

FRANÇOIS RASTIER

DES CONCEPTS AUX CONTENUS :
POUR UNE SÉMANTIQUE DES TEXTES THÉORIQUES

*J'ai souvent rêvé d'un ouvrage dont le texte,
sur la page de gauche, serait accompagné
d'un commentaire perpétuel sur la page de droite.*

Ferdinand Gonseth, *Mon itinéraire philosophique*, 1995, p. 50.

Les corpus de textes théoriques sont massivement investis par les problématiques du *data-mining*, de l'extraction et de la représentation des connaissances. Toutefois, elles sous-estiment voire ignorent les spécificités de ces textes, considérés comme de simples gisements d'informations.

Après une mise au point sur le statut des textes théoriques, nous aborderons leur typologie pour chercher à caractériser leur textualité propre. Le problème de la *mimésis théorique* permettra d'interroger notamment la structure de l'article scientifique et son hétérogénéité sémiotique.

L'étude des formes textuelles s'attachera d'abord à la formation et à l'évolution des concepts, puis s'étendra aux formes globales de la dialectique. Elle conduira ainsi à une typologie des univers sémantiques propres aux discours scientifique et philosophique.

1. Le statut des textes et discours théoriques

La question du texte scientifique reste difficile à poser : la tradition scientifique occidentale fait de la science une affaire de concepts et de termes et non de textes, car elle tient que l'objectivité est indépendante de la différence des langues et des normes textuelles. Le développement de la linguistique de corpus permet cependant depuis peu d'explorer les caractéristiques propres des textes scientifiques. Elle pourra véritablement les restituer à condition de disposer d'une conceptualisation permettant la comparaison des discours, champs génériques et genres ; et aussi d'hypothèses informées, de façon à pouvoir unifier sous des propositions globales les observations de régularités partielles, comme les définitions, les marques de « polyphonie », etc.

Au-delà du discours scientifique revendiqué comme tel, nous étendrons notre champ d'investigation à l'ensemble des discours théoriques. Leur étude relève de la linguistique comparée, dès lors qu'elle tient compte comme elle le doit de la dimension textuelle : elle s'assigne pour but de caractériser contrastivement les spécificités des discours et genres, en comparant notamment les discours théoriques (sciences *vs* philosophie), les discours scientifiques entre eux, enfin les genres au sein de chaque discipline.

Dans cet exposé, nous privilégierons ainsi la comparaison entre discours scientifique et discours philosophique, le discours littéraire servant au besoin de terme de comparaison. Certes, il peut sembler quelque peu abusif de mettre sur le même plan des discours aux objectifs différents et qui font tout pour se différencier : mais, outre que pour une sémantique des textes tous les discours sont justiciables d'une étude contrastive, c'est là un moyen de mieux caractériser ce qui les distingue et d'éprouver les catégories descriptives. Nous chercherons enfin à contraster les régimes de textualité, tant globalement, par exemple pour ce qui concerne la mimésis, que localement pour le statut des unités textuelles.

Discours théoriques. — En sémantique des textes, l'usage de la notion de *discours* tient aux deux hypothèses que les pratiques sont constituantes et que toute pratique comprend un niveau sémiotique¹. Le niveau sémiotique d'une pratique comporte en règle générale une instance linguistique : par exemple, la religion, entendue comme pratique sociale, comprend un discours qui lui est propre. Il reste bien entendu à chercher des critères linguistiques de caractérisation des discours ; outre les spécificités lexicales, traditionnellement repérées par les indicateurs lexicographiques et décrites par la lexicométrie, on peut explorer, nous le verrons, l'analyse thématique et d'autres composantes textuelles.

Le projet d'une caractérisation linguistique des discours revient à Bronckart et à ses collègues (1985), mais ils n'avaient pas à l'époque les moyens de passer à la caractérisation expérimentale ; depuis, la caractérisation linguistique des discours a pu devenir une question empirique. Comme telle, elle comprend la formulation d'une hypothèse, qui se traduit par l'indication du discours comme variable globale caractérisant chaque texte (ex. juridique vs littéraire)². Une vérification expérimentale par classification automatique sur les textes étiquetés confirme le bien-fondé de l'hypothèse, les divisions entre discours dominant les divisions entre champs génériques et entre genres (cf. l'auteur, 2011, ch. 3).

Dans le monde anglo-saxon, on caractérise les discours théoriques comme des discours académiques. À fondement sociologique, la notion de discours académique renvoie essentiellement aux institutions universitaires. Au plan des fonctions qui distinguent la recherche de l'enseignement, on pourrait estimer que les discours didactiques et les discours scientifiques doivent être traités séparément ; mais en fait les disciplines priment, car elles déterminent leurs champs d'application, y compris didactiques.

Dans le domaine de recherche *English for Special Purposes (ESP)*, les textes scientifiques et techniques sont étudiés dans une perspective normative : leur analyse « rhétorique » entend caractériser des discours et des genres, mais conduit aussi à des applications didactiques qui intéressent les normes rédactionnelles, au demeurant beaucoup plus précises et mieux observées que dans le monde francophone.

En règle générale, l'analyse sémantique des textes théoriques se limite à l'extraction de concepts pour la terminologie et la représentation des connaissances. Les propriétés mises en évidence par la typologie des genres et discours « gênent » cette extraction dans la mesure où ils la relativisent : par exemple, un corpus technique sur la cogénération d'électricité manifeste 40 fois plus de relations hypéronymiques qu'un corpus scientifique sur le même domaine (cf. Grabar et Jeannin, 2002). Une contradiction oppose ainsi les études qui cherchent à caractériser les variations de genres et de discours à la problématique de la représentation des connaissances qui entend construire des ontologies par définition indépendantes de ces variations.

