

*François Rastier**

La sémantique des textes — **concepts et applications**

Cette étude présente synthétiquement un point de vue argumenté par ailleurs dans diverses publications¹. Je commencerai par rappeler les principales approches, pour formuler ensuite des propositions. Je mentionnerai ensuite les besoins nouveaux qu'une sémantique des textes doit contribuer à satisfaire, puis les remaniements épistémologiques qu'elle engage.

1. Problématiques

1.1. La linguistique comme science des textes

Admettons tout d'abord que les textes sont l'objet empirique de la linguistique. Isoler des phrases, *a fortiori* des mots, résulte d'une décision méthodologique ; de même pour abstraire des phrases et des significations, formes typiques décrites par la linguistique.

Il n'y a pas en droit deux linguistiques, l'une qui serait centrée sur la morphosyntaxe, et l'autre sur le texte : ce sont seulement deux paliers de description complémentaires.

Pour avoir négligé cela, la linguistique restreinte ne s'est pas donné les moyens de penser sa restriction. Quand par exemple Fauconnier estime que *Elle a de bonnes jambes* est une phrase ambiguë, parce qu'elle peut aussi référer à une scène de cannibalisme, il souligne invo-

¹ Malgré son caractère personnel, il doit beaucoup aux débats conduits avec divers collègues et amis, notamment au sein de l'équipe *Sémantique des textes* de l'Université Paris-Sorbonne.

* *François RASTIER*
Directeur de Recherche (INaLF-CNRS)
Centre de Linguistique Française
Université de Paris-Sorbonne
1, rue Victor Cousin
F-75230 Paris Cédex 05

lontainement combien de faux problèmes peut susciter l'étude de phrases sans attestation, sans contexte, sans texte, sans situation.

De façon converse, la sémiotique discursive a tenté jadis de prendre son autonomie par rapport à la linguistique restreinte, pour se constituer en discipline autonome. Elle a connu très vite une involution spéculative qui l'a très vite coupée de la description linguistique et des textes eux-mêmes.

La légitimité de l'extension aux textes de la description linguistique reste contestée de diverses façons, car la diversité et l'hétérogénéité des textes ne laissent pas d'inquiéter.

On peut réagir par dénégation : Molino souligne ainsi "l'hétérogénéité des textes, pour lesquels nous ne croyons pas [...] qu'il existe de science unique", (il la nomme plus loin *Science Magnifique des Textes* (1989, p. 40)). Or, la relation médiante des textes divers domaines d'objectivité et même divers mondes n'entraîne pas que la linguistique devienne la Science Universelle (cf. *contra*, Molino, p. 42). L'hétérogénéité (comme diversité) de l'objet est le lot de toute science descriptive, et une condition, plutôt qu'un obstacle, à son existence. De l'hétérogénéité des textes on ne peut conclure à l'impossibilité d'une science des textes, mais bien au contraire à sa nécessité.

Une autre stratégie, accueillante celle-ci, est suggérée par Van Dijk, quand il affirme : "L'étude scientifique des textes ne peut être que d'ordre interdisciplinaire" (1981, p. 90). L'éclectisme jouit certes de solides traditions universitaires, mais la notion de science interdisciplinaire mériterait d'être examinée. La confusion épistémologique qui a régné en sémiotique, en IA, en sciences cognitives invite en effet à la prudence. Par exemple l'étude du discours juridique n'entraîne aucunement qu'un mixte de linguistique et de droit se constitue à cet effet. Que les connaissances nécessaires à l'interprétation d'un texte aient été produites par d'autres disciplines que la linguistique, cela n'empêche aucunement la mise en œuvre d'une sémantique interprétative.

Nous estimons que la linguistique est la sémiotique des langues et des textes, et comme telle fait partie de la sémiotique des cultures.

1.2. Penser la diversité des textes

Chaque texte procède d'un genre, et chaque genre est relatif à un discours (politique, religieux, etc.). Les normes génériques et discursives

sives se surimposent au système fonctionnel de la langue ; en d'autres termes, on appelle *langue* le résultat d'une description systématique qui fait abstraction de ces normes. Elles n'en ont pas moins une incidence à tous les niveaux : norme orale et norme écrite diffèrent considérablement dans leur syntaxe, et même parmi les genres écrits, les règles ou du moins les usages syntaxiques varient (par exemple, les substantifs épithètes, communs dans les genres journalistiques, sont rarissimes dans le roman).

Les définitions lexicographiques gardent trace du contexte, par les constructions et les exemples, et du texte par les indicateurs de domaines (ex. *cuis.* pour cuisine) voire de discours (*poét.* pour poétique). Cette information indispensable pourrait être encore systématisée.

Alors que la recherche a beaucoup progressé aux paliers du mot et de la phrase, le développement d'une linguistique du texte paraît une entreprise prioritaire. La dimension textuelle a été sous-estimée en linguistique, ou déléguée à des disciplines voisines.

Cependant son étude permet d'ouvrir l'espace des normes et le lien entre les textes et les pratiques sociales où ils sont produits. Certes, la description linguistique des textes a connu une brève période d'essor voici vingt ans, mais les recherches ont tourné court, faute sans doute de théories linguistiques adéquates et de possibilités institutionnelles et académiques.

1.3. Les approches sémantiques des textes

On peut distinguer deux sortes principales de théories du texte : sémiotiques et linguistiques. Les théories sémiotiques les plus connues considèrent le niveau linguistique comme une variable de surface (Greimas), et se sont naguère constituées en opposition à des théories linguistiques restreintes à la morphosyntaxe. Avec l'extension du champ des études linguistiques, le principe d'une sémiotique discursive autonome devient de plus en plus difficile à défendre. Pour leur part, les théories linguistiques se laissent ramener à quatre sortes.

a) Les théories issues de la *sémantique formelle*, dont la plus connue est celle de Kamp. Leur complexité technique est notable. En revanche, elles ne se prêtent pas à une description comparative des textes. Par exemple, le concept de genre n'a pu être reformulé dans ce cadre. Les descriptions, de fait, ne s'étendent pas au-delà du paragraphe.

b) Les théories *pragmatico-énonciatives* issues de l'analyse du discours. Elles se sont attachées à identifier des marques de l'énonciation, comme les indexicaux, à classer les actes de langage, à étudier les structures argumentatives des textes qui s'y prêtent. Ces théories se prêtent à l'analyse des interactions microsociologiques, notamment à l'étude des conversations. Elles sont liées à certains genres oraux, mais ne se sont pas prêtées à la typologie des textes, notamment quand elles revendiquent une définition transcendantale de la communication (cf. Grice et sa référence au kantisme, Sperber et Wilson et leur postulation d'un principe *a priori* de la pertinence communicative).

