

Des genres à l'intertexte

From texts genre to intertextuality

François Rastier et Bénédicte Pincemin



Édition électronique

URL : <http://journals.openedition.org/praxematique/1974>

DOI : [10.4000/praxematique.1974](https://doi.org/10.4000/praxematique.1974)

ISSN : 2111-5044

Éditeur

Presses universitaires de la Méditerranée

Édition imprimée

Date de publication : 10 janvier 1999

Pagination : 83-111

ISSN : 0765-4944

Référence électronique

François Rastier et Bénédicte Pincemin, « Des genres à l'intertexte », *Cahiers de praxématique* [En ligne], 33 | 1999, document 3, mis en ligne le 01 janvier 2014, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/praxematique/1974> ; DOI : <https://doi.org/10.4000/praxematique.1974>

Tous droits réservés

François RASTIER
CNRS
Bénédicte PINCEMIN
INRP — Technologies Nouvelles et Education

Des genres à l'intertexte

Avertissement. — Bien qu'elle traite des genres, cette étude ne relève d'aucun. Est-ce un article à deux voix, tardif écho académique d'une rhapsodie ? Au lecteur d'inventer une catégorie *ad hoc*, car nous ne prétendons pas créer ici un genre nouveau. Si nous n'avons pas décelé de contradiction flagrante entre les sections que nous avons pris en charge, nous en assumons sereinement les différences. F.R. & B.P.

1. Définition des corpus et structures de l'intertexte (B. Pincemin)

1.1. Effectivité des genres textuels

a) L'intertextualité première

De même que le texte est l'objet effectif de la linguistique (toute pratique langagière est de l'ordre d'un texte)¹, l'intertexte est le milieu naturel de toute relation au texte. Cela s'observe pour des approches aussi différentes que la linguistique, l'herméneutique, ou les sciences de la documentation. De fait, un texte n'est qu'artificiellement perçu

F. Rastier, 57, rue de Paris — 94340 Joinville-le-Pont
lpe2@ext.jussieu.fr

B. Pincemin, 91, rue Gabriel-Péri — 92120 Montrouge
Benedicte.Pincemin@inrp.fr

1 Certes, Benveniste préférerait « fixer dans la phrase le niveau ultime » de l'analyse linguistique (Benveniste 1966 : 131), mais notamment Rastier (1989) met en évidence que le palier du texte, toujours présent, est un déterminant de la description linguistique.

isolément : il se profile sur d'autres textes, par rapport auxquels il prend sens.

b) Contexte de la situation et contexte des textes

Faut-il faire du contexte situationnel et du contexte linguistique deux objets d'étude indépendants, le premier relevant de la pragmatique, le second d'une linguistique supraphrastique ? Une sémantique des textes invite à penser qu'il s'agit de deux faces d'un seul objet, que ces deux contextes se correspondent et se révèlent l'un par l'autre. Le genre, qui se manifeste dans la matière linguistique même du texte, est comme l'empreinte significative et caractéristique d'une situation socioculturelle et du rapport au texte qu'elle prévoit². Autrement dit, le genre est le répondant, en tant que pôle intrinsèque du texte, des circonstances de rédaction et de lecture, qui font partie des pôles extrinsèques du texte.

La reconnaissance d'un système de pôles intrinsèques du texte, traces d'un système de pôles extrinsèques qui instaurent le texte en le situant, fonde la valeur d'une approche s'en tenant au texte (Rastier 1996). Celle-ci ne consiste pas à nier l'entour du texte, mais à le percevoir à travers ses incidences dans le texte.

1.2. Constitution et significativité du corpus

a) Le contexte comme point de vue : le corpus de référence

Le contexte n'est ni une totalité infinie et informe (un « tout le reste »), ni un entourage qu'une exploration méthodique permettrait de cerner. Ni infini, ni déterminé, il est *a priori* indéfini : multiple, mouvant, mais prenant place dans un champ de contraintes, de lignes de force.

De même, le corpus, qui se fait l'image d'un contexte saisi sous un certain angle de vue, est toujours relatif, sans pour autant être arbitraire. L'analyste qui procède à des statistiques textuelles assume cela, à juste raison, en distinguant en pratique : (i) un *corpus existant*, correspondant aux textes accessibles dont il peut disposer, (ii) un *corpus de référence*,

² Les autres typologies évoquées au § 2.1.a) (langues, discours, etc.) entrent d'ailleurs dans la même interrelation avec le contexte situationnel, à d'autres degrés de généralité.

constituant le contexte global de l'analyse, ayant le statut de référentiel représentatif, et par rapport auquel se calcule la valeur de paramètres (pondérations...) et se construit l'interprétation des résultats, (iii) un *corpus de travail*, ensemble des textes pour lesquels on veut obtenir une caractérisation, et le cas échéant (iv) un *corpus d'élection*, sous-corpus du corpus de travail, contrasté par rapport à celui-ci (Bommier-Pince-min 1999, § V.A). Le *corpus existant* est un état de fait : « corpus » est ici pris dans son sens le plus large, d'ensemble de textes. C'est à partir du *corpus de référence*, construit en vue de telle ou telle étude ou famille d'études, qu'apparaissent des conditions de signifiante (pertinence, cohérence), d'acceptabilité (représentativité, régularité, complétude), d'exploitabilité (homogénéité, volume). Ainsi, le corpus de référence, le corpus de travail et le corpus d'élection se distinguent non pas tant par leur constitution (interne) que par leur rôle, leur position respective dans l'analyse : le premier ancre une référence interprétative, le second définit le champ à explorer, le troisième correspond à un objet à caractériser.

b) Finitudes, frontières et perception sémantique : les unités textuelles au sein du corpus

Le contexte n'est pas une réalité à côté du texte, dont la prise en compte serait un raffinement, intervenant sur le mode d'ajouts ou de corrections. Le contexte est l'extériorité qui dessine le texte. Et c'est dans ce jeu de contraste forme / fond, de répétition et de différence, que le texte devient saisissable comme tel, s'articule en unités de sens et devient lui-même unité de sens. Le contexte précède le texte, bien plus qu'il ne l'accompagne.

Si bien que, comme le souligne très justement Péry-Woodley (1995), un corpus n'est véritablement exploitable que s'il est constitué par des textes soient donc des *unités* motivées, correspondant à une pratique langagière, à un usage interprétatif, et non pas par du texte amalgame obtenu en piochant dans les productions linguistiques comme dans un matériau, comme dans une masse pouvant être savamment sondée ou encore débitée au poids (Kilos- ou Méga-octets...). Car même si l'objectif de description vise la langue, celle-ci n'est observable qu'à travers des textes. Un corpus de phrases ou un corpus

d'échantillons ne sont pas utilisables en tant que tels³, puisqu'il manque des frontières sémantiques, qui d'une part organisent des paliers et des zones de localité concourant directement à la construction du sens (paragraphe, parties, texte), et d'autre part autorisent (par un acte qui n'est autre qu'un jugement herméneutique) à conférer une certaine autonomie aux textes ainsi définis. Or c'est cette autonomie des textes, puis du corpus, qui fait que le corpus peut être soumis à l'analyse, devenir un objet d'étude à part entière.

Dans les calculs sur corpus, l'inventaire et le décompte des séquences, des cooccurrences, de la dispersion d'un « mot » sur l'ensemble des textes, est le moyen *a minima* de pondérer et qualifier des unités d'analyse que sont par exemple les mots graphiques. Une perspective plus ambitieuse est de réajuster et de reconstruire ces unités locales, ayant connaissance de leur comportement dans un contexte global : ainsi donne-t-on un contenu opérationnel au principe de contextualité mobilisé par l'activité interprétative (cf. § 2.2.b). Bommier-Pincemin (1999, § VI) s'appuie ainsi sur la propriété sémantique de détermination du local par le global pour mettre au point une nouvelle forme d'indexation automatique adaptée au texte intégral. Lafon et Salem (1983) avaient ouvert la voie, en envisageant l'application de leur algorithme des *segments répétés* à la réalisation d'une « segmentation non minimale ».