L'analyse contrastive des genres et discours scientifiques reste peu pratiquée ; les textes philosophiques sont encore moins étudiés.

Qu'est-ce qu'une théorie ? — Sans prétendre répondre comme il le faudrait à cette question téméraire, nous retiendrons ici qu'une théorie peut être caractérisée à trois niveaux principaux.

a) Le premier regroupe trois *instances instituant*es. (i) La *préconception du réel*, que l'on a pu appeler sa « métaphysique »³. Guidant la création des hypothèses et, pour une part, leur vérification, elle peut se transposer sans modifications fondamentales dans d'autres champs disciplinaires : par exemple, ce qu'on a nommé le structuralisme a témoigné, dans sa généralité, d'une préconception du réel largement héritée de Saussure. (ii) La *gnoséologie*, ou théorie de la connaissance, règle le rapport entre la préconception du réel, le système conceptuel et la méthodologie qui ont pour fonction de produire, dans les deux sens du terme, le champ d'objectivité. (iii) L'*instance critique* assume une double fonction : de rétrospection pour écrire son

¹ La notion de discours reste cependant affectée par deux méconnaissances. (i) Il s'agirait simplement d'une dimension du langage, comme en témoigne l'opposition *en langue* vs *en discours* traditionnelle en linguistique française, et qui se superpose à la dualité saussurienne entre *langue* et *parole* ; à quoi s'est ajoutée l'idée d'inspiration harrissienne que toute suite voire tout couple de phrases constitue un discours. (ii) Il ne s'agirait pas d'un concept linguistique mais sociologique, voire sociopolitique (comme le postule l'école française d'Analyse du discours). À l'opposé, mais complémentirement, le tenants de l'individualisme méthodologique (qui pratiquent une microsociologie issue du positivisme logique) tiennent que seuls existent des jeux de langage et que genres et discours se résument à des jeux interpersonnels. Cette conception reste bien représentée en analyse conversationnelle.

Dans les deux cas, le discours est reconduit à la sociologie. Le même type d'objection est d'ailleurs opposé à l'existence linguistique des genres : Greimas tenait qu'il s'agit de formations « idéologiques » qui échapperaient comme telles à la description linguistique ; Biber les renvoie également à une sociologie et entend faire émerger par des critères linguistiques des « types de textes » sans rapport défini avec les genres (cf. l'auteur 2011, ch. 3).

² Faute de reconnaître l'autonomie des discours, on serait conduit par exemple à ranger dans la même catégorie le commentaire sportif et le commentaire politique, avec les résultats que l'on devine.

³ Voir Bouquet, 1997.

histoire et désigner ses ascendants et de projection pour formuler un agenda et déterminer une prospective. Elle a un rôle tout à la fois historique et épistémologique.

b) Le second niveau théorique comprend trois *instances instituées* : (i) Le *système conceptuel* articule entre eux des concepts interdéfinis de manière à définir un domaine de légalité sinon de légitimité ; (ii) la *méthodologie* apparie ces concepts à (iii) un *champ d'objectivité*.

c) Enfin, aux discours théoriques sont associées des *instances auxiliaires* qui permettent diverses transpositions. (i) Une instance d'*interdisciplinarité* permet le dialogue avec les disciplines voisines. (ii) Une instance *applicative* permet les applications, qui imposent des traductions simplifiantes vers des disciplines techniques : une théorie doit pouvoir donner des versions partielles et affaiblies d'elle-même, voire partager son champ avec d'autres. (iii) Enfin, une instance *didactique* règle la transmission des savoirs acquis. Soit :

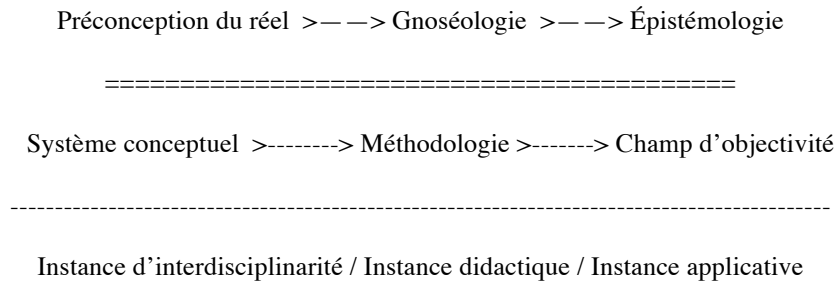


Tableau 1 : Les trois niveaux d'instances théoriques

Si trois sortes de discours théoriques (scientifiques, philosophiques et esthétiques) prétendent comme tels à une validité indépendante d'une situation déterminée, nous ne retenons ici que les deux premiers, en soulignant que dans les discours philosophiques, l'épistémologie, le champ d'objectivité et l'instance applicative demeurent facultatifs. Nous proposerons dans la troisième section de cet exposé des critères pour différencier les concepts des deux derniers niveaux.

Textes théoriques et métalangage. — Le peu d'intérêt manifesté pour une étude linguistique des textes scientifiques et plus largement théoriques tient sans doute à leur statut métalinguistique présumé⁴ ; cependant, ce statut ne les affranchit évidemment pas d'une étude linguistique⁵. À l'opposition entre langage et métalangage, il nous paraît préférable, à la suite de Ferdinand Gonseth, de substituer la distinction entre phase d'*investigation* et phase de *textualisation*. Successives, ces deux phases n'en sont pas moins inséparables⁶. Certains auteurs qui se réclament du positivisme logique, comme Jean-Claude Gardin en archéologie, ont certes pu rêver d'une science sans textes, les langues étant accusées d'imprécision, de polysémie, d'équivoque, etc. Mais l'usage d'autres sémiotiques, comme la photographie ou la vidéo, n'émancipe cependant pas les productions scientifiques d'une forme de textualité : un texte multimédia reste un texte, plus complexe encore.