Ces deux premiers types de théories sont compatibles entre elles, du moins par le cousinage de la pragmatique et de la sémantique vériditionnelle au sein du positivisme logique.

c) Les théories *sémantiques issues du courant saussurien* (Coseriu, le premier Greimas). Leurs domaines de prédilection sont la sémantique lexicale, la théorie des isotopies et l'analyse narrative. Dans ce cadre général, le programme de la sémantique unifiée vise à intégrer les trois paliers de description linguistique (mot, phrase, texte) comme des paliers de complexité différents, mais qui peuvent être décrits en utilisant les mêmes concepts de base (comme le sème).

Par leurs recherches sur l'interprétation, ces sémantiques se sont éloignées de l'objectivisme qui a marqué jadis le structuralisme, et se sont rapprochées de certaines théories herméneutiques, comme l'herméneutique matérielle de Szondi, qui est une forme élaborée mais restée lacunaire de l'herméneutique philologique.

d) Les théories "*rhétoriques*" issues de l'étude des langues de spécialité (Swales, Bhatia). Liées pour la plupart à la tradition anthropologique (Firth, puis Halliday), elles sont soucieuses de décrire la diversité des textes en fonction des pratiques socialisées, et ont accumulé des observations précieuses dans des domaines comme le langage juridique.

On pourrait assister à une convergence de ces trois derniers types de recherche. Cela suppose d'une part un refus commun de l'objectivisme et de l'immanentisme, et corrélativement le passage d'une problématique logico-grammaticale que la sémantique a héritée du voisinage millénaire de la grammaire et de la logique au sein du trivium à une problématique rhétorique-herméneutique. Peu importe qu'elle soit moins

scientiste ou moins scientifique, comme on voudra : par sa prise en compte de la complexité textuelle, elle a déjà montré une capacité descriptive supérieure, ce qui à moyen terme permettra de trancher pour des raisons d'efficacité.

Un état de l'art reste à faire, une synthèse à rechercher. Parallèlement, il faut confronter, évaluer, conjoindre les diverses méthodes d'analyse utilisées pour décrire les textes scientifiques et techniques, et les textes littéraires ou mythiques. Une typologie textuelle doit être éditée avec des concepts issus de la sémantique (puisque les structures textuelles sont essentiellement sémantiques). La description linguistique des genres scientifiques et techniques n'a pas encore fait l'objet d'un travail de recherche systématique, du moins en France, à la différence par exemple du Danemark ou de la RFA (*Fachtextlinguistik*)².

2. Tâches et propositions d'une sémantique des textes

2.1. Champ d'étude et principes de recherche de la sémantique des textes

Voici quelques orientations que nous assumons. Nous prenons pour objet d'étude les textes français modernes, considérés en particulier dans leurs structures transphrastiques cela impose la constitution de corpus dans le respect des pratiques et des situations, la typologie des discours et des genres, et l'analyse sémantique, car les structures transphrastiques sont sémantiques pour l'essentiel.

Sans exclure au besoin les monographies, la perspective adoptée est *sérielle*. En d'autres termes, elle doit s'appuyer sur des corpus importants, auxquels les moyens informatiques donnent à présent un accès aisé.

² En France, les chercheurs manquent encore d'un corpus de référence numérisé et suffisamment diversifié. Certes, la base Frantext de l'Institut National de la Langue Française a l'avantage de recueillir aussi bien des textes littéraires que des textes techniques et scientifiques (en nombre insuffisant mais croissant). Elle est la seule, dans le domaine français, à permettre un accès conjoint à ces textes, de manière à les contraster. Mais ce corpus n'a pas été exploité systématiquement dans sa variété générique (ce qui ne recoupe pas les études lexicographiques en cours). Par exemple, les romans et les recueils de nouvelles sont indexés ensemble ; une classification nouvelle ne serait-elle pas nécessaire ? Dans une étude thématique sur les sentiments (1995), nous avons pu ainsi établir que les sentiments du roman diffèrent de ceux de l'essai.

Elle est prioritairement *synchronique*, mais n'exclut pas des recherches diachroniques ponctuelles (par exemple, sur l'évolution thématique d'un genre).

Elle s'appuie pour l'essentiel sur des textes *écrits* — sans négliger les discours transcrits, comme par exemple les entretiens d'experts, ou les discours politiques.

Pour les textes écrits, la situation des typologies est contrastée. Distinguons grossièrement les textes pratiques et les textes mythiques. Les textes littéraires sont fort étudiés, mais par la critique plutôt que par la linguistique, et un gros travail reste à faire pour unifier les conceptualisations linguistiques de la stylistique, de la rhétorique et de l'herméneutique. D'autre part, la critique contemporaine a prêté fort peu d'attention à la question des genres, car elle doit sans doute à la tradition romantique le goût des œuvres d'exception. La poétique s'efforçait traditionnellement de classer les genres littéraires ; à mesure que le concept de genre perdait de l'importance pour les créateurs et les critiques eux-mêmes, l'entreprise typologique passait à l'arrière plan.

Les textes scientifiques et techniques sont naturellement étudiés dans une tout autre perspective, pour l'essentiel en terminologie, c'est-à-dire à un palier d'analyse qui ne dépasse pas le syntagme, et n'est donc pas celui du texte. Mais l'étude comparative de leurs genres reste rare.

On sait que la *méthode comparative* issue de la linguistique historique n'a aucunement démerité. Trois entreprises typologiques simultanées sont à concevoir selon ses principes : (i) Celle des *genres* traite de la diversité externe des textes (cf. langues). (ii) Celle des *textes* traite de la diversité interne des genres (cf. dialecte ; cette typologie transcende les frontières linguistiques (ex. les motifs en folkloristique). (iii) 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, mais aussi d'autres qui ne sont pas nommées, comme les molécules sémiques (structures stables de traits sémantiques, qui n'ont pas nécessairement de lexicalisation privilégiée).