L'effectivité et la légitimité de telles procédures automatiques de construction tiennent à la remarquable propriété de convergence en faisceaux des critères. Autrement dit, la réalité, notamment linguistique, n'est pas uniformisante et égale, mais se plaît à façonner des formes originales et contrastées.

³ Corpus de phrases et corpus d'échantillons ne sont pas utilisables en tant que tels pour la description linguistique, ce qui n'exclut pas par exemple l'apport heuristique et méthodologique de batteries de tests destinés à cerner le comportement formel d'un analyseur morphosyntaxique. Toute la question est ici de préciser la nature et la portée des résultats qui peuvent être légitimement tirés de tel ou tel corpus. Voir aussi (Pincemin 1999).

c) Définir des regroupements de textes : les genres textuels et la construction de corpus

Les genres textuels sont un paramètre important de la construction d'un corpus. Leur prise en compte s'impose dès qu'il s'agit d'avoir une vue d'ensemble sur un certain nombre de textes : ceux-ci ne sont comparables et commensurables que dans la mesure où leur analyse et leur parcours sont guidés par la connaissance des genres concernés. Chaque genre a son vocabulaire de construction, ses formes d'organisation, ses contenus attendus, ses modes rédactionnels. La portée sémantique d'un mot⁴, ou le sentiment de longueur ou de brièveté, varient selon le genre.

Il faut bien garder leur autonomie aux deux concepts de *corpus* et de *genre*. Les genres sont déterminés par les pratiques sociales. Ils sont reconnus et décrits par la linguistique, car c'est une réalité intertextuelle, par laquelle peuvent s'expliquer certaines affinités et certaines régularités entre des textes. Mais, notamment pour les besoins du codage et de la structuration des grandes bases textuelles, le genre court toujours le risque d'une définition théorique figée. Celui qui veut étiqueter et classer les textes d'un corpus par genre risque de voir la délimitation de ses « paquets » se dérober. Où commence et où finit le genre ? Parions que les discussions soient encore pour longtemps ouvertes.

Le corpus, lui, relève d'un point de vue, contingent, — parmi une multiplicité d'autres points de vue possibles⁵ —, donnant un cadre à la constitution d'un objet. C'est un paramètre méthodologique, car l'étude veut que l'on se donne un domaine d'action, relatif à une recherche particulière⁶. Le corpus s'inscrirait davantage du côté de l'herméneu-

⁴ Par exemple, dans le manuel d'initiation à l'utilisation d'un traitement de texte, l'usage de « mot » (« Sélectionnez un mot en double-cliquant dessus... ») ne retiendra pas spécialement l'attention du lecteur ; il en irait tout autrement des emplois de « mot » dans un traité de lexicologie.

⁵ Sur la multiplicité des corpus et intertextes, voir notamment Thlivity (1998) : § 1.2.2, p. 22 sq. (intertexte centré texte, ou centré auteur et plutôt descriptif, ou centré lecteur et plutôt productif) ; § 2.1.3, p. 39 (incidence de l'intertexte sur ce que l'on perçoit dans la lecture d'un texte), et p. 41 (l'intertexte comme point de vue).

⁶ Le caractère subjectif et singulier d'un corpus peut être relativisé, dans la mesure où il conduit à une exploitation et à des résultats s'inscrivant dans un cadre d'analyse

tique que de celui de la linguistique. Il est défini par un objectif de lecture, d'analyse. Toute étude (de statistique textuelle, de texte), explicitement ou implicitement, se donne tôt ou tard un corpus, dans lequel elle va sélectionner, calculer, contraster. Le corpus a été fixé, la manière dont cela a été fait peut être discutée, argumentée, mais de toutes façons la définition du corpus est effective.

En somme, un ensemble de textes relevant d'un même genre peuvent constituer un corpus intéressant et fructueux. Pour autant, ce n'est pas le seul corpus valide⁷. Le genre est un facteur qui contribue à l'homogénéité du corpus, mais d'autres modes de cohésion sont possibles.

1.3. Structures intertextuelles

a) L'introduction de structures élémentaires pour modéliser un intertexte

L'opération de codage de corpus électroniques en vue de traitements sémantiques invite à expliciter ce qui fait sens comme représentation d'un intertexte.

La concaténation est rédhibitoire : elle ignore les unités textuelles et considère le corpus comme *du* texte, au lieu d'y reconnaître *des* textes. La juxtaposition \emptyset le corpus comme ensemble de textes \emptyset est admissible, mais donne une vision « plate » : tous les textes ont des relations indifférenciées les uns avec les autres.

commun et réutilisables. Voir par exemple l'effort de Thlivitit (1998) dans ce sens, pour la réutilisation d'« interprétations » (classes sémantiques) basées sur un corpus : « la méthodologie de la Sémantique Interprétative Intertextuelle [...] oblige à l'explicitation des *sources sémiqes* à l'origine de la constitution d'une classe sémantique. Le lecteur est donc indirectement incité à réutiliser les interprétations existantes, en y apportant sa propre interprétation. De cette manière, nous proposons le dépassement, à la fois, de la volonté d'atteindre une juste *objectivité descriptive* et de la liberté d'une *subjectivité descriptive* totale en les remplaçant par un *consensus inter-subjectif*, issu de l'interaction *multi-utilisateur* avec un *espace commun d'analyses* et soutenu par une méthode de travail interprétatif qui incite à la *consultation* et à la *réutilisation* récurrentes » (Thlivitit 1998 : 28).

⁷ En l'occurrence, une analyse statistique visant la caractérisation des genres travaillera sur un échantillonnage reflétant la diversité des modes d'expression textuelle.

b) Rangements et boîtes

Une première forme de structuration intertextuelle peut être proposée suivant l'image de *rangements* dans des *boîtes* archive. Un rangement est une répartition des textes dans les boîtes. Un texte peut au besoin être classé dans plusieurs boîtes (comme si on pouvait en avoir plusieurs exemplaires), ou ne pas être classé (s'il ne relève pas des critères du rangement). Le corpus est alors structuré comme une série de possibilités de regroupements de textes (une série de rangements).

Cette modélisation évite donc le double écueil de la conception classificatoire et de la conception typologique (cf. § 4 2.1.b) puisqu'elle admet la multiplicité des classements, l'absence de hiérarchie stricte, et échappe à la problématique de désignation d'éléments centraux et de l'organisation qu'ils induisent.

Concrètement, la définition d'un rangement et de boîtes n'est que l'explicitation d'un acte interprétatif, le repérage de ce qui est perçu comme une organisation porteuse de sens pour l'analyse en cours. Un rangement, en tant qu'il reflète un point de vue sur le corpus, n'est donc jamais exclusif : plusieurs points de vue peuvent coexister. Il n'est pas non plus définitif : un rangement (point de vue) peut être réaménagé ; c'était par exemple une hypothèse interprétative *a priori* significative, que de premiers résultats d'analyse amènent à rectifier de même que la signification accordée à un texte, et le contexte dans lequel on l'inscrit, peuvent évoluer au fil des lectures.

Comment « fonctionne » cette structure ? Dans un rangement donné, chaque regroupement (boîte) a une double signification : il pose la cohésion des textes qu'il rassemble, et un contraste de ce groupe de textes avec les autres regroupements du même rangement. Tout l'enjeu consiste à reconnaître là l'opération de rapprochement et de contextualisation génératrice de sens (cf. §2.2.b), *via* les principes perceptifs fondamentaux d'assimilation et de dissimilation⁸. Le principe sémantique est donc différentiel : homologation des textes rassemblés dans une boîte, et différenciation des groupes de textes entre eux.