Une autre distinction de Gonseth caractérise les stratégies scientifiques, quand il oppose la stratégie de *fondement* et la stratégie d'*engagement*. Sans argumenter ici, nous admettons que la textualisation impose nécessairement une stratégie d'engagement, car elle construit de toute manière une représentation de l'énonciateur, tout aussi personnelle quand on emploie le *je*, le *nous*, le *on*, ou que l'on supprime tout pronom de l'énonciation représentée. Elle construit parallèlement une image du lecteur et suppose que l'argumentation est non seulement démonstrative, mais persuasive, comme s'il allait de soi qu'une démonstration doive persuader et une argumentation convaincre.

⁴ La notion de métalangage, élaborée par la logistique russellienne, convient parfois aux langages formels, mais malgré son succès reste sans doute inadéquate pour les langues.

⁵ À la différence des langages formels, les langues sont gouvernées non seulement par des règles, mais par des normes qui configurent la textualité de leurs performances. Les formes de la textualité des discours scientifiques méritent une étude spécifique qui relève de la linguistique comparée et comme telle doit tenir compte de la différence des langues et des usages, y compris nationaux (cf. Fløttum, 2003).

⁶ Loin de résider dans le rapport à un objet donné qui serait observable indépendamment de toute théorie, la « vérité » s'établit dans le rapport entre les investigations et les textes scientifiques. Comme telle, elle suppose une déontologie qui reconnaisse l'Altérité : soit le Réel comme Autre pour les sciences de la nature et l'Autre comme Réel pour les sciences de la culture.

Nous en resterons ici à la textualisation, pour comparer avec une méthodologie commune les textes scientifiques et les textes philosophiques, voire littéraires, et chercher ainsi à spécifier leurs différences.

Genres et composantes sémantiques. — En matière de genres, la linguistique doit élaborer ses propres catégories descriptives, car les disciplines voisines, études littéraires et philosophie, procèdent d'autres problématiques et leur apport reste malgré tout limité.

Comme il n'y a pas d'isomorphisme entre les plans du langage, le problème de la sémosis doit être posé au palier du texte, plutôt qu'en termes de référence au palier du mot ou de vérité au palier de la phrase. La sémosis textuelle est bien une sémosis, c'est-à-dire un acte d'expression et d'interprétation. En tant qu'acte, elle relève d'une praxéologie des performances linguistiques, non d'une ontologie ou d'une théorie des représentations. La poétique, entendue comme théorie des genres, décrit des normes qui relèvent de pratiques sociales dont elles constituent le niveau sémiotique.

On a pu décrire un genre, au plan sémantique, par l'interaction de quatre composantes⁷ non hiérarchisées. Par *composantes*, on entend des ensembles de normes de même type : par exemple, celles de la thématique. Dans les textes théoriques, les thèmes correspondent à des concepts et des notions. De la dialogique relèvent notamment l'éthos, la représentation de l'énonciateur et du lecteur, les évaluations. De la dialectique relèvent les démonstrations et argumentations, voire en certains cas les structures narratives ; de la tactique, l'étude des positions des unités sémantiques.

Le genre contraint l'interaction des composantes sémantiques mais aussi leur articulation aux composantes de l'expression. Il définit en effet un rapport normé entre le plan de l'expression et le plan du contenu au palier textuel : par exemple, dans le genre de l'article scientifique, le premier paragraphe, sur le plan de l'expression, correspond ordinairement une introduction, sur le plan du contenu.

Bref, les genres permettent voire déterminent l'établissement entre les deux plans du langage des relations complexes constitutives de la textualité. La production du texte établit en outre un compromis entre les contraintes de la pratique, dans laquelle se définit le genre, et la situation — c'est-à-dire, au-delà du *hic et nunc*, la position historico-culturelle de l'énonciateur et de l'interprète.

La *sémosis textuelle* normée par le genre détermine 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 l'attestent les textes gnomiques ou religieux.

Comme tout usage modifie et configure potentiellement la langue, chaque texte modifie son genre ; mais les perturbations locales ne sont sensibles qu'au sein d'une stabilité globale et en retour le genre configure le texte.

En tant qu'unité minimale d'analyse, le texte agit comme instance globale par rapport à ses éléments, mais aussi comme instance locale par rapport à son corpus. Par le biais notamment de son genre, le texte pointe sur son corpus et sélectionne en quelque sorte les autres textes qui permettent de l'interpréter. Le premier cercle du corpus, corpus nécessaire mais pas toujours suffisant, est ainsi constitué des textes du même genre. Outre le mode de régulation du contexte interne (par la textualité), le genre détermine ainsi le mode de recours au contexte externe, celui de l'intertextualité — qui apparaît dans l'article scientifique par les références, et moins évidemment par les allusions voire les larcins.

Un exemple : les genres de la philosophie. — Comme on croit que le discours de la vérité n'a pas de genres, rien de surprenant si la philosophie n'a guère étudié les siens⁸. Si l'on a plutôt étudié le style des philosophes que les genres dont ils usent, c'est pourtant dans ce domaine que les réflexions sur la textualité philosophique paraissent le plus prometteuses. On pourrait songer à lier des genres et des traditions philosophiques : le dialogue serait privilégié par la tradition idéaliste, comme chez Platon, Leibniz, Berkeley ; ou à des postures : le traité serait un genre favori des dogmatiques, ou favoriserait le dogmatisme. Mais cette voie ne peut être suivie qu'avec circonspection.

Pour caractériser les voies linguistiques de la philosophie classique, on pourrait toutefois distinguer les *genres argumentatifs*, qui recherchent la vérité par la quête des apories et l'articulation des arguments et les *genres révélationnels* qui interprètent une vérité déjà exprimée ou la laissent brièvement entrevoir.