En bref, les tâches principales d'une sémantique des textes se disposent sur trois lignes convergentes :

(i) Elaboration d'une sémantique unifiée pour les trois principaux paliers de description (mot, phrase, et texte). (ii) Elaboration de catégories pour un typologie des textes (littéraires et mythiques, scientifiques et techniques). (iii) Développement de ces théories descriptives en liaison avec les traitements automatiques des textes.

Voici donc quelques propositions théoriques (développées par ailleurs, 1989, 1994), qui permettent de préciser ces objectifs.

2.2. Les composantes

Pour établir le cadre conceptuel d'une typologie des textes, on peut concevoir la production et l'interprétation des textes comme une interaction non-séquentielle des composantes autonomes : thématique, dialectique, dialogique et tactique.

1. La *thématique* rend compte des contenus investis, c'est-à-dire du secteur de l'univers sémantique mis en œuvre dans le texte. Elle en décrit les unités. Par analogie, et bien qu'elle ne décrive pas spécifiquement le lexique, on peut dire qu'elle traite du "vocabulaire" textuel, dont nous détaillerons plus loin les unités (molécules sémiques, faisceaux d'isotopies, etc.).

2. La *dialectique* rend compte des intervalles temporels dans le temps représenté, de la succession des états entre ces intervalles, et du déroulement aspectuel des processus dans ces intervalles.

3. La *dialogique* rend compte des modalités, notamment énonciatives et évaluatives, ainsi que des espaces modaux qu'elles décrivent. Dans cette mesure, elle traite de l'énonciation représentée.

4. La *tactique* rend compte de la disposition séquentielle du signifié, et de (l'ordre linéaire ou non) selon lequel les unités sémantiques à tous les paliers sont produites et interprétées.

Chaque unité sémantique, aux différents paliers d'analyse, peut être caractérisée en fonction de ces quatre composantes : en bref, située par sa position dans l'univers sémantique, par un repérage identitaire, modal, temporel, ou distributionnel.

Seule une décision méthodologique peut isoler ces quatre composantes. Elles ne sont pas indépendantes, mais en interaction. Dans la quasi-totalité des textes, chacune des composantes est *simultanément* en interaction avec les autres. Aucune directionnalité n'est imposée à ce dispositif *hétérarchique*.

La sémantique du texte a notamment pour objectif de décrire cette interaction. Elle doit le faire en outre selon les quatre ordres de la description linguistique (paradigmatique, syntagmatique, référentiel et herméneutique). En effet, chacun des types de repérage d'une unité sémantique que permettent les quatre composantes est susceptible de quatre sortes de description. On peut décrire ainsi une forme sémantique quelconque : par rapport à un répertoire de formes, et l'on en fait alors une description paradigmatique ; comme une part d'un enchaînement de formes (description syntagmatique) ; comme le résultat d'un parcours de constitution ou de reconstitution (description herméneutique) ; par rapport à des formes non linguistiques (description référentielle).

Chacune des composantes est par ailleurs susceptible de connaître les trois degrés de systématisme : système fonctionnel, normes sociolectales, normes idiolectales.

A. *La thématique*

a) La notion de *thème* a divers usages, en critique thématique, en linguistique d'inspiration pragoise (par opposition à *rhème*), en analyse du discours (*topic* par opposition à *focus*). La sémantique descriptive peut clarifier cette notion en définissant le thème comme un groupement structuré de sèmes. Leur statut et leur nombre, les relations établies entre eux, tout cela varie selon les genres. Selon le statut de ces sèmes, il convient de distinguer les thèmes génériques et les thèmes spécifiques.

b) Un *thème générique* est défini par un sème ou une structure de sèmes génériques récurrents. Cette récurrence définit une isotopie ou un faisceau d'isotopies génériques (c'est-à-dire un groupement de sèmes génériques co-récurrents). Les isotopies génériques, et particulièrement domaniales, déterminent le « sujet » (*topic*) du texte, ou plus précisément elles induisent les impressions référentielles dominantes. Par exemple, les textes techniques étant contraints par un domaine d'application, ils ne manifestent qu'un domaine sémantique, alors que les textes littéraires peuvent en juxtaposer plusieurs. Selon les types de classes sémantiques qu'ils manifestent, on peut distinguer quatre sortes.

(i) Les thèmes taxémiques sont portés par les membres d'un même *taxème*.

(ii) Les thèmes domaniaux sont portés par les membres d'un même *domaine* sémantique. Les langues de spécialité décrivent et structurent des domaines sémantiques.

(iii) Les thèmes dimensionnels sont portés par les membres d'une même *dimension* sémantique. Les dimensions évaluatives ou thymiques, les tons, les espaces modaux, les plans temporels ou *chronotopes*, définissent dans les textes autant de sortes d'isotopies dimensionnelles.

(iv) Les thèmes liés à un champ sémantique sont plus variables, puisque les champs ne sont pas des classes de langue. Les plus faciles à identifier sont ceux qui sont codifiés (topique littéraire) ou définitoires d'un corpus.

c) Les *thèmes spécifiques* sont des groupements récurrents de sèmes spécifiques. Nous avons proposé de nommer *molécules sémiques* ces groupements. Ils ne sont pas nécessairement dépendants d'une lexicalisation particulière. En règle générale, dans les textes techniques, les molécules sémiques ont cependant une lexicalisation privilégiée, voire exclusive, car les disciplines techniques répugnent à l'équivocité.

B. *La dialectique*

Comme elle traite des intervalles de temps représenté et des évolutions qui s'y déroulent, la dialectique rencontre les théories du récit, qui ne participent pas de théories générales du texte, mais dont les acquis doivent être sauvegardé. La dialectique est définie à deux niveaux, dont nous rappelons simplement les principaux concepts. Le premier niveau, dit *événementiel*, apparaît dans tous les textes structurés par une composante dialectique. Ses unités de base sont les *acteurs*, les *rôles* et les *fonctions* (au sens de Propp, soit un type d'action représentée).