Un corollaire remarquable est la mise en suspens possible des métadonnées (qualifiant le texte mais externes à lui). Ainsi, les critères de rangement peuvent être variés : auteurs (une boîte regroupe les

⁸ Cf. Rastier 1987 : ch. III et VII ; 1991 : ch. VIII.

textes d'un auteur), type de texte (une boîte correspond à un genre textuel)... Pour autant, l'effet sémantique intertextuel sera modélisé de façon analogue, quel que soit le critère organisateur : ce qui contribue à construire le sens dans le cadre du corpus, *au niveau des textes eux-mêmes*, ce n'est pas de connaître le nom de l'auteur (donnée métatextuelle, qui renvoie à tout un savoir encyclopédique), mais d'indiquer l'organisation des textes qu'induit la prise en compte de cette information.

c) Piles et strates

La structuration en rangements et en boîtes convient par exemple à la description des discours normatifs, produits et interprétés sur le registre de la permanence (cf. § 2.1.b). La structure rangements / boîtes ne permet cependant pas de rendre compte des effets de progression, de dépendances ordonnées, d'échelle de valeur, par exemple d'effets de récence ou de lignée (cf. § 2.1.b). Une seconde forme de structuration est nécessaire : la *pile*, divisée en *strates*. Là encore, selon le critère organisateur que l'on se donne, le corpus se modèle comme telle ou telle pile, laissant de côté peut-être certains textes, et répartissant les autres en strates, hiérarchisant ainsi les groupes de textes. Cette structuration permet par exemple de traduire une gamme de niveaux de saillance des textes : importance accordée par une interprétation, récence valorisée dans le contexte applicatif considéré, etc.

Ces deux structururations intertextuelles élémentaires, rangements / boîtes et piles / strates, constituent une proposition simple et expressive pour le codage de corpus électroniques en vue de traitements sémantiques. Elles se fondent sur l'organisation du texte lui-même, et constituent donc un apport sémantique complémentaire aux informations explicitées dans les en-têtes (métadonnées).

2. Pour une poétique généralisée (F. Rastier)

Si la question des genres a été traitée par la poétique, le nom même de cette discipline évoque traditionnellement la littérature ; or, bien au-delà, c'est l'ensemble des normes et des usages linguistiques, oraux et écrits, artistiques ou non, qui relève de ce que l'on pourrait appeler une *poétique généralisée*. Comme la distinction platonicienne entre *praxis*

et *poiesis* n'est pas ici de mise, *poétique* s'entendra au sens pratique du terme — et non plus seulement esthétique. Adopter un point de vue unifié sur les genres des discours littéraires et non littéraires permet aussi de reconsidérer les objections souvent convergentes élevées par la critique contemporaine contre l'étude des genres littéraires, et par la linguistique textuelle contre l'étude des genres en général.

Si elle part de la tradition littéraire pour s'étendre aux genres non littéraires, une telle poétique pourra être dite *généralisée*. Mais elle ne sera pas pour autant universelle, car elle devra rompre avec l'universalisme traditionnel de la poétique transcendantale.

En outre, dès lors qu'elle prend quelque recul à l'égard de la poétique littéraire, elle doit assumer une tâche nouvelle : décrire la diversité des discours (littéraire, juridique, religieux, scientifiques, etc.) et leur articulation aux genres⁹.

⁹ Comme la linguistique ne traite pas de la diversité des discours (sauf dans des applications lexicographiques) et que la poétique s'interroge sur une problématique littéraire, la diversité des discours est encore moins étudiée que celle des genres. En effet, le discours est généralement conçu comme une dimension d'analyse (cf. la *discourse analysis* de Harris, *l'analyse du discours*) selon une dichotomie langue/discours. Nous employons, on le voit, le terme *discours* dans une toute autre acception, distributive, en admettant qu'un genre est ce qui rattache un texte à un discours (cf. 1989) : la pluralité des discours reflète celle des pratiques sociales.

Outre la fonction médiatrice du sémiotique, l'humanité est caractérisée par la spécification des pratiques et la division corrélatrice du travail. A chaque type de pratique sociale correspond un domaine sémantique et un discours qui l'articule. Avec une singulière profondeur, Dante, dans le *De vulgari eloquentia* (I, 7), explique pourquoi l'humanité a surmonté la confusion des langues : sur le chantier de Babel, les ouvriers des divers corps de métier ont pu continuer à se comprendre parce qu'ils partageaient des pratiques communes. Alors même que le nombre des langues décroît rapidement, un second chantier s'ouvre à la linguistique : celui de la diversité irréductible des discours et des genres dont traite la poétique généralisée.

Comme il n'y a pas de pratique générale non spécialisée, le concept de langue générale reste une formation de la philosophie du langage. Même les échanges linguistiques qui paraissent les plus spontanés sont réglés par les pratiques sociales dans lesquelles ils prennent place, et relèvent donc d'un discours et d'un genre. Le genre assure la médiation entre texte et situation, tels qu'ils sont unis dans une pratique. Le rapport entre la pratique et le genre détermine celui qui s'établit entre l'action en cours et le texte écrit ou oral qui l'accompagne.

Le concept de pratique appelle certes quelques éclaircissements. Toutes les pratiques, y compris celles qui mettent principalement en jeu le sémiotique, ont été

L'enjeu n'est pas mince, car les textes sont configurés par les situations concrètes auxquelles ils participent ; en outre, par la médiation des genres et les discours, ils s'articulent aux pratiques sociales dont les situations d'énonciation et d'interprétation sont des occurrences.

2.1. Déontologie et méthodologie

a) Comparatisme et genres : sept typologies

Comme la *méthode comparative* issue de la linguistique historique n'a aucunement démerité, nous poserons le problème des critères descriptifs dans son cadre épistémologique. Sept entreprises typologiques paraissent en relever.

(i) La typologie des *langues*, évidemment. (ii) Chaque langue, aux diverses étapes de son histoire, connaît des usages propres à des types de pratiques sociales ; pour en rendre compte, il faut une typologie des *discours*. (iii) Chaque discours compte un nombre déterminé de *genres*, dont la typologie rend compte de la diversité externe des textes. (iv) La typologie des *textes* traite de la diversité interne des genres. (v) Celle des parties d'un texte traite des *sections* (parties de textes délimitées par des critères d'expression) et *configurations* (parties de textes définies par des critères de contenu). (vi) Celle des *morphologies* traite de la parenté des textes, indépendamment des genres : il y a dans les textes un « vocabulaire » de formes sémantiques, dont certaines ont été reconnues et inventoriées par les traditions rhétoriques et poétiques (les figures non tropes, par exemple), mais aussi d'autres qui ne sont pas nommées, comme les *molécules sémiques* (structures stables de traits

pensées naguère sur le modèle des pratiques de production — d'où par exemple en Analyse du discours la notion de *conditions de production*. Le marxisme « réel » ne laissait aucune place spécifique au langage, et n'a d'ailleurs produit que des études de politique linguistique ou de sociolinguistique. Il ne pouvait en effet penser le niveau sémiotique des pratiques dans le cadre de son anthropologie basée sur la division entre physique et idéologique. Aussi à ses débuts l'Analyse du discours, issue de l'althussérisme, a programmé ainsi la recherche : à une linguistique formalisante (s'appuyant sur la *discourse analysis* asémantique de Harris) de repérer les marques de l'énonciation dans le discours, à une théorie des idéologies détenue par une instance politique de les interpréter. Aucune place, dans ce dispositif, pour une sémantique des textes ; en effet, la théorie des pratiques sociales relevait de la théorie des idéologies et du matérialisme historique, ce qui a naturellement rendu oiseuse voire suspecte la théorie des genres.

sémantiques, qui n'ont pas nécessairement de lexicalisation privilégiée). Cette typologie peut transcender les frontières linguistiques (ex. les motifs en folkloristique). (vii) Enfin, la typologie des *usages génératives* différencie des classes d'usagers ou des « styles »: par exemple, comme l'a relevé Sueur (1982), les locuteurs sans diplôme emploient à l'oral, dans des réponses ouvertes à une enquête, plus de pronoms, de négations et de verbes que la moyenne, les diplômés de l'enseignement supérieur plus de noms et d'adjectifs ; Verlaine, dans les *Romances sans paroles*, emploie quatre fois plus de pronoms de seconde personne que Rimbaud dans les *Illuminations*, etc.

b) Trois conceptions du genre

Bien qu'elles soient souvent confondues, on peut distinguer trois conceptions du genre : la *classe*, le *type* et la *lignée*.