Le *dialogue* et le *traité* relèvent de la première voie. Ils partagent des structures dialectiques comparables et les mêmes *topoi* de démonstration, mais leurs structures dialogiques diffèrent évidemment, car le dialogue,

⁷ Cf. l'auteur, 1989 et 2011, ch. 3.

⁸ Cf. l'auteur, 2001a.

par la pluralité des énonciateurs représentés, peut suggérer qu'il existe plus d'un point de vue légitime sur l'Être, en quoi il sera préféré par les subjectivistes ; alors que le traité, qui n'a qu'un narrateur au point de vue uniforme, suit en général une voie thétique et objectiviste.

Le genre ancien du *commentaire* et le genre moderne du *fragment* illustrent de manière antithétique la voie révélationnelle. Le *commentaire* et son succédané la *paraphrase* supposent qu'une révélation humaine ou divine a eu lieu : la connaissance réside dans des textes anciens, mais il faut encore la discerner et l'exposer aux contemporains. Par le *fragment*, en revanche, des philosophes contemporains, notamment à la suite de Nietzsche, ont cru rivaliser avec les présocratiques – dont on ne connaît plus que des textes fragmentaires. Le fragment ne situe pas nécessairement la révélation dans un passé, mais il veut la faire jaillir dans le présent instantané ou éternel de sa lecture. Par sa structure, il interdit en tout cas une mimésis de la complétude, ce qui en fait un genre antidogmatique par excellence. Dans le domaine de la philosophie, l'investigation ne semble pas précéder la textualisation et les genres constituent alors les gnoséologies, argumentative ou révélationnelle par exemple, qu'ils sont réputés simplement illustrer.

L'histoire des genres scientifiques reste presque aussi peu étudiée que celle des genres philosophiques. Ainsi, l'article scientifique vient de la lettre (ex. la lettre de Pascal au Père Noël), qui circulait à l'âge classique dans les milieux savants, avant la création des revues universitaires. Dans certaines disciplines, il a laissé sa place au *paper*, plus court, ponctuel et moins réflexif. Chaque domaine élabore et hiérarchise ses genres. Par exemple, en France, au Centre National de la Recherche Scientifique, dans le département des Sciences de la vie, les livres ne sont pas retenus dans les bilans d'activité des chercheurs, car ils sont considérés comme relevant de la vulgarisation. Là encore, la hiérarchie des genres reflète une préconception normative de la connaissance.

2. La mimésis des textes théoriques

Ces points de typologie engagent à mieux caractériser la textualité propre des écrits théoriques. Traditionnellement employée pour les textes littéraires depuis la *Poétique* d'Aristote, la notion de *mimésis* peut cependant être étendue à l'ensemble des textes, car l'on peut considérer que chaque genre détermine un mode génétique, un mode herméneutique et un mode mimétique (pour un développement, cf. l'auteur, 2001 a). Dans les termes de la tradition logico-grammaticale, le mode mimétique correspondrait à la fonction référentielle.

Censés représenter une réalité, les textes théoriques exercent une fonction mimétique qui nous importe indépendamment de leur valeur scientifique ou philosophique. Une sémantique des textes doit décrire les conditions des impressions référentielles (ou « effets de réel ») suscitées par des structures linguistiques : elles configurent en une « ontologie » locale la sémantique des domaines d'investigation que l'on appelle les objets scientifiques.

La séparation conventionnelle entre sujet et objet impose aux textes scientifiques une mimésis de l'objectivité, d'où par exemple les stratégies d'effacement ou d'euphémisation de l'énonciateur, que renforce la multiplication des citations et références : elles ne jouent pas seulement un rôle doxographique, mais s'appuient sur la doxa scientifique pour prévenir l'auteur de toute accusation de subjectivité « haïssable ».

Les niveaux d'analyse des textes scientifiques. — On dit ordinairement qu'un texte appartient à un genre. Toutefois, comme un genre n'est pas une essence, ni même un type, mais une sommation évolutive de régularités, cette proposition mériterait d'être renversée : le genre appartient au texte, qui contient des indications de son genre (dans son titre, dans son support, mais aussi dans son lexique, dans son mode compositionnel, etc.). On peut distinguer trois niveaux principaux d'analyse du texte : le *péritexte* (titres, titres courants, liens, etc.) ; puis l'*intra-texte* (colonnes graphiques, « casiers ») ; enfin, l'*infratexte* subordonné (notes, etc.). Ces niveaux valent pour l'écrit mais trouvent des analogues à l'oral : le péritexte est alors épilinguistique (cf. *je vais t'en raconter un bien bonne*), de même que l'infratexte (expansions sur un point de détail).

Dans les textes scientifiques, tels qu'on peut les coder conformément aux recommandations de la *Text Encoding Initiative (TEI)*, les variables globales, sociologiques (comme l'école), éditoriales (ouvrage, revue), linguistiques (discours, champ générique, genre) peuvent figurer dans l'en-tête (*header*). Par ailleurs, le corps du texte (*body*) est décomposé en sections balisées (sur la base de critères d'expression) comme le titre, les résumés, les exemples, les notes, la bibliographie. Enfin, il faut détailler les hors-texte.

Il convient de questionner le rôle mimétique de la diversité des configurations ou « séquences » textuelles propres au texte scientifique (résumés, notes, bibliographie, hors-textes divers) : les différentes sections des genres scientifiques, l'article en premier lieu, décrivent ou construisent par des voies mimétiques diverses des aspects d'un « même » objet. Ainsi, l'objectivation scientifique passe non seulement par la concordance sociale des points de vue des spécialistes, mais aussi par la concordance des mimésis propres aux différentes sections du texte.