1. — Le niveau événementiel

Un *acteur* peut être défini comme une classe d'actants : il est constitué par totalisation d'actants anaphoriques des périodes (au palier inférieur du texte, mésosémantique). Dans la période, les actants peuvent être nommés ou recevoir diverses descriptions (définies ou non). Chaque dénomination ou description lexicalise un ou plusieurs sèmes de l'acteur. L'acteur se compose de trois sortes de structures sémiques :

— Sa molécule sémique est constituée des sèmes spécifiques de ces actants.

— Outre leur molécule sémique, les acteurs comportent par ailleurs des sèmes génériques, ici /humain/ ; mais dans certains genres les acteurs ne comprennent pas nécessairement le trait /animé³. Les actants constituants d'un acteur peuvent n'avoir pas de sèmes communs stables.

— Enfin, un acteur comprend des sèmes afférents, qui sont des *rôles*. Ce sont des cas sémantiques associés aux actants qu'il subsume. L'ensemble des rôles d'un acteur définit sa sphère interactionnelle. Elle évolue avec les intervalles temporels. On peut classer les acteurs selon le nombre et la nature des rôles qui les caractérisent.

b) Les *fonctions* sont des interactions typiques entre acteurs : ce sont des classes de processus. Comme les acteurs, elles sont définies par une molécule sémique et des sèmes génériques : par exemple, le *don* est une fonction irénique (de transmission, à valence ternaire), le *défi* une fonction polémique (d'affrontement, à valence binaire). Les fonctions correspondent à des *valences* actorielles. Les fonctions peuvent se grouper en syntagmes fonctionnels (ex : *échange* se compose de deux transmissions, *affrontement* d'une attaque et d'une contre-attaque). Les *scripts* sont des syntagmes fonctionnels typés par une isotopie générique de champ ou de domaine.

L'ensemble du niveau événementiel peut être représenté par des graphes macrosémantiques. En bref, leurs nœuds représentent les acteurs et les fonctions, et leurs liens représentent les rôles.

2. — Le niveau agonistique

Ce niveau est hiérarchiquement supérieur au niveau événementiel. Comme il n'apparaît que dans les textes mythiques, et que les applications techniques concernent pour l'essentiel des textes pratiques (notamment technologiques), sa présentation sera succincte. Ses unités de base sont les *agonistes* et les *séquences*.

a) Un *agoniste* est un type constitutif d'une classe d'acteurs. Dans les textes mythiques au moins, il est fréquent que les acteurs relevant d'un même agoniste soient indexés sur des isotopies génériques différentes, voire dans des univers et des mondes divers, mais se trouvent cependant en relation métaphorique. Par exemple, dans *Toine* de Maupassant, la Vieille est un acteur sur l'isotopie humaine, qui se trouve en relation

³ Le concept d'*acteur* ainsi défini n'a pas de rapport privilégié avec la notion de personnage. Les personnages d'une épopée sont par exemple des acteurs, mais au même titre que les pièces et les outils d'une notice de montage.

métaphorique avec le Coq sur l'isotopie animale, le Vent du large sur l'isotopie météorologique, et la Mort sur l'isotopie métaphysique.

Les agonistes sont définis d'une part par les composants invariants ou équivalents des molécules sémiques de leurs acteurs, et aussi par les rôles identiques ou équivalents dans les sphères interactionnelles de ces acteurs. À la différence des acteurs, ils ne comportent pas nécessairement de sèmes génériques qui les indexent sur des isotopies.

b) Les *séquences* sont définies par homologation de syntagmes fonctionnels isomorphes. Comme les syntagmes qu'elles homologuent occupent des positions différentes dans le temps dialectique, les séquences sont ordonnées par des relations de logique narrative (présuppositions) mais non chronologiques.

La distinction entre niveau événementiel et niveau agonistique permet une typologie dialectique et conduit à détailler la notion de récit. En général, la composante dialectique des textes pratiques ne comporte que le niveau événementiel alors que les textes de fiction le redoublent par un niveau agonistique.

C. La dialogique

La dialogique rend compte de la modalisation des unités sémantiques à tous les paliers de complexité du texte.

a) Un *univers* est l'ensemble des unités textuelles associées à un acteur ou à un foyer énonciatif : toute modalité est relative à un site (un univers) et un repère (un acteur). Par exemple, quand le narrateur de *la Cousine Bette* parle d'une *mauvaise bonne action*, 'bonne' renvoie à l'univers de deux acteurs, et 'mauvaise' à son propre univers.

Chaque univers est susceptible de se diviser en trois *mondes* : (i) Le *monde factuel* est composé des unités comportant la modalité assertorique ; (ii) Le *monde contrefactuel*, des unités comportant les modalités de l'impossible ou de l'irréel ; (iii) Le *monde du possible*, des unités comportant la modalité correspondante⁴. Mondes et univers peuvent évoluer selon les intervalles de temps dialectique.

⁴ Nous ne prétendons pas pour autant que la théorie logique des mondes possibles puisse s'appliquer à la sémantique des textes. Le monde du possible n'est pas un monde possible au sens de Leibniz — qui dans sa *Théodicée* ne voyait aucun obstacle à ce que les romans de Mademoiselle de Scudéry se réalisent un jour.

Dans le cadre ainsi esquissé, le récit relève de la dialectique et la narration de la dialogique. Ces deux composantes sont évidemment liées. Par exemple, une fonction *contrat* consiste en un échange de processus de transmission situé dans les mondes du possible associés aux acteurs contractants : en tant qu'échange de promesses, il relève de la dialectique, mais ces promesses elles-mêmes, en tant qu'unités modalisées, relèvent de la dialogique.

La dialogique fonde la typologie des énonciateurs représentés. En général, les textes d'instructions techniques ne comportent qu'un foyer énonciatif et un foyer interprétatif non nommés, et se présentent souvent comme une suite de phrases à l'impératif. Les articles scientifiques en revanche multiplient les énonciateurs délégués par le biais de citations ou d'allusions, et précisent les foyers interprétatifs par de rituels cliniques d'œil aux initiés.