(i) La conception classificatoire doit affronter tous les problèmes ordinaires des taxinomies, dont le moindre n'est pas la variabilité des critères. Certes, on a tenté de réformer cette conception en utilisant la notion d'« air de famille », au risque de multiplier et d'affaiblir les critères sans avoir le moyen de les hiérarchiser.

(ii) La conception typologique doit caractériser le rapport entre type et occurrences. Or, les types de textes sont des modèles hypothétiques, et leurs occurrences font sens tout autant parce ce qu'elles instancient le type que parce qu'elles s'en écartent. D'ailleurs, aucune théorie des types n'a pu constituer une sémantique de la variation des occurrences à l'égard des types. La théorie des prototypes a certes introduit du flou dans les taxinomies, mais sans parvenir à qualifier cette variation, car elle ne décrit le rapport entre exemplaires centraux et périphériques que par la métrique quantitative du nombre de traits (la *cue validity*).

Les deux premières conceptions du genre, la classe et le type, relèvent de la problématique logico-grammaticale : la première, de l'imaginaire classificatoire de la grammaire ; la seconde, de la logique.

(iii) La troisième conception considère le texte comme une « génération » dans une lignée de réécritures (cf. l'auteur, 1995). Dans une problématique rhétorique / herméneutique, l'opposition logico-grammaticale entre type et occurrence le cède à l'opposition philologique entre source et reprise. Bien entendu, des sources peuvent devenir canoniques et se trouver promues au rang de parangons ; cependant les

reprises modifient et transforment inévitablement les sources, puisque de fait le changement des contextes rend toute répétition impossible. Le rapport entre les textes est alors médiatisé par une série de réécritures et d'interprétations qu'elles concrétisent. Si bien que le problème de l'interprétation ne trouve plus à se poser à propos du rapport atemporel entre type et occurrence, mais dans le temps, scandé de ruptures, d'une tradition. Ainsi, un thème littéraire n'est pas un type (au sens ontologique), mais une famille de transformations. La textualité elle-même est faite de ces expositions, développements, reprises et variations¹⁰.

Au sein même des genres, ce que l'on a nommé des sous-genres sont des lignées génétiques spécifiques : les auteurs de romans picaresques se sont par exemple imités les uns les autres, comme de raison ; l'*Arcadia* de Sannazar ouvre la lignée européenne du roman pastoral, la *Nadja* de Breton réécrit l'*Aurélia* de Nerval, etc.

Le statut même des genres, comme le rapport entre les textes et leurs genres, diffèrent sans doute selon les discours. Dans les discours normatifs, les textes sont produits et interprétés sur le mode de la permanence, comme des occurrences de leur type ; tandis que dans les discours normés, mais non normatifs, ils sont produits et interprétés comme des transformations de leurs sources.

Cependant, la caractérisation d'un texte, dans la mesure où elle procède d'un parcours interprétatif *a priori* indéterminé, peut partir de n'importe quel niveau de « profondeur » de la classification : d'une configuration pour arriver directement au discours, d'une section pour identifier le genre, du genre pour caractériser une section, etc. Enfin, la sélection des critères pertinents varie évidemment selon les objectifs de l'application.

c) Que les genres ne sont pas des « types de textes »

Gardons-nous de confondre les « types de textes » et les genres. Les types de textes sont des classes qui ne reposent que sur un critère ; par exemple, les fameux *genera* de Diomède sont distingués par le critère

¹⁰ Nous ne retenons pas ici les notions de *jeu de langage* et d'*air de famille* qui semblent trop vagues pour définir une quatrième conception du genre. La première n'est pas spécifique aux textes, et peut valoir pour de simples propositions (voir cependant Bouquet, 1998) ; la seconde doit peut-être son succès à son imprécision, qui permet d'en user à tout propos.

énonciatif : seul le poète parle, seuls les personnages parlent, ou ils alternent leur propos. Le privilège exorbitant donné à ce critère a d'ailleurs entravé le développement de la poétique jusqu'à Hegel inclus.

Mesurer les diversités suppose une réflexion sur les critères : à supposer même qu'elles soient fondées, des oppositions comme *fiction vs non fiction* sont trop grossières et départagent, au mieux, des classes de discours mais non des genres. Sans prétendre que les espèces culturelles ressemblent aux naturelles, prenons l'exemple de la classification des espèces : que vaudrait une théorie qui regrouperait le chat, le poulpe et la chèvre sous le juste prétexte que ces animaux supérieurs ont la pupille fendue verticalement ? Les typologies qui prêtent à un seul critère un caractère déterminant, comme celle de Benveniste¹¹, procèdent pourtant ainsi, et deviennent normatives — avec de moins solides raisons que le Lévitique, quand il classe les animaux selon qu'ils ont le sabot fendu ou portent des écailles.

En assimilant la théorie des genres à la typologie des textes, on oublie que la définition d'un type de texte dépend de l'analyste : pour les besoins d'une cause ou d'une application, il peut inventer une catégorie quelconque qui divise un corpus (roman en *je* ou en *il*, textes longs ou courts, d'avant ou d'après 1945, comme à la Bibliothèque Nationale de France). Les adversaires de l'étude des genres glosent d'ailleurs sur la relativité sinon la vanité de telles typologies. Si, pour

¹¹ Dans un article illustre (1966 : 237 sq.), mais dont on a négligé la généalogie platonicienne, Benveniste oppose ainsi l'usage des historiens (*L'histoire grecque* de Glotz) et des romanciers (*Gambara* de Balzac), pour distinguer entre deux régimes d'énonciation, l'*histoire* et le *discours*. Le critère demeure le critère platonicien (qui parle ?) : « personne ne parle ici » écrit Benveniste à propos de Glotz (p. 241). À l'inverse, dans le discours, quelqu'un parlerait ? c'est Balzac, apparemment confondu avec le narrateur de *Gambara*. À partir de « marques » énonciatives, minces et discutables (un historien peut dire *je*, un romancier employer le parfait, cf. *L'Étranger* de Camus), Benveniste oppose ainsi un genre, le traité historique (et non l'histoire) à une notion, le discours, qui reste sans consistance, à moins de subsumer un groupe de genres non autrement défini. La définition benvenistienne du discours reprend en outre le thème platonicien de la supériorité de l'oral sur l'écrit (cf. Platon, *Lettres*, VII). N'importe, et bien que Pêcheux ne l'ait pas reprise, elle restera fondatrice pour l'école française d'analyse du discours, et permet encore aujourd'hui d'opposer le discours (témoignant de l'énonciation) au texte (assimilé à la structure de l'énoncé).

ainsi dire, la typologie peut assumer un principe de plaisir, la théorie des genres doit obéir à un principe de réalité, car les genres ne devraient point être créés par les poéticiens.

La poétique ne vise pas seulement une typologie des textes. Elle doit certes produire et hiérarchiser des critères descriptifs, mais surtout rechercher leurs interactions. Les genres sont en effet définis non par un critère, mais un *faisceau* de critères. Ils doivent d'ailleurs leur caractère d'objectivité à cette multiplicité des critères, et les nouveaux critères que l'on propose en linguistique de corpus ne s'appliquent efficacement qu'en fonction de la différence des genres, et la confirment au lieu de l'infirmier.

Un genre se définit en effet par (i) la cohésion d'un faisceau de critères, tant au plan du signifié qu'à celui du signifiant, et par (ii) son incidence sur la textualité, sur ces deux plans également (cf. l'auteur, 1989). Aussi, il détermine ce mode de corrélation entre plan du signifiant et plan du signifié que l'on peut nommer *sémiosis textuelle* (cf. *infra*).

Ce n'est donc pas la typologie des textes, mais celle des genres qui nous importe. Or cette typologie est subordonnée à celle des discours. L'existence de genres transdiscursifs reste douteuse, car le voisinage d'autres genres (ou, s'il s'agit de genres inclus, d'autres contextes d'inclusion) suffit à les modifier : un proverbe par exemple n'a pas le même sens dans un discours ludique ou dans un discours juridique ; la lettre commerciale n'a presque rien de commun avec la lettre personnelle du discours privé, car la corrélation entre contenu et expression demeure critique pour la définition du genre.

