *chair* reste associé au *péché*¹³. Quoi qu'il en soit, l'expression *œuvre de chair* permet de poser le problème du caractère formulaire de la doxa. Les normes sémantiques se concrétisent dans des formules sémiotiques : elles ne sont pas séparables de leurs expressions privilégiées, même si l'on se plaît à innover en leur trouvant de nouvelles formulations. Les formules prégnantes sont en fait des passages de mythes oubliés ou du moins ritualisés.

Alors que le mariage et le travail sont liés par la loi civile, l'œuvre et l'adultère sont liées par la loi religieuse. Ces deux lois ne sont évidemment pas sans rapport, mais retenons que la littérature, surtout la mauvaise, reflète la doxa, formée et sous-tendue par ces lois. Enfin, l'esthétique romantique tardive, du moins sous la forme affadée et vulgarisée qui fait à présent partie du sens commun, ne pouvait qu'opposer l'œuvre au travail, comme l'amour à l'argent, la possession à la propriété, le poète au bourgeois et l'amant au mari.

3. Conclusion

Les objets culturels ont beau dépendre de leurs conditions d'élaboration et d'interprétation, les valeurs qu'ils concrétisent peuvent cependant être objectivées comme des faits. Partout, l'on a affaire maintenant à des corpus numériques, qu'il s'agisse de musiques, d'images fixes ou animées, de danses, de performances polysémiotiques comme le cinéma, l'opéra, les rituels, etc. L'exigence scientifique de décrire de tels corpus rencontre ici la demande sociale. Avec les corpus numériques, les sciences de la culture trouvent ainsi de nouvelles perspectives épistémologiques et méthodologiques, voire un projet fédérateur.

La philosophie du langage qui a toujours tenu lieu de sémiotique jusqu'à nos jours, qu'elle soit d'inspiration néothomiste, augustinienne-phénoménologique, ou logico-positiviste, n'est hélas ici d'aucun secours, en raison de son inadéquation épistémologique et de son indigence empirique, tout particulièrement quand il s'agit d'établir un corpus de façon critique, de varier les critères d'interrogation et d'interpréter les résultats des requêtes. La notion même de corpus lui reste étrangère.

En revanche, un saussurisme renouvelé, ou néo-saussurisme, par son exigence méthodologique même, a fait la preuve de sa pertinence empirique. Il le doit notamment aux liens maintenus avec la philologie, pour ce qui concerne le recueil critique et l'indexation des documents, comme avec l'herméneutique, pour ce qui intéresse l'interprétation des œuvres. Ainsi, au sein même de la sémiotique, se reflète la contradiction majeure que doivent affronter les sciences de la culture : c'est dire l'importance du débat épistémologique qui s'ouvre à présent.

¹¹ Cf. aussi Georges Duhamel, dans sa préface au *Notaire du Havre* : « Issu de paysans très pauvres, il acquit, moitié par son travail et moitié grâce à son mariage, un petit bien dont il vécut. » (1933, p. 14).

¹² On trouve aussi diverses mentions obliques, comme celle de Joséphin Péladan (*Le vice suprême*, 1884, p. 234) : « celles qui désirent d'accomplir l'œuvre de chair hors le mariage, bien que l'effet ne s'ensuive pas, pèchent mortellement. »

¹³ L'expression *le péché de la chair* désignait l'adultère. Le droit canon ne condamne aucunement l'œuvre de chair, mais met en garde contre la fornication dans le mariage, dès lors que l'œuvre de chair ne prend pas pour but la procréation ; cependant, ces distinctions ne sont évidemment plus comprises par les contemporains.

N. B. — Cet aide-mémoire puise pour l'essentiel dans *La mesure et le grain*, Paris, Champion, 2011, ouvrage auquel nous nous permettrons de renvoyer pour les références et développements nécessaires. La bibliographie des sources littéraires est celle de Frantext, consultable sur le site de l'Atilf.