D. *La tactique*

Cette dernière composante rend compte de la disposition linéaire des unités sémantiques à tous les paliers. La linéarité du signifié entretient certes des rapports étroits avec celle du signifiant, elle ne se confond cependant avec elle à aucun des paliers : rappelons les morphèmes à signifiant zéro, les ellipses, les unités macrosémantiques comme les isotopies ou les séquences, dont les manifestations sont discontinues. En outre, des unités sémantiques de toute dimension sont restituées par des inférences : elles appartiennent alors au sens du texte, sans qu'on puisse pour autant leur assigner une position dans sa linéarité.

La position relative des unités du signifiant peut être évidemment utilisée comme indice des relations distributionnelles entre unités du signifié. Bref, quand le texte a une structure dialectique, et quand le temps représenté est linéaire, les positions des unités sémantiques ne s'ordonnent nécessairement ni selon la linéarité du signifiant, ni selon celle du temps représenté.

E. *Les formes sémantiques*

(i) *Unités textuelles et formes sémantiques.* — Au palier textuel comme aux autres, les unités résultent de segmentations et de catégorisations sur des formes et des fonds sémantiques, que l'on peut désigner du nom général de *morphologies*. Leur étude se divise en trois sections : liens entre fonds (dans le cas par exemple des genres qui

comportent plusieurs isotopies génériques), liens entre formes (dont il faut faire une description différentielle), et surtout liens des formes aux fonds (cruciaux pour l'étude de la perception sémantique).

Selon les composantes, ces morphologies sémantiques peuvent faire l'objet de diverses descriptions. Par exemple, rapporté aux quatre composantes, un groupement stable de traits sémantiques (ou molécule sémique) peut être décrit comme thème, comme acteur, comme but ou source d'un point de vue modal, comme place dans la linéarité du texte. En outre, à chaque composante correspondent des types d'opérations productives et interprétatives.

(ii) *Description dynamique*. — Traditionnellement, la linguistique textuelle a étendu au texte les procédures de segmentation issues de la morphosyntaxe. Cependant, il ne s'agit pas d'interpréter des unités qui se donneraient comme discrètes ou déjà discrétisées, mais de les discrétiser comme des moments de parcours interprétatifs. Si la description statique peut convenir à certaines applications, en didactique par exemple, une description plus fine doit restituer l'aspect dynamique de la production et de l'interprétation des textes. La première étape consiste à décrire les dynamiques de ces fonds et de ces formes : par exemple, la construction des molécules sémiques, leur évolution, et leur dissolution éventuelle.

Ces dynamiques et leurs optimisations sont paramétrées différemment selon les genres et les discours. D'une part les formes et les fonds sont constitués et reconnus par rapport à des présomptions, et comparés à des stéréotypes différents selon les pratiques. En outre, les contrats de production et d'interprétation qui sont associés aux genres et aux discours norment le parcours de ces morphologies. Pour en tenir compte, la sémantique des textes doit adapter ses descriptions aux régimes morphologiques des discours et des genres.

Enfin, la perception des formes et celle des fonds posent des problèmes différents : celle des fonds semble liée à des rythmes, celle des formes à des contours (dont les contours prosodiques peuvent présenter une image).

2.3. L'interaction entre composantes sémantiques et les genres

Nous venons de définir des unités et des relations dans les quatre composantes : elles restent analysables en sèmes. Nous avons procédé dans une perspective de production ou d'interprétation, qui en reste à la sémantique : les objets décrits sont tous constitués d'unités et de relations sémantiques organisant des fonds ou des formes.

Chacune des composantes peut être la source de critères typologiques. On peut classer les textes selon les critères les plus divers, nombre et nature des isotopies génériques, représentation de l'énonciation, etc. Aussi proposons-nous cette hypothèse : sur le plan sémantique, les *genres* sont définis par des *interactions normées* entre les composantes que nous venons de décrire.

a) La thématique d'un texte peut d'abord se décrire par son étendue, c'est-à-dire la part de l'univers sémantique mise en jeu, comme par ses restrictions de fait. Rapportée à la tactique, cette étendue est linéarisée en isotopies ou spécifiée en molécules sémiques, les unes comme les autres pouvant être caractérisées par leur position relative dans le texte. L'étendue thématique peut varier au cours du texte.

Rapportées à la dialectique, les molécules sémiques deviennent, par adjonction de traits casuels afférents, des acteurs ou des fonctions, voire — après homologation — des agonistes ou des séquences. Rapportées à la dialogique, les isotopies et molécules sémiques se placent dans des espaces modaux. Pour les textes narratifs, cela ouvre la possibilité de décrire des sphères possibles ou irréelles du récit. Mais encore, pour les textes descriptifs, cela permet de rendre compte des « points de vue » et des évaluations qui en sont inséparables.

b) La dialectique d'un texte peut être spécifiée par le type de processus qu'elle met en œuvre, notamment selon qu'ils sont réversibles ou non. Parmi les processus irréversibles, certains constituent des molécules sémiques nouvelles. L'évolution interne de l'univers sémantique dépend ainsi des structures dialectiques du texte.

Les textes pratiques, du moins ceux qui décrivent des procédures, n'utilisent pas l'ellipse des fonctions, puisqu'ils ont des objectifs didactiques. En revanche, les textes mythiques utilisent ces ellipses, dans la mesure où ils renforcent l'appartenance à une culture en présupposant la connaissance de ses règles. La succession stéréotypée des fonctions

dans les syntagmes fonctionnels sert ici d'interprétant et permet de suppléer les fonctions attendues mais manquantes. Sans en être définitoire, la structure dialectique peut être spécifique d'un genre.

c) La dialogique a été bien étudiée pour les textes littéraires, mais peu pour les autres. Rapportées à la thématique et à la dialectique, les variations dialogiques introduisent des dénivellations entre mondes et des ramifications du temps.

d) Rapportée aux autres composantes, la composante tactique permet de définir des rythmes sémantiques, définis par la mise en rapport d'une unité ou d'une classe d'unités, et d'une série de positions dans la syntagmatique du signifié. On peut distinguer ainsi des rythmes thématiques (utilisés notamment en poésie), des rythmes dialectiques (mis à profit dans les arts du récit), des rythmes dialogiques (dans les romans polyphoniques, et le théâtre notamment). Au delà de leur utilisation esthétique, ces rythmes jouent vraisemblablement un rôle dans la compréhension, en favorisant d'une part la mémorisation, et d'autre part la reconnaissance des formes sémantiques.