De fait, il semble qu'aucune typologie des textes fondée sur des critères définis indépendamment des genres (comme oral vs écrit, public vs privé, etc.) n'a permis d'isoler des genres : par exemple, les tentatives de classification automatique conduites par Biber ont conclu à des variations très importantes selon les corpus, mais sans parvenir à isoler des genres (notamment parce qu'il se limite à des critères morphosyntaxiques)¹². On objectera certes qu'un genre n'est qu'un

¹² Au moyen d'une analyse multidimensionnelle, Biber (1993 a) étudie les cooccurrences entre 67 traits morphosyntaxiques dans les mille premiers mots de textes divers d'anglais contemporain.

type parmi d'autres, voire un « genre de type ». La question sera tranchée quand on aura produit des faisceaux de critères stables et cohérents, indépendants de toute connaissance préalable des genres et des discours, qui ne recourent ni les genres ni les discours, et qui aient une pertinence théorique et pratique. Alors, — le défi est lancé —, la typologie des textes sera devenue indépendante de la théorie des genres.

Comme, dans notre hypothèse, c'est l'étude des genres qui permet de déterminer la pertinence des critères, toute typologie rigoureuse des textes procéderait d'une typologie des genres. Sur ce point cependant, la tâche scientifique propre à la linguistique diffère des tâches techniques (didactiques, etc.), pour lesquelles on pourra admettre des typologies diverses, et elle diffère également des tâches d'autres disciplines comme la sociologie, l'histoire (de l'édition, par exemple) ou la psycholinguistique.

Pour une sémantique des genres, on peut rechercher des critères de corrélation entre composantes sémantiques, comme ceux-ci, à titre indicatif : thématique ouverte / fermée, concentrée / diffuse ; dialectique ordonnée, désordonnée, impertinente, orientée positivement ou négativement ; dialogique variant ou non les foyers de l'énonciation et de l'interprétation représentées ; tactique pertinente, ou non pertinente, etc. On cherche à caractériser les modes de leur co-variation (cf. l'auteur, 1989 : 35-109), un genre étant alors défini comme un mode d'interaction normé entre composantes.

La combinatoire des composantes n'est aucunement libre, car beaucoup de combinaisons possibles ne sont pas attestées : par exemple, on ne rencontre nulle part de ballades non narratives, qui compteraient un nombre variable de strophes, et dont le nombre de vers serait fixe ; ou encore, comme le notait Zumthor, les structures narratives de la chanson de toile auraient pu être, en théorie, reprises dans des contes, mais elles lui sont restées spécifiques. Pour sembler contingentes, les combinaisons de traits définitoires du genre n'en sont donc pas moins contraignantes.

d) Genres et sémiotique textuelle

On définit ordinairement la sémiotique au palier du signe, et comme un rapport entre signifié et signifiant ; mais on ne s'interroge guère sur les paliers supérieurs, comme si leur sens se déduisait par composition de la signification des signes. Or, un genre définit précisément un rapport normé entre signifiant et signifié au palier textuel : par exemple, dans le genre de l'article, au premier paragraphe (sur le plan du signifiant) correspond ordinairement une introduction (sur le plan du signifié) ; dans le genre de la nouvelle, il s'agira plutôt d'une description.

Outre les régimes de production et d'interprétation des textes qui relèvent du genre, la sémiotique textuelle détermine, semble-t-il, le mode de mimésis. En règle générale, plus les rapports entre les deux plans du texte sont normés, plus son effet de réel (empirique ou transcendant) est intense, comme en attestent les textes gnominiques ou religieux¹³.

Le genre est donc le facteur fondamental de la sémiotique textuelle. Dans certains discours, comme le discours littéraire, des normes additionnelles peuvent être élaborées et mises en œuvre (cf. les « styles d'auteur ») ; mais que ce soit pour les spécifier ou pour les contester, elles s'appuient sur les normes génériques.

Le problème de l'arbitraire du signe peut se reformuler au palier du texte à propos de ce que l'on pourrait nommer *l'arbitraire du texte*. Il résulte de l'appariement *a priori* contingent entre les structures génériques du plan du signifié et du plan du signifiant : par exemple, la ballade française a un nombre fixe de strophes, mais le nombre de leurs vers ne l'est pas, et elle n'est presque jamais narrative ; la ballade anglo-saxonne a en revanche un nombre de strophes variable, mais à nombre de vers fixe, et elle est en général narrative. Les poétiques essentialistes et normatives, comme celle de Boileau, ont évidemment considéré ces corrélations génériques apparemment contingentes comme des convenances fondées en nature ; ce qui n'est pas faux, si comme pour le « sens propre » du signe, on appelle *nature* les habitudes invétérées d'une société.

¹³ Sur la typologie des effets de réel, cf. Rastier (1992).

e) La poétique commence au palier des discours

À une typologie *a priori* des genres, on préférera la recherche des axes typologiques qui dépendent des discours. Le projet d'une typologie transdiscursive paraît en effet illusoire, dès lors que les genres sont spécifiques aux discours : par exemple, un texte technique ne peut être assimilé à un texte scientifique, et même dans des discours aussi proches que les discours scientifiques, les genres ne sont pas exactement comparables, car chaque discipline a ses traditions et ses normes, et par exemple un traité de physique n'obéit pas aux mêmes normes qu'un traité de linguistique.

À chaque discours, on peut faire correspondre un système ou *synmorie*¹⁴ générique. Chaque groupe de pratiques sociales correspondant à un discours se divise en activités spécifiques (ex. le jury de thèse, la conférence, le cours, la correction de copie, etc.), qui ont chacune leurs genres. Par exemple, dans le discours médical, on peut distinguer les genres écrits dont dispose un professeur des hôpitaux dans sa pratique professionnelle : ils sont au nombre de trois, le résumé d'observation clinique, l'article scientifique et la lettre au collègue ; chacun correspond à une de ses pratiques.

Pour relier les genres aux discours, la poétique généralisée a pour tâche d'étudier les *synmories* dans leur spécialisation et leur co-évolution. Nous ne sommes sur ce point qu'au début des recherches, et l'expérience montre que les étudiants en lettres, voire leurs professeurs, sont fort déconcertés quand on leur demande d'inventorier les genres qu'ils utilisent. La difficulté est d'autant plus grande que tous les discours n'ont pas le même régime générique : il est évident que celui de la littérature est des plus complexes, et sans doute le plus révélateur. Si les discours peuvent s'influencer les uns les autres et partager des traits communs au sein de ce que Foucault appelait une *épistémé*, chacun des systèmes génériques reste cependant autonome et évolue selon ses propres lois, comme en témoignent leurs évolutions diachroniques différentes.

Le problème des *synmories* se retrouve, transposé, quand on étudie les relations entre genres dans l'œuvre des auteurs qui en ont pratiqué

¹⁴ Je transpose ici un terme que Longin utilisait pour désigner les associations de tropes, à l'image des groupements de citoyens.

plusieurs. Par exemple, chez Primo Levi, les poèmes inclus dans les témoignages, romans et essais, à titre de citations ou d'exergues, permettent de faire communiquer les thématiques de ces œuvres de genres différents et de restituer l'unité de son univers sémantique, bien qu'elles abordent des thèmes propres, qui n'auraient pas pu figurer dans les autres genres. Contraignant une sorte d'intertextualité propre à l'œuvre, chaque poème cité évoque dans chaque livre la présence des autres. En recueillant ces poèmes, *Ad ora incerta* (1984) permet de relier des récits de témoignage comme *Se questo è un uomo* (1957), et *La tregua* (1963), un roman comme *Se non ora, quando ?* (1982), et un essai comme *I sommersi e i salvati* (1986).

f) Des genres suprêmes et des sous-genres

Tout classement, dès lors qu'il prétend refléter une « nature des choses », retrouve les problèmes classiques de l'ontologie, car l'ontologie occidentale s'est formée à partir des problèmes de catégorisation, d'abord posés par la théorie de la prédication vraie. Par contraste, on peut considérer le genre comme le niveau de base dans la classification des textes, pour trois raisons convergentes.