Les étapes du déroulement du texte peuvent même correspondre à des phases de l'objectivation. Au plan de la structure d'ensemble, les différents niveaux de généralité (titre, résumé, introduction, position théorique, argumentation, expérimentation) construisent dans leur succession une structure hiérarchique qui rappelle fort la structure arborescente d'une ontologie.

La multiplicité des sémiotiques. — Les textes scientifiques se caractérisent par la présence de hors-texte, quasiment absents des discours philosophique et littéraire. Rompant la linéarité linguistique, ils introduisent d'autres formes de mimésis : ainsi, le caractère frontal et non modalisé des illustrations leur permet de passer pour la « chose même ». En outre, la séparation entre texte et hors-textes redouble la séparation entre sujet et objet. En effet, la multiplicité des sémiotiques induit un effet d'objectivation, car c'est un apanage de l'Être même que de pouvoir être dit de plusieurs façons⁹.

Si tout discours a sa propre « raison graphique » (Goody), le rôle des figurations, croquis et diagrammes, peut être rapporté au fait que les différents moments génétiques de la textualisation scientifique (imaginer, concevoir, argumenter) sont associés à des sémiotiques différentes. La variété sémiotique du texte scientifique reflète sans doute, dans sa textualisation même, la variété des approches adoptées dans la phase d'investigation. Par exemple, en linguistique de corpus, les tableaux lexicométriques, les analyses factorielles, etc., témoignent d'opérations méthodologiques concordantes pour objectiver des propriétés du corpus étudié.

Enfin, pour le lecteur qui interprète le texte scientifique, la diversité des sémiotiques à l'œuvre assure aussi une fonction d'objectivation, comme, toutes proportions gardées, un document multimédia suscite un effet de réel empirique plus intense qu'un médium pauvre comme le croquis.

Faute d'une sémiotique adaptée, nous manquons cependant d'une théorie suffisamment élaborée des effets de réel théorique ; elle serait d'autant plus nécessaire que les textes scientifiques multimédia se multiplient et se complexifient.

Cercle herméneutique et relations intersémiotiques. — Certains auteurs, comme René Thom, estiment que la modélisation graphique résout certains problèmes épistémologiques des sciences sociales : « By using the “distanciation” effect of geometric representation, to break the hermeneutic circle which has kept imprisoned so many social science thinkers » (1980). Cette mise à distance permettrait, confirme Petitot, de « briser le cercle herméneutique » (1985, p. 83) : une représentation serait alors conforme aux choses mêmes, parce qu'elle en exprime la forme, structurant selon les mêmes principes les deux « couches de l'Être », physique et sémiotique. Mais on peut voir dans ce vœu une illusion diagrammatique, car les représentations géométriques elles-mêmes doivent être interprétées, et, loin de le briser, elles renforcent et complexifient le cercle herméneutique.

Comme il n'y a pas de hiérarchie *a priori* des langages, les langues « naturelles », les langages artificiels, les sémiotiques graphiques se distinguent par des capacités mimétiques différentes. Aucun langage n'est inadéquat, mais chacun a ses régimes herméneutiques : ceux des langues étant fort ouverts, elles semblent des interprétants absolus, puisqu'elles permettent d'instituer les autres langages et de les interpréter, d'exprimer voire de fixer le sens des autres sémiotiques et leurs conventions d'interprétation. Si le cercle herméneutique est maintenu, la récursion qui empile indéfiniment les métalangages qui se définissent est évitée : le texte scientifique peut être objectivé sans récursion métalinguistique, puisque la « métalangue » fait pleinement partie du discours. Outre que le discours scientifique n'est pas techniquement un métalangage, il n'existe pas de métalangage qui prendrait le discours scientifique comme langage objet et qui serait autre que la langue.

Dans ces conditions, le problème de la référence des textes scientifiques devient un problème intersémiotique de constitution des impressions référentielles. Comme nul n'a accès à la chose en soi, le

⁹ Cf. Aristote, *Métaphysique*, Gamma, 1006.

problème du rapport des discours théoriques à leur empirie se transpose ainsi en problème des relations intersémiotiques au sein de leurs textes et entre leurs textes. Si chaque système de signes crée son objet propre par son mode mimétique particulier, les textes polysémiotiques comme le texte scientifique permettent d'unifier ces objets partiels par la synthèse qu'opère leur textualité même. Les textes scientifiques sont en effet des performances polysémiotiques complexes et leur textualité détermine, notamment par les normes de genre, le mode d'interaction des sémiotiques en leur sein.

Pour une typologie des mimésis. — Comme chaque genre, champ générique voire discours crée cet effet de réel particulier qu'on appelle son objet, nous avons proposé pour en rendre compte la notion de *régime mimétique*. Un tel régime dépend d'une part des structures textuelles et notamment dialectiques (narratives ou argumentatives), mais aussi de l'homogénéité ou de l'hétérogénéité sémiotiques. Elles induisent des mimésis du monde empirique (« concret ») ou transcendant (« abstrait »). On aurait par exemple cette répartition :

Discours	<i>Romanesque</i>	<i>Philosophique</i>	<i>Scientifique</i>
<i>Dialectique</i>	Narrative	Argumentative	Argumentative
<i>Sémiotique</i>	Homogène	Homogène	Syncrétique
<i>Mimésis</i>	Empirique	Transcendante	Mixte

Tableau 2 : Typologie des mimésis

Ainsi, les textes philosophiques semblent-ils en quelque sorte intermédiaires entre les textes scientifiques et les textes littéraires. La mimésis des textes littéraires n'est pas unifiée et l'on a pu opposer par exemple la mimésis empirique du roman réaliste à la mimésis transcendante de la poésie métaphysique (cf. l'auteur, 1992). À cet égard, la mimésis des textes scientifiques semble syncrétique, car ils figurent tout à la fois des relations entre des concepts et entre les objets qu'ils configurent. Fût-ce sur le mode de l'hypothèse, le texte scientifique part d'un postulat réaliste de vérité qui le distingue idéalement de tout régime fictionnel. Il induit cependant une mimésis mixte, car son abstraction renvoie au transcendant, alors même qu'il vise l'empirique dans l'objet qu'il construit. Comme le texte poétique, mais par une voie tout autre, il unit ainsi le sensible et l'intelligible pour créer un effet de vérité.