3. Besoins et applications

L'essor en cours des recherches textuelles est lié à des progrès théoriques en sémantique, à des enjeux économiques nouveaux, et aux capacités accrues de traitements informatiques. Il est facilité par des corpus étendus, et des dictionnaires informatisés maintenant disponibles.

Mais l'interprétation des résultats des traitements automatiques et le retour aux textes posent des problèmes complexes. Comme un corpus n'est évidemment pas donné mais construit, et dépend du point de vue qui a présidé à sa constitution, il convient de proposer des moyens d'accroître sa représentativité et d'affiner son indexation.

3.1. Besoins et stratégies

a) *Besoins sociaux*

On sait que l'informatique donne à présent accès à d'énormes masses de textes, accessibles sur des supports de moins en moins coûteux. Parallèlement se crée le besoin, entre les applications classiques d'intelligence artificielle et d'automatique documentaire, de créer des instruments de recherche et d'exploration assistées dans les textes.

Les besoins des milieux professionnels vont croissant, notamment dans les domaines de l'extraction d'expertise et de l'analyse du contenu. Par exemple, telle compagnie d'assurance cherche des linguistes pour analyser les structures narratives dans 9.000 déclarations d'accidents mortels ; tel service dépendant du Ministère de la Défense veut analyser 120.000 pages de documents historiques sur les Balkans pour faire de la prospective à partir de séquences événementielles. Dans le domaine du marketing enfin, on cherche à dépasser les méthodes lexicométriques par analyse factorielle, aussi éprouvées que limitées.

b) *Stratégies*

De plus en plus, devant la masse croissante des données disponibles, on recherche des moyens informatiques pour éliminer les données non pertinentes pour l'application. Deux stratégies sont exploitables. La première, issue de la tradition documentaire, médiatise le rapport aux textes par un thesaurus, et/ou une indexation. Elle a fait ses preuves, mais conserve un aspect normatif. En outre, il est difficile de faire évoluer des thesaurus et des indexations. Or, exploiter les bases textuelles exige sans doute de pouvoir modifier les corrélats en fonction des besoins de la recherche, au lieu de proposer toujours les mêmes.

La seconde stratégie se développe avec l'accès informatique au plein texte. Elle permet de caractériser un texte en fonction d'une requête ponctuelle (par analyse de la question) ou permanente (profil de l'utilisateur) ; ou encore de l'indexer contrastivement par rapport aux autres textes du corpus ou d'un sous-corpus de travail.

D'une part l'évolution et la normalisation des formats de saisie permet d'améliorer le balisage des unités de l'expression textuelle (ex. norme SGML) ; la description des standards sémantiques peut s'appuyer sur cette normalisation, et la compléter.

D'autre part, l'informatisation des dictionnaires va permettre à brève échéance des étiquetages sémantiques rudimentaires (indicateurs de domaines, par exemple). L'étiquetage par traits génériques, voire ultérieurement par traits spécifiques, ouvre une direction de recherche prometteuse.

3.2. Le traitement sémantique assisté

Dans ce type d'application, l'interaction avec l'informatique se situe plutôt sur le plan méthodologique que sur le plan théorique : il ne s'agit

pas de construire des maquettes qui fonctionnent sur quelques phrases-test, mais d'utiliser les logiciels existants, en combinant différents outils (analyseurs morphologiques, statistiques, etc.) pour l'aide au traitement sémantique de gros corpus textuels. Par exemple, il importe de mettre au point un système d'aide à l'analyse sémantique qui dépasse les méthodes fondées sur les co-occurrences de mots clé, et qui permette de sélectionner les sous-corpus pertinents en fonction des tâches à accomplir (cf. e.g. l'auteur, 1995).

Pour cela on dispose de deux sortes d'outils : les analyseurs, qui sont mieux dits étiqueteurs (*taggers*) et les navigateurs.

(i) Les étiqueteurs se divisent en deux classes : les analyseurs morphosyntaxiques et les dictionnaires informatisés. Ils permettent d'enrichir le texte d'indexations morphologiques, syntaxiques, sémantiques, qui attachent à des mots ou expressions la mention de leur catégorie dans l'occurrence rencontrée. Ces étiqueteurs marquent pour l'essentiel l'incidence du syntagme et de la phrase sur le mot. Par ailleurs, les indexations statistiques ou certaines indexations sémantiques marquent l'incidence du texte sur le mot.

(ii) Les navigateurs sont des systèmes d'interface qui établissent des liens entre parties d'un texte, textes divers, voire manifestations sémiotiques diverses. Les plus connus sont de type hypertexte. Ils donnent accès à des données, mais ne les produisent ni ne les traitent. Il importe donc de les utiliser en leur donnant un contenu approprié aux applications.

Les systèmes d'analyse sémantique assistée doivent répondre à des besoins trop divers pour pouvoir être développés hors d'une application. Soit par exemple une application de Diffusion ciblée⁵: en fonction du profil de chaque utilisateur, établi par son résumé d'activité, les textes numérisés lui sont acheminés par courrier électronique. L'objectif est d'améliorer l'application en dépassant les calculs de proximité entre documents, et en tenant compte pour cela de la structure des textes diffusés.

⁵ A laquelle nous collaborons avec Xavier Lemesle et Bénédicte Bommier (Direction des Etudes et Recherches d'Electricité de France).

3.3. Genres textuels et traitements automatiques du langage

Comme les Traitements automatiques du langage (TAL) ont affaire à des textes, non à des phrases. Leur typologie est une condition de leur analyse.

Les besoins en amont sont le recueil du corpus et sa préanalyse. D'une part, il faut vérifier l'homogénéité du corpus d'étude, quel que soit le traitement envisagé, et en premier lieu le traitement statistique. Si le corpus n'est pas homogène, il faut poser le problème de la caractérisation interne du corpus par la création de sous-classes. Les critères de cette sous-classification dépendront de l'objectif.

Dans un corpus homogène, on peut poser la question de la structure du genre. En premier lieu, on peut étudier la structure syntagmatique. Certaines parties des textes peuvent en effet être systématiquement éliminées, pour constituer des sous-corpus pertinents.