(i) Il n'y a pas de genres suprêmes (pas de genre de genres), car les critères de groupement des genres sont les discours — et les pratiques qui leur correspondent. Aussi, de grandes catégories de l'expression, comme la prose ou l'oral, conduisent à des regroupements oiseux (l'oral, de la brève de comptoir au réquisitoire, n'a évidemment pas plus d'unité que la prose).

(ii) Les parties de genres sont elles-mêmes relatives à ces genres : par exemple, la description inaugurale dans la nouvelle du XIX^e n'est pas une simple occurrence de *la* description.

(iii) Les sous-genres, comme le roman « de formation » ou le roman policier sont définis par diverses restrictions qui intéressent soit le plan de l'expression (par exemple le roman par lettres, le traité versifié), soit celui du signifié. Elles doivent être spécifiées selon les composantes mises en jeu : thématique et dialectique pour le roman policier, par exemple, dialogique pour les romans fantastiques, tactique pour les

sonnets liminaires, etc. On évitera de retenir des critères contingents : le roman du XVIIe est par exemple une création purement académique¹⁵.

2.2. Du corpus à l'intertexte

a) La constitution des corpus

La caractérisation raisonnée des genres reste un préalable à la constitution de corpus pleinement utilisables pour des tâches de description linguistique. Quels que soient les critères choisis, on ne peut tirer grand-chose d'un corpus hétérogène, car les spécificités des genres s'annulent réciproquement, et les disparates qui demeurent ne peuvent être interprétées pour caractériser les textes.

Or, avec l'essor de la numérisation, on se trouve fréquemment devant des corpus hétérogènes : par exemple, les corpus issus d'une même entreprise de presse (cf. Illouz et coll., 1999, à propos du journal *Le Monde*), voire d'autres, qui sont simplement recueillis sur la Toile. Aussi faut-il envisager le *profilage* des corpus (cf. *op. cit.*, 1999), ce qui exige la mise au point d'outils permettant de travailler sur des sous-corpus homogènes. La création de tels outils donnera sans doute des moyens d'avenir à une poétique expérimentale.

b) Paliers du contexte et de l'intertexte

Si l'on convient que l'interprétation est une activité, et non une représentation ou une contemplation, elle procède principalement par contextualisation. Elle rapporte le passage considéré, si bref soit-il — ce peut être un mot : (i) à son voisinage, selon des zones de localité (syntagme, période) de taille croissante ; (ii) à d'autres passages du même texte, convoqués soit pour des tâches d'assimilation, soit de contraste ; (iii) enfin à d'autres passages d'autres textes, choisis (délibérément ou non) dans le corpus de référence, et qui entrent, par ce choix, dans le corpus de travail.

¹⁵ Voici comment un ouvrage universitaire subdivise le théâtre : « Le genre dramatique : La Tragédie. La Comédie. La farce. La commedia dell'arte. Le vaudeville. Les comédies spécialisées : La comédie-ballet. La Comédie de caractère. La comédie héroïque. La Comédie de mœurs. Le Drame : Le Drame bourgeois du XVIII^e siècle. Le Drame romantique. Le Drame symboliste. Le mélodrame ».

Aucune de ces trois contextualisations n'est déterministe, au sens de mise en Intelligence artificielle, qui suppose un parcours linéaire mot à mot. La première peut être rétrograde ; les deux autres sont peu contraintes par la linéarité du texte ou des textes qui font l'objet des rapprochements.

Qu'elle contextualise ou recontextualise, dans tous les cas la pratique des rapprochements génère du sens, de manière d'ailleurs inévitable sinon compulsive, selon un *principe de contextualité* qui pourrait s'énoncer ainsi : deux signes — ou deux passages d'un même texte mis côte à côte — sélectionnent réciproquement des éléments de signification (sèmes). Cet échange transforme leur signification en sens (soit par validation de traits inhérents, soit par actualisation et/ou propagation de traits afférents).

Ce principe de contextualité est la base du *principe d'intertextualité* : deux passages de textes différents, si brefs soient-ils, et fussent-ils réduits à la dimension d'un signe, sélectionnent réciproquement, dès qu'ils sont mis côte à côte, des éléments de signification (sèmes). Cet échange surdétermine leur sens (par actualisation et/ou propagation de traits afférents).

À un palier encore supérieur, on peut formuler un principe d'*architextualité*¹⁶ : tout texte placé dans un corpus en reçoit des déterminations sémantiques, et modifie potentiellement le sens de chacun des textes qui le composent.

c) Les parcours intertextuels

De la déontologie qui a présidé à la constitution du corpus dépendent évidemment les parcours interprétatifs au sein de ce corpus, et les modes de l'intertextualité. Les parcours privilégiés s'étendent en premier lieu entre les textes de la même lignée, puis entre ceux du même genre, puis entre les genres d'un même discours. Les relations entre discours ne sont jamais directes, mais toujours médiatisées par des transpositions (cf. par exemple, l'image du discours juridique dans les romans de Balzac).

Les rapports au sein du genre dominant les rapports entre lignées, comme au sein du discours : les relations d'un genre à l'autre supposent

¹⁶ Ou, en hommage pince-sans-rire à Genette, *principe d'Architexte*.

également des transpositions (comprenant les inclusions, citations, etc.). Par exemple, chez Primo Levi, la citation du chant XXVI de l'*Enfer* de Dante au chapitre XI de *Se questo è un uomo* permet le lien avec le poème *Il superstite*, qui cite par allusion le chant XXXIII, dont on retrouve au demeurant des traces dans le même chapitre XI.

Bref, les parcours intertextuels, tant génétiques qu'interprétatifs, s'établissent préférentiellement¹⁷ entre textes de même genre.

2.3. Directions de recherche

a) Problèmes et perspectives en linguistique de corpus

La demande sociale d'une théorie opératoire des genres est croissante, aussi bien pour la linguistique de corpus que pour l'accès aux banques textuelles. L'étude des corpus en situation montre que le lexique, la morphosyntaxe, la manière dont se posent les problèmes sémantiques de l'ambiguïté et de l'implicite, tout cela varie avec les genres. Les systèmes d'analyse et de génération doivent tenir compte de ces spécificités. Les projets de systèmes universels sont ainsi irréalistes, linguistiquement parlant¹⁸.

Pour parvenir à des traitements automatiques spécifiques et efficaces de corpus, il convient de spécifier les fonctionnements propres aux différents genres textuels, pour adapter les stratégies d'interrogation à ces genres.

¹⁷ Les recherches linguistiques portant sur le système de la langue ne tiennent pas compte des variations de genre : tout texte français peut appartenir à leur corpus. Si le degré minimal de l'homogénéité est l'unité de langue, il reste légitime, même hors des recherches linguistiques « unifiantes », de permettre des recherches interdiscursives. Comparons par exemple Rousseau, dans les *Confessions*, à propos d'une dame dont nous taïrons le nom : « 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) ; et dans un tout autre genre, Proudhon, dans la préface de *Qu'est-ce que la propriété*, 1840, p. 157 : « Si j'ose me servir de cette comparaison, un amant est possesseur, un mari est propriétaire ». Jamais une recherche limitée à des textes littéraires n'aurait permis de constater l'ubiquité de ce topos.

¹⁸ Ils reposent en effet sur le préjugé que la langue est homogène et identique à elle-même dans tous les textes et dans toutes les situations de communication.

Le paramétrage préalable des genres peut permettre d'ailleurs de simplifier les traitements automatiques, par exemple en éliminant les ambiguïtés. Ainsi, exemple élémentaire, dans les comptes rendus d'hospitalisation, *pense* ne correspond qu'à la troisième personne ; il en irait tout autrement dans un corpus de lettres aux collègues.

Dans un corpus homogène, connaître la structure du genre peut permettre de simplifier les traitements. Certaines parties des textes peuvent en effet être systématiquement éliminées, pour constituer des sous-corpus pertinents pour une tâche donnée.

Cependant, nous ne privilégions pas exclusivement une démarche déductive. Par un mouvement inverse, la linguistique de corpus qui se développe actuellement doit permettre de refonder ou d'affiner la distinction intuitive et empirique en genres. Les variations morphosyntaxiques selon les genres sont notables. Par exemple, les textes littéraires contiennent trois fois moins de passifs que les autres ; la position de l'adjectif, la nature des déterminants, des pronoms et des temps, l'usage du nombre varient aussi notablement. Ou encore, dans le domaine technique même, les variations sont importantes entre un manuel et une brochure commerciale : au premier, les acronymes, les impératifs, les ellipses de déterminants ; au second, les phrases longues, les pronoms nombreux, etc. (cf. Slocum, 1986). La typologie textuelle peut ainsi apporter du nouveau en morphosyntaxe. Bref, l'étude des normes linguistiques complète utilement celle des règles, et permettra sans doute de préciser leurs conditions d'application.

Pour la caractérisation assistée des textes, on peut envisager trois stratégies : (i) L'indexation contrastive de tous les textes d'un corpus sans préanalyse sémantique, au moyen de logiciels qui permettent de mettre en évidence des pics et dépressions statistiques. (ii) La caractérisation des séquences (paragraphe, par exemple) pertinentes pour une application, et des critères de leur mise en évidence (position¹⁹, indices). (iii) La création de sous-corpus sémantiquement enrichis pour répondre aux objectifs de la tâche en cours.

Encore faut-il, en linguistique de corpus, adopter une déontologie touchant les textes (le British National Corpus partout cité en exemple

¹⁹ Par exemple, le premier paragraphe d'un résumé d'observation médicale a une sémantique ouverte, car il décrit le malade et non le mal.

ne comprend que des extraits, mais pas un seul texte intégral !) et bien entendu les genres. Or les critères typologiques proposés par les recommandations de la *Text Encoding Initiative (TEI)* restent passablement insuffisants dans leur généralité. Ils se limitent pour l'essentiel à des types de « factuelité », d'interactions et de fonctions²⁰.

Par ailleurs, la TEI prévoit huit types de genres fondamentaux, pour le codage desquels sont ménagées des balises de base : prose, poésie (versifiée), théâtre, transcriptions de l'oral, dictionnaire, terminologie ; à quoi s'ajoutent deux genres mixtes. Il est clair qu'il ne s'agit pas de genres, mais de formes de l'expression (prose, catégorie qui transcende tous les discours ; transcription de l'oral), de champs pratiques du discours littéraire (le théâtre, la poésie). Seul le dictionnaire, et la terminologie (au sens de *glossaire*), sont des genres (dans l'acception que nous retenons).

Or, les recherches qu'autorise une banque textuelle seront plus ou moins productives selon qu'elle classe correctement ou non les textes qu'elle recueille ; un exemple permettra d'évaluer l'état de l'art en France et les problèmes à résoudre.

La classification des genres dans la banque Frantext de l'Institut national de la langue française compte 76 catégories dans la version la plus étendue. Sans préjuger de sa refonte annoncée, elle appelle diverses observations, dont nous ne mentionnerons que quelques-unes.

(i) Elle utilise un critère transdiscursif en séparant d'emblée tous les discours et genres en deux catégories (vers *vs* prose). Ce critère d'expression reste improductif, car les critères de genre sont d'abord sémantiques ; mais surtout, il divise les corpus de poésie, de théâtre, et même de traités (la banque en compte un en vers).

(ii) Pour ce qui concerne la division des discours, elle ne distingue pas les sciences et les techniques, alors que ces deux discours n'ont rien de commun, car ils ne relèvent pas des mêmes pratiques. Elle n'établit pas non plus de différence claire entre discours littéraire et non littéraire

²⁰ Les voici : (i) *factuality type* : "fiction, fact, mixed, inapplicable"; (ii) *interaction type* : "none, partial, complete (e.g. face to face conversation), inapplicable"; (iii) *purposes* ("purpose type") : *persuade* : didactic, advertising, propaganda ; *express* : self-expression, confessional ; *inform* : convey information, educate ; *entertain* : amuse, entertain.

(les traités et la presse, rangés dans la littérature, dépendent cependant d'autres discours).

(iii) Quant à la division des genres, si par exemple on conserve dans une seule catégorie les romans et les nouvelles, l'utilisateur ne pourra jamais constituer de sous-corpus permettant de les contraster. Une banque textuelle devrait cependant permettre de vérifier les hypothèses sur les différences génériques.

(iv) Enfin, l'identification et la nomenclature gagneraient à une révision. Par exemple, la classification actuelle confond les catégories d'ouvrages et les genres d'œuvres : ainsi un *mélange* est une indication bibliographique, non un genre²¹.

La classification restreinte de Frantext, pour l'accès par internet, ne distingue plus que dix catégories : correspondance, éloquence, mémoires, pamphlet, récit de voyage, roman, théâtre, poésie, traité, essai. Ces catégories transcendent les frontières entre discours : l'éloquence peut être politique, religieuse ou judiciaire ; les traités et les essais entrent également dans divers discours (philosophique, scientifique) ; la correspondance et les mémoires relèvent soit des écrits intimes, soit du discours littéraire. Enfin, théâtre et poésie sont des champs pratiques du discours littéraire, et comprennent bien évidemment plusieurs genres.

Pour favoriser des recherches différenciées, on peut proposer cinq niveaux de description, qui pourraient figurer dans une description de type de document (*Document Type Definition* ou *DTD*), bien que jusqu'à présent les DTD ne tiennent pas compte du plan sémantique en tant que tel : (i) Le *discours* correspond linguistiquement à un domaine sémantique, et extralinguistiquement à un ensemble de pratiques sociales (ex. religion, littérature, politique)²². (ii) Le *champ pratique*, sous-ensemble du discours, correspond à un ensemble de pratiques spécialisées (ex. liturgie, morale, poésie). (iii) Le *genre* proprement dit (ex. oraison funèbre, sonnet). (iv) Les *sections* : parties et genres inclus

²¹ À cela s'ajoutent des problèmes propres à toute taxinomie : par exemple, sur 76 catégories, 22 sont improductives, et ne correspondent qu'à une seule œuvre. Ou encore, les catégories mixtes comme *Traité ou Essai* pourraient être évitées, surtout quand on dispose déjà de *Traité* et *Essai*.

²² Quelles que soient par exemple les qualités littéraires de Bossuet ou de Paul-Louis Courier, elles ne suffisent pas à inclure leurs écrits dans la littérature.

(ex. : titre, chapitres). (v) Les *configurations*, comme l'exemple, l'anecdote, la description.

Le discours et le champ pratique sont à coder au niveau du corpus et de l'en-tête du corpus (<teiCorpus>), et le genre proprement dit au niveau de l'unité de texte et de l'en-tête du texte (<teiHeader>). Pour les niveaux inférieurs, un premier pas pourrait être de distinguer les parties (comme le premier paragraphe) et les fonctions de ces parties (comme l'introduction, articulée dans le premier paragraphe) ; un second, de coder, en fonction des besoins, les unités sémantiques. Cela demande évidemment une préanalyse, car ces unités (par exemple, les molécules sémiotiques) ne sont pas codables *a priori*, car elles ne sont pas isolables par les méthodes de segmentation classiques.

b) Genres et médiations symboliques

Les genres sont des moyens (i) de la médiation symbolique (au sens proposé par Clifford Geertz, 1972) qui articule l'individuel et le social, et (ii) de la médiation sémiotique, celle qui sépare le physique du représentationnel.

La poétique généralisée engage dans son ensemble la médiation symbolique : le genre partage tout à la fois le caractère public de l'action individuelle socialisée et de la norme sociale où elle prend place. Ne pas encore appartenir à la société, comme l'enfant qui apprend à parler, ou en être rejeté, comme l'aliéné, c'est utiliser des genres idiosyncrasiques. Relativement à leurs genres, les performances sémiotiques sont sanctionnées comme convenantes ou non, qu'il s'agisse d'écrire des rapports d'activité ou de danser le *paso-doble*.

Partout, le mélange des genres a été suspecté comme moralement répréhensible, signe de duplicité ou de confusion, voire de subversion, et les œuvres durablement subversives restent en général réputées inclassables. Ainsi, peut-être, le roman était-il considéré comme immoral (on l'interdisait aux jeunes filles) car il parlait certes d'amour, mais aussi parce que sa structure rhapsodique lui permettait de subvertir les normes, même les siennes.

Sans prétendre qu'il n'y a de lois que du genre, celui-ci demeure quand même l'instance historique de normalisation de la langue — dont les règles ne dépendent pas directement des situations sociales ou politiques déterminées, alors que les normes des genres s'y conforment

(ou les conforment) par diverses médiations. En d'autres termes, le genre pourrait être considéré comme le lieu sémiotique de l'intersubjectivité en tant qu'elle est médiatisée par la Loi.

L'étude du genre revêt son plus grand intérêt quand elle permet de percevoir la singularité des textes — de même que les normes sociales constituent le fond qui permet de comprendre les actions individuelles. C'est dans l'usage singulier des genres que se constitue la personnalité, alors que les genres dessinent « en creux », par les positions énonciatives et interprétatives qu'ils codent, la personne comme ensemble de rôles sociaux²³.

Dans ses écrits pédagogiques, Schleiermacher a parfaitement saisi les enjeux de l'apprentissage des genres. Si apprendre c'est abandonner les genres idiosyncrasiques, « le langage malgré tout est quelque chose de partagé en commun et la médiation de la communauté, et [les jeunes enfants] abandonnent rapidement leurs inventions linguistiques pour s'intégrer dans la vie de la langue commune » (1876 : 231 ; trad. Berner, 1995 : 207). En pratique, et tout particulièrement en didactique, la théorie des genres trouve ainsi des applications immédiates, puisque la langue ne se perçoit qu'à travers des genres.

BIBLIOGRAPHIE

- Adam J.-M. 1990, *Eléments de linguistique textuelle*, Paris : Mardaga.
 Adam J.-M. 1992, *Les textes : types et prototypes*, Paris : Nathan.
 Adam J.-M. 1999, *Linguistique textuelle. Des genres de discours aux textes*, Paris : Nathan.
 Bakhtine M. [1952-53] 1984, *Esthétique de la création verbale*, Paris : Gallimard.
 Benveniste E. 1966, *Problèmes de linguistique générale*, Paris : Gallimard.

²³ Entre le sociologisme bakhtinien et l'individualisme wittgensteinien, il faut faire retour à l'histoire de la dialogique : Schleiermacher, Feuerbach, Kierkegaard, Stirner. Elle a été obscurcie passablement par les débats entre Apel et Habermas, et l'on n'a pas pris garde que la dialogique de Bakhtine est une reprise tardive des principaux thèmes de la dialectique de Schleiermacher.

- Berner C. 1995, *La philosophie de Schleiermacher*, Paris : Cerf.
- Biber D. 1988, *Variations across Speech and Writing*, Cambridge : CUP.
- Biber D. 1992, « The multi-dimensional approach to linguistic analysis of genre variation : an overview of methodology and findings », *Computers and the Humanities*, 26 (5-6), 331-345.
- Biber D. 1993 a, « Using Register-Diversified Corpora for General Language Studies », *Computational Linguistics*, 19 (2), 219-241.
- Biber D. 1993 b, « Co-occurrence Patterns among Collocations : A Tool for Corpus-Based Lexical Knowledge Acquisition », *Computational Linguistics*, 19 (3), 531-538.
- Bommier-Pincemin B. 1999, *Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*, Thèse de Doctorat, Linguistique, Université de Paris IV (Sorbonne).
- Bouquet S. 1998, « Linguistique textuelle, jeux de langage et sémantique du genre », *Langages*, 129, 112-123.
- Burnard L., Sperberg-McQueen, C.M. 1996, « La TEI simplifiée : une introduction au codage des textes électroniques en vue de leur échange », *Cahiers GUTenberg*, 24, 23-151.
- Combe D. 1992, *Les genres littéraires*, Paris : Hachette.
- Coseriu E. 1981, *Textlinguistik Ø Eine Einführung*, Tübingen : Narr.
- Fowler A. 1982, *Kinds of Literature. An Introduction to the Theory of Genres and Modes*, Oxford : Oxford University Press.
- Geertz G. 1972, *The Interpretation of Cultures*, New York : Basic Books.
- Genette G. 1986, « Introduction à l'architexte », in Genette G. et Todorov T., 1986, *Théorie des genres*, Paris : Seuil, 89-16.
- Genette G. (éd) 1986, *Théorie des genres*, Paris : Seuil.
- Guilhaumou J., Maldidier D. 1979, « Courte critique pour une longue histoire — L'analyse du discours ou les (mal)leures de l'analogie », *Dialectiques*, 26, 7-23.
- Heinemann W., Viehweger D.

- 1991, *Textlinguistik— Eine Einführung*, Tübingen : Niemeyer.
- Illouz G. *et al.* 1999, « Maîtriser les déluges de données hétérogènes », in Condamines A. *et al.* (éds.), *Corpus et traitement automatique des Langues : pour une réflexion méthodologique*, Actes de l'atelier thématique TALN, Cargèse, 37-46.
- Lafon P., Salem A. 1983, « L'inventaire des segments répétés d'un texte », *M.O.T.S.*, 6, 161-177.
- Péry-Woodley M.-P. 1995, « Quels corpus pour quels traitements automatiques ? », *Traitement Automatique des Langues*, 36 (1-2), 213-232.
- Pincemin B. 1999, « Construire et utiliser un corpus : le point de vue d'une sémantique textuelle interprétative », in Condamines A. *et al.* (éds.), *Corpus et traitement automatique des Langues : pour une réflexion méthodologique*, Actes de l'atelier thématique TALN, Cargèse, 26-36.
- Raible W. 1988, « Qué son los géneros ? Una respuesta desde el punto de vista semiótico y de la lingüística textual », in Garrido Gallardo, Miguel A. (éds.), *Teoría de los géneros literarios*, Madrid : Arco/Libros S.A., 303-339.
- Raible W. 1995, « Arten des Kommentierens — Arten der Sinnbildung — Arten des Verstehens. Spielarten der generischen Intertextualität », in Assmann Jan, Gladigow, Burkhard (éds.), *Text und Kommentar*, Munich : Fink, 51-73.
- Rastier F. 1987, *Sémantique interprétative*, Paris : P.U.F.
- Rastier F. 1989, *Sens et textualité*, Paris : Hachette.
- Rastier F. 1991, *Sémantique et recherches cognitives*, Paris : P.U.F.
- Rastier F. 1992, « Réalisme sémantique et réalisme esthétique », *TLE*, 10, 81-119.
- Rastier F. 1994, « Le problème du style pour la sémantique du texte », in Molinié G. et Cahné P. (éds.), *Qu'est-ce que le style ?*, Paris : P.U.F.
- Rastier F. 1995, « Communication ou transmission », *Césure*, 8, 151-195.
- Rastier F. 1996, « Pour une sémantique des textes-questions d'épistémologie », in *Textes & Sens*, François Rastier (dir.), Paris : Didier Erudition, 9-35.
- Schaeffer J.-M. 1989, *Qu'est-ce qu'un genre littéraire ?*, Paris : Seuil.

- Schleiermacher F. 1876, *Schleiermachers Pädagogische Schriften*, éd.C. Platz, Langensalza, Beyer and Söhne.
- Slocum J. 1986, « How one might automatically identify and adapt to a sublanguage », in Grihman R. et Kittredge R. (éds.), *Analysing Language in restricted domains*, Hillsdale : Erlbaum, 195-210.
- Sueur J.-P. 1982, Pour une grammaire du discours, *Mots*, 5, 145-185.
- Swales J. M. 1990, *Genre Analysis. English in Academic and Research Settings*, Cambridge : Cambridge University Press.
- Thlivitis T. 1998, *Sémantique Interprétative Intertextuelle : assistance informatique anthropocentrée à la compréhension de textes*, Thèse de Doctorat, Informatique, Université de Rennes 1.