3. Formes textuelles et thématique

Peut-on transposer la théorie et la méthodologie utilisées pour les thèmes littéraires aux thèmes philosophiques ? Les premiers résultats semblent positifs ; mais peut-on définir de même des thèmes scientifiques, et si c'est le cas, dans quels genres (leçons transcrites, traités, articles) ?

Des termes et concepts. — Pour la théorie terminologique, les termes ne sont pas des signifiés mais des concepts ; elle confère donc à certains mots le statut de termes et en fait alors l'expression d'unités conceptuelles non liées à la langue. Nous l'avons vu, la doctrine terminologique de Wüster considérait ainsi les concepts comme des représentations indépendantes des langues et le cognitivisme perpétue aujourd'hui cette tradition. À présent, les projets WordNet et EuroWordNet ont précisément pour but de représenter les lexiques des langues par une ontologie qui en reste indépendante — bien que ses concepts soient figurés par des mots anglais en capitales (pour une discussion, cf. l'auteur, 2004b).

Les logiciels d'extraction terminologique utilisent pour leur part des indices phraséologiques, mais cette approche sous-estime inévitablement la textualité propre des écrits théoriques. Comment les termes et les non termes interagissent-ils ? Par ailleurs, à quelles conditions un mot devient-il un terme ou cesse-t-il de l'être ? La construction d'ontologies évite ces questions par une conception achronique et non textuelle des corpus théoriques.

Les concepts sont ordinairement abordés par leurs définitions et par les relations sémantiques à valeur définitionnelle repérables dans les textes (hypéronymie, hyponymie, méronymie, synonymie). La première indique leur genre, la deuxième et la troisième leurs parties, la dernière leurs lexicalisations. Ces relations sont en fait une projection de l'article de dictionnaire : on commence par le genre commun (ex. *action de...* ou *objet qui...*), puis on passe à la différence spécifique qui mentionne une ou plusieurs parties caractérisantes ; enfin l'on renvoie aux expressions jugées équivalentes. On utilise pour constituer ainsi des « ontologies » des patrons « lexico-syntaxiques », comme : +X ET ADVERBE DE SPECIFICATION Y (hypéronymie), qui permet d'identifier par exemple un syntagme comme *Intégration efficace des sources d'énergie renouvelables, en particulier l'énergie solaire* (cf. Grabar et Jeannin, 2002). Cependant, cette méthode ne voit évidemment que ce qu'elle retient : elle sélectionne *a priori* certains types de contextes favorisant l'expression des relations hiérarchiques retenues. Par ailleurs, elle ne tient aucunement compte de la situation du concept dans le texte, puisqu'elle entend l'intégrer à une ontologie du domaine. Bref, elle permet de le repérer, mais non de le définir par rapport à ses contextes effectifs. Elle est ainsi l'expression aboutie d'une problématique logico-grammaticale qui utilise les patrons formels comme des instructions « contextuelles » indépendantes du contexte. En somme, les patrons sont normatifs et locaux, alors que nous avons besoin d'une démarche tout à la fois descriptive et globale, qui tienne effectivement compte des régularités et des spécificités du texte et de son corpus. L'extraction de « concepts » par patrons terminologiques gagnerait à être complétée par une *approche globale* qui tienne compte du texte et de ses structures. Dès que l'on dispose de corpus suffisants, l'approche statistique peut être utilisée pour la caractérisation, par des techniques comme la classification automatique (pour les textes) ou le test de l'écart réduit (pour les corrélats de « concepts »).

Pour la sémantique textuelle, le concept est une molécule sémique (groupement structuré et relativement stable de traits sémantiques) qui correspond dans le cas le plus simple, à un sémème à lexicalisation unique et sans traits afférents. Mais il s'agit là d'un cas privilégié, car une molécule sémique peut avoir une manifestation diffuse par des corrélats récurrents. Par ailleurs, tous les éléments du texte concourent à la construction des unités textuelles : un « concept » se définit alors aussi bien par des lexèmes cooccurrents à ses lexicalisations privilégiées, que par sa distribution, par des ponctuations associées, voire des italiques, etc.

Faisant retour au mentalisme qui précédait la formation de la linguistique, les linguistes cognitivistes appellent *concepts* les signifiés. En renonçant au préjugé mentaliste, on peut par concession appeler *concepts* certaines unités sémantiques des textes théoriques : ce léger abus de langage suppose qu'on les considère comme des unités linguistiques, non des représentations mentales indépendantes des langues.

Le rapatriement des concepts dans les textes a plusieurs conséquences. (i) En premier lieu, nous prenons nos distances avec la conception terminologique : les concepts ne sont pas les signifiés de certains mots, même si certains peuvent jouir d'une lexicalisation privilégiée. (ii) De même façon, toutes proportions gardées, que les personnages d'un récit ne se réduisent pas à leur nom, les concepts sont des unités textuelles, et en tant que telles hautement contextualisées : relevant de la méso-sémantique (palier intermédiaire entre la micro-sémantique du mot et la macro-sémantique du texte), elles peuvent être décrites comme des thèmes ou des acteurs de la dialectique ou de la dialogique. (iii) Comme telles, elles sont associées à des traits descriptifs relevant de tous les niveaux de l'analyse linguistique. Ainsi, comme toutes les formes sémantiques, un concept peut être associé à des éléments de formes expressives : position, catégories morphosyntaxiques, ponctuations, etc.