Pour la caractérisation assistée, on peut envisager trois stratégies :

- L'indexation contrastive de tous les textes d'un corpus sans préanalyse sémantique, au moyen de logiciels (comme le THIEF de Brunet), qui permet de mettre en évidence des pics et dépressions statistiques.
- La caractérisation des séquences (paragraphe, par exemple) pertinentes pour une application, et des critères de leur mise en évidence (position, indices).
- La création de sous-corpus sémantiquement enrichis pour répondre aux objectifs de la tâche en cours.

Pour parvenir à des traitements automatiques spécifiques et efficaces de gros corpus, il convient de spécifier les fonctionnements propres aux différents genres textuels, pour adapter les stratégies d'interrogation à ces genres. Le paramétrage préalable des types textuels permet enfin de simplifier les traitements automatiques, par exemple en éliminant les ambiguïtés. Ainsi, dans les compte rendus d'hospitalisation, *pense* ne correspond qu'à la troisième personne ; il en irait autrement dans un corpus de lettres aux collègues.

3.4. Incidences en retour sur la théorie linguistique

Ce type d'étude permet de refonder ou d'affiner la distinction intuitive et empirique en genres, et de répondre à des questions comme : le roman par lettres est-il bien un roman ? Faut-il classer ensemble romans et nouvelles ?

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. La typologie textuelle peut ainsi apporter du nouveau en linguistique théorique, notamment en morphosyntaxe.

En lexicologie, sur les valeurs positionnelles des unités lexicales, et leur répartition varie également selon les genres : par exemple une étude contrastive de Biber (1993 a) a permis de mettre en évidence une liste de 6.000 mots, pour la plupart concrets, propres aux textes fictionnels (cf. *impatiently* ou *sofa*).

Bref, l'étude des normes linguistiques complète utilement celle des règles, et permettra sans doute de préciser leurs conditions d'application.

Les besoins à satisfaire ne sont pas seulement techniques. Pour définir le cahier des charges d'une station de travail, il faut prévoir les stratégies d'interrogation. Et pour cela, il faut une vue claire des structures sémantiques des textes et de leurs processus de compréhension. Cela exige un approfondissement théorique pluridisciplinaire en linguistique, en psychologie et en informatique.

Aucune des applications n'échappe cependant à ces contradictions : (i) on manipule des mots pour étudier des textes ; (ii) on manipule des chaînes de caractères pour étudier des signifiés.

4. Le problème de l'interprétation et l'apport herméneutique de la sémantique des textes

Alors qu'au cours de ce siècle la linguistique a accéléré son involution positiviste et accru corrélativement son déficit herméneutique, le moment est venu de formuler des propositions qui tiennent compte du caractère herméneutique des sciences du langage.

L'herméneutique matérielle réunit l'herméneutique et la philologie, et place donc la problématique de l'interprétation au cœur des sciences du langage. Cela conduit à privilégier le palier du texte, et à reformuler dans ce cadre le rapport entre global et local. Enfin, à reformuler le lien entre les textes et leur entour social et historique, ce qui suppose une théorie des genres et des pratiques sociales qui les définissent.

4.1. Les ordres herméneutique et référentiel

Cet objectif général nous a incité à proposer une reformulation des ordres de la description linguistique.

Outre l'*ordre syntagmatique* (celui de la linéarisation du langage, dans une étendue spatiale et/ou temporelle), et l'*ordre paradigmatique* (celui de l'association codifiée), les langues mettent en œuvre deux autres ordres, qui à la différence des deux premiers, ne sont pas régis principalement par leur système fonctionnel.

(i) L'*ordre herméneutique* est celui des conditions de production et d'interprétation des textes. Il englobe les phénomènes de communication, mais il faut souligner que les textes ne sont pas simplement des messages qu'il suffirait d'encoder puis de décoder pour en avoir fini avec la langue. Il englobe aussi ce que l'on appelle ordinairement les facteurs pragmatiques, qui affectent la situation de communication *hic et nunc* ; mais il les dépasse car il inclut les situations de communication codifiées, différées, et non nécessairement interpersonnelles. Il est inséparable de la situation historique et culturelle de la production et de l'interprétation.

(ii) L'*ordre référentiel* engage traditionnellement les rapports entre d'une part les signes, les concepts et les choses, d'autre part les phrases, les propositions et les «états de choses». Ce rapport n'a rien d'immédiat : il s'établit par la médiation d'impressions référentielles, sortes d'images mentales que nous avons définies comme des *simulacres multimodaux* (cf. 1991a). Ce que nous appelons ici *référence* n'est pas un rapport de représentation à des choses ou à des états de choses, mais un rapport entre le texte et la part non linguistique de la pratique où il est produit et interprété. Là où l'ordre herméneutique marque l'incidence de la pratique sur le texte, l'ordre référentiel fait retour du texte à la pratique, et traite l'incidence inverse du linguistique (et plus généralement du sémiotique) sur les strates non sémiotiques de la pratique. La référence ainsi définie ne relève pas de la représentation mais de l'action, telle qu'elle est structurée par une pratique. Plus généralement, l'ordre référentiel met en jeu, de façon différenciée, au sein de chaque pratique sociale, les rapports variables entre la sphère sémiotique (ici les suites linguistiques), la sphère des représentations (ici les impressions référentielles), et la sphère physique (ici les « objets »). Il faut donc préciser les modes de référenciation propres aux pratiques so-

ciales. Pour y contribuer, nous avons formulé quelques propositions concernant les textes littéraires (cf. 1989, II).

4.2. Vers une herméneutique critique

Une sémantique des textes, quel que soit son statut au sein de la linguistique, se voit affrontée à deux ordres de problèmes qui définissent son objet et ses objectifs : d'une part, lier la "lettre" du texte, entendue au sens philologique et grammatical, avec son "esprit", c'est-à-dire les diverses interprétations qu'il contraint et suscite. Cela conduit à rompre avec deux attitudes unilatérales, que nous nommerons le "littéralisme" et le "spiritualisme".

Le "littéralisme" aujourd'hui à une linguistique restreinte, dominée par le positivisme et le formalisme, qui se conjoignent dans le positivisme logique et la philosophie du langage anglo-saxonne. Elle peut suivre deux voies. La voie formaliste, asémantique, réduit le signe à son seul signifiant : elle ne considère que la forme des "symboles", et décalque leur sens de leur composition syntaxique. La voie substantialiste rapporte le mot à son sens littéral, censé s'imposer d'évidence, alors qu'il est le résultat d'une construction interprétative et que personne n'a su proposer de méthode pour le produire. Comme les usages effectifs font peu de cas du sens littéral, on a créé la notion de sens dérivé : pour en pallier les insuffisances, il se distingue certes du sens littéral, mais il ne fait que confirmer sa précellence.

Le "spiritualisme" à son tour emprunte deux voies. D'une part le mentalisme classique aujourd'hui revivifié par la sémantique cognitive, qui délie le sens des langues et des textes pour le fonder dans un espace transcendantal (Langacker) ou dans une phénoménologie expérientielle (Johnson).

La seconde voie est celle d'une herméneutique regrettamment coupée de son substrat textuel, et ontologisée par la phénoménologie heideggerienne. Elle a abouti au déconstructionnisme.

Or il existe une place entre une philologie positiviste et une philosophie spéculative : celle d'une linguistique ouverte sur les textes et consciente de son statut herméneutique. Elle entend contester, réduire, voire annuler, la séparation de la lettre et de l'esprit, gagée sur l'allégorisme paulinien, qui prolonge lui-même le dualisme platonicien. Entre la forme matérielle du texte et ses interprétations, tout un système

de normes, dont témoignent les structures textuelles, assure la médiation indispensable. Mieux, l'identification même des moindres signes n'échappe pas aux conditions herméneutiques ; et la reconnaissance des contraintes linguistiques à leur tour peut seule libérer l'herméneutique de son involution spéculative.

Bibliographie

- Adam, J.-M. (1990) : *Eléments de linguistique textuelle*. Bruxelles : Mardaga.
- Barthes, R. (1972) : Article Texte, *Encyclopaedia universalis*, t. 15.
- Beaugrande, R. A. de, et Dressler, W. (1984) : *Introduzione alla linguistica testuale*. Bologne : Il Mulino.
- Bellert, I. (1971) : On a condition of coherence of texts. In : *Semiotica*, III, 4, pp. 253-288.
- Bhatia, V. K. (1993) : *Analysing Genre : Language Use in Professional Settings*. Londres : Longman.
- Berkenkotter, C. & Huckin, T. N. (éd.) (1995) : *Genre Knowledge in Disciplinary Communication*. Hillsdale (N. J.) : Lawrence Erlbaum.
- Biber, D. (1988) : *Variations across Speech and Writing*. Cambridge : CUP.
- (1992) : The multi-dimensional approach to linguistic analysis of genre variation : an overview of methodology and findings. In : *Computers and the Humanities*, 26 (5-6) 331-345.
- (1993 a) : Using Register-Diversified Corpora for General Language Studies. In : *CL* 19 (2)219-241.
- (1993 b) : Co-occurrence Patterns among Collocations : A Tool for Corpus-Based Lexical Knowledge Acquisition". In : *CL* 19 (3) 531-538.
- Brown, G., Yule, G. (1983) : *Discourse Analysis*. Cambridge : Cambridge University Press.
- Coseriu, E. (1981) : *Textlinguistik — Eine Einführung*. Tübingen : Narr.
- Culioli, A. (1984) : Sur quelques contradictions en linguistique. In : *Communications* 20.
- Eco, U. (1979) : *Lector in fabula*. Milan : Bompiani.
- Harris, Z., 1969 : Analyse du discours. In : *Langages*, 61, pp. 8-45.
- Jucquois, G. (1986) : Aspects anthropologiques de quelques notions philologiques. In : François, F. (éd.) *Le texte parlé*. Louvain : CILL, pp. 183-248.
- Kempson, R. (1982) : *La Semantica*. Bologne : Il Mulino.
- Lallot, J., éd. (1989) : *La grammaire de Denys le Thrace*. Paris : Éditions du CNRS.
- Lardet, P. (1992) : *Travail du texte et savoirs des langues*. In : Auroux, S. (éd.) *Histoire des idées linguistiques*. Bruxelles : Mardaga, t. II, pp. 187-205.

- Leach, E. (1980) : *L'unité de l'homme et autres essais*. Paris : Gallimard.
- (à paraître) : Herméneutique matérielle et sémantique des textes, in Salanskis, J.-M. et al. (éds) *Herméneutique : textes, sciences*.
- Rastier, F. (1987) : *Sémantique interprétative*. Paris : PUF.
- (1989) : *Sens et textualité*. Paris : Hachette.
- (1991) : *Sémantique et recherches cognitives*. Paris : PUF.
- (1994) : Sur l'immanence en sémantique. In : *Cahiers de Linguistique Française* 15, pp. 325-335.
- éd. (1995) : *L'analyse thématique des données textuelles — L'exemple des sentiments*. Paris : Didier.
- (à paraître) : Herméneutique matérielle et sémantique des textes. In : Salanskis, J.-M. et al. (éds) *Herméneutique : textes, sciences*.
- Rastier, F. et al. (1994) : *Sémantique pour l'analyse*, avec la collaboration de Marc Cavazza et Anne Abeillé. Paris : Masson.
- Ricœur, P. (1986) : *Du texte à l'action. Essais d'herméneutique II*. Paris : Seuil.
- Rorty, R. (1985) : Texts and Lumps. In : *New Literary History*, XVII, 1.
- Rutten, F. (1980) : Sur les notions de texte et de lecture dans une théorie de la réception. In : *Revue des sciences humaines* 177.
- Slakta, D. (1985) : Grammaire de texte : synonymie et paraphrase. In : Fuchs, C., (éd.) *Aspects de l'ambiguïté et de la paraphrase dans les langues naturelles*, Berne, Peter Lang.
- Stati, S. (1990) : *Le transphrastique*. Paris : PUF.
- Swales, J. M. (1990) : *Genre Analysis. — English in Academic and Research Settings*. Cambridge : Cambridge University Press.
- Van Dijk, T. (1984) : Texte, in Beaumarchais, éd. *Dictionnaire des littératures de langue française*. Paris : Bordas.
- Weinrich, H. (1989 [1982]) : *Grammaire textuelle du français*. Paris : Didier.