Des concepts aux thèmes. — Sans s'arrêter à des préoccupations ontologiques, on peut définir, au sein d'un corpus qui relève d'un discours et d'un genre déterminé, des formes sémantiques, en premier lieu des thèmes¹⁰. Qu'il possède ou non une lexicalisation unique, un thème peut être décrit en deux étapes. (i) On recrute une clique de ses *cooccurrents*, les chaînes de caractères (lexèmes, grammèmes, ponctèmes) qui dans ses contextes définis par une fenêtre de 20 mots dépassent un écart réduit seuillé à 3 (en corpus multigenre) voire à 2 (en corpus unigenre). (ii) On qualifie ensuite certains cooccurrents de l'expression comme des *corrélats* sémantiques, dès lors que leur contenu partage au moins un sème commun avec celui du mot qui

¹⁰ Les acquis de la thématique assistée sur corpus numériques permettent en effet d'aborder la description des « concepts », requalifiés comme des signifiés, par un nouveau biais mis à l'épreuve sur des corpus littéraires dans le recueil intitulé *L'analyse sémantique des données textuelles* (1995a).

supporte l'interrogation initiale. La structure constituée par ces sèmes est la *molécule sémique* qui définit le thème (cf. l'auteur, 2001 b, ch. 7 et 2011, ch. 2, § 4).

Le modèle idéalisé du signe qui suppose une correspondance entre signifiant et signifié est ici encore moins valide qu'ailleurs : d'une part, les formes sémantiques comme les formes expressives peuvent être manifestées de manière compacte ou diffuse ; d'autre part les relations qui assurent la sémiosis des formes sémiotiques résultent d'un parcours interprétatif complexe, non séquentiel, qui comporte : (i) le passage des cooccurrents aux corrélats ; (ii) la structuration des corrélats en molécules sémiques ; (iii) la recherche onomasiologique des formes expressives ; (iv) la qualification des cooccurrents ponctuels pour interpréter leurs faisceaux au sein de configurations de l'expression.

Des études récentes indiquent que les concepts fonctionnent comme des thèmes dans les discours théoriques, qu'ils soient scientifiques ou philosophiques (cf. Valette, 2003, Loiseau, 2003, Poudat, 2006). Les thèmes théoriques gardent cependant leurs spécificités par rapport aux thèmes littéraires : par exemple, là où la littérature varie les dénominations et descriptions, les théories tendent à les figer.

L'étude des corrélats sémantiques des lexicalisations privilégiées permet de préciser le statut des thèmes théoriques, en distinguant les unités qui participent à l'évolution des *formes* sémantiques le long du texte de celles qui appartiennent au *fond* sémantique général¹¹. Les premières sont caractérisantes, car elles correspondent aux « concepts » effectivement élaborés, débattus et utilisés, alors que les secondes restent d'une grande généralité. Par exemple, Rinck, dans une étude contrastive avec des études de critique littéraire (2005), relève 21 mots spécifiques de la linguistique, qui tous cependant pourraient parfaitement se retrouver dans n'importe quelle science humaine : *analyse, analyser, approche, catégorisation, concept, conceptuel, discussion, donnée, étude, étudier, exemple, hypothèse, interprétation, notion, objectif, observation, observer, problématique, problème, recherche, système*¹².

Outre la fréquence, d'autres critères peuvent être retenus : les unités de formes sont épisodiques, celle de fond sont réparties assez uniformément. Les premières ont en outre un nombre plus élevé de cooccurrents et par suite de corrélats. Soit, en somme :

	<i>Éléments de forme</i>	<i>Éléments de fond</i>
<i>Fréquence</i>	Inférieure	Supérieure
<i>Distribution</i>	Épisodique	Uniforme
<i>Nombre de cooccurrents</i>	Supérieur	Inférieur
<i>Nombre de corrélats</i>	Supérieur	Inférieur

Tableau 3 : Éléments sémantiques de forme et de fond

Ces distinctions permettent d'éclairer les difficultés rencontrées par les logiciels d'extraction terminologique, car tous les candidats termes ne sont pas à égalité, et les plus en évidence restent les moins caractérisants, car ils appartiennent pour ainsi dire au « fond de sauce » de la discipline.

Un exemple commenté permettra de discuter quelques critères de caractérisation des thèmes théoriques. Les concepts sont ordinairement conçus comme des noms d'entités et leurs flexions grammaticales ne sont pas considérées comme pertinentes. Cependant, les travaux en linguistique de corpus permettent de différencier les acceptions en fonction de la catégorie du nombre. Ainsi, une étude de Poudat (2006), conduite sur 224 articles de linguistique étiquetés par 141 catégories avec le logiciel TnT¹³, permet de juger que *sens* au singulier et *sens* au pluriel, *langue* et *langues* n'ont pas les même cooccurrents, tant en qualité qu'en quantité, si bien que les singuliers et les pluriels n'ont pas le même statut textuel.

¹¹ Parallèlement à des recherches sur la perception sémantique (cf. 1991, ch. 7), nous avons proposé une théorie des formes sémantiques (l'auteur, 1989, 1994, ch. 6, 2001b, ch. 1) : en bref, les thèmes sont des formes qui se profilent sur des fonds dont les isotopies génériques sont un exemple.

¹² *Système* peut revêtir toutefois une acception spécifique.

¹³ Ils sont extraits de 32 numéros de 11 revues de linguistique, publiés entre 1995 et 2003 et occupent un volume de 9 Mo en format *txt*. Le nombre d'occurrences s'élève à 2082 pour *langue* et 973 pour *langues*, 2186 pour *sens* (sg.) et 179 pour *sens* (pl.). Le logiciel a été entraîné pour l'occasion sur le discours scientifique, de manière à ajuster le jeu d'étiquettes.

	<i>Sens (sg.)</i>	<i>Sens (pl.)</i>	<i>Langue</i>	<i>Langues</i>
Lexèmes	13	3	7	4
Grammèmes	2	2	6	1
Ponctèmes	1 (nég.)	1	1 /1 (nég.)	0
Autres	2 (nég.)	0	3 (nég.)	0

Tableau 4 : Types de cooccurrents dépassant un seuil d'écart réduit de 2.

N. B. : Les chiffres indiquent le nombre de variables qui dépassent un seuil d'écart réduit positif ou négatif fixé à 2.

Détaillons ces résultats synthétiques. Ils conduisent à penser qu'à la différence de leurs pluriels *sens* (sg.) et *langue* sont de bons candidats pour figurer parmi les lexicalisations privilégiées de concepts. Ce n'est pas en raison du nombre comparativement plus important de leurs cooccurrents lexématiques, car, les lois ordinaires d'isotopie favorisant la répétition et la paraphrase, la plupart appartiennent à la même famille morphologique (comme *sémantique* ou *langagier*, respectivement). Une minorité de cooccurrents lexématiques seulement peuvent être qualifiés comme corrélats sémantiques : pour *sens* (sg.), *contexte* (0,47) et *contextuel(le)* (0,23), *interprétation* (0,41) et *interprétatif* (0,26), *texte* (0,32), *langages* (0,31) ; pour *langue*, on ne trouve que *parole* (0, 27).

Quant aux autres cooccurrents, les grammèmes ne distinguent que *langue* : ensemble des déterminants de 3^e personne du pluriel (0,31), ensemble des prépositions (0,3), ensemble des déterminants (0,23), imparfaits (0,2), connecteurs d'addition (0,21).

La ponctuation attire d'autant plus l'attention que les débats séculaires sur la nature des concepts n'en ont évidemment jamais tenu compte. Ici, les singuliers *sens* et *langue* se distinguent de *sens* (pl.) et *langues* par le déficit de certaines ponctuations (parenthèses, -0,23 et -0,24 respectivement) et d'autres symboles (numéraux (-0,28 ; -0,25), symboles, sigles, abréviations : -0,23 ; -0,24). En revanche et complémentairement, le tiret d'incise caractérise *sens* (pl.) avec un excès de 0,37 – ce qui reste considérable dans un corpus aussi homogène. Comment se fait-il que les singuliers « repoussent » ces symboles et que les pluriels les « attirent » ? Il est vraisemblable que le déficit des ponctèmes au singulier, et leur excès (relatif) au pluriel soient liés à la différence des instances : le *sens* et la *langue* seraient des concepts de l'instance instituée ; les *sens* et les *langues*, des notions des instances auxiliaires, notamment l'instance applicative. En effet, dans les articles ou les sections d'articles à vocation descriptive, les ponctèmes comme les parenthèses ou les tirets d'incise sont plus nombreux, de même que les symboles comme les astérisques, etc. ; pour introduire et discuter des exemples, on utilise en effet d'autres configurations textuelles que pour discuter des concepts.

Indice concordant, la longueur moyenne des textes dans lesquels figurent *sens* (sg.) est corrélée à la longueur des phrases (+0,27), contrairement à *sens* (pl.). Or, au sein des genres scientifiques, les textes à dominance théorique (qui relèvent des instances instituante et instituée, cf. *supra* tableau 13) sont en moyenne plus longs que les textes à dominance applicative (qui relèvent des instances auxiliaires). Ce paradoxe s'explique sans doute parce que les textes à dominance théorique restent attachés au modèle du discours philosophique, qui se caractérise par la longueur moyenne plus étendue de ses mots, de ses phrases et ses textes.

Complémentairement, à la différence de *langue* (sg.), *langues* (pl.) est positivement corrélé aux noms concrets étiquetés comme tels par le logiciel Cordial (+0,18), et donc négativement corrélé aux noms abstraits. Ainsi, bien qu'à l'instar des dictionnaires les terminologies lemmatisent leurs entrées, singuliers et pluriels méritent au contraire d'être nettement distingués. Même si l'abstraction théorique tend à privilégier les singuliers, ne nous hâtons pas de conclure qu'en général les mots au singulier auraient vocation à lexicaliser des concepts à la différence des pluriels. Un travail spécifique reste à faire pour éprouver à ce propos les critères que nous utilisons comme indices d'abstraction : longueur moyenne des mots, des phrases et des textes, association avec des mots abstraits, détermination, etc.

En conclusion provisoire de cette analyse, retenons que dans ce corpus *sens* (sg.) lexicalise un concept (comportant une molécule sémique), alors que 'langue', notion résiduelle ne figure plus parmi les objets de

débat et ne se définit que par son opposition à celle de *parole*¹⁴. Enfin, *sens* (pl.) et *langues* ne manifestent que des éléments du fond sémantique, à l'instar des lexèmes fréquents et peu spécifiques, comme *propriétés*, *exemples*, *relations*, qui dans ce corpus sont du même ordre de fréquence absolue.

Sans prétendre la généraliser, éprouvons enfin cette conclusion par un exemple tiré d'un corpus relevant d'un autre discours. D'après les données recueillies par Sylvain Loiseau, dans *L'Anti-Œdipe* de Deleuze et Guattari, les 20 premiers cooccurrents d'*inceste*, comme *refoulement*, *mère*, *frère*, *prohibition*, *filiation*, etc., appartiennent tous au domaine de la psychanalyse, comme le confirment les contextes, et leurs taux de corrélations réciproques sont très élevés. En revanche, les 20 premiers cooccurrents d'*image* sont répartis entre plusieurs domaines : économique (*capitalisme*), psychanalytique (*inceste*, *refoulé*), mais aussi artistique (*théâtre*, *Chaplin*), etc. À la différence d'*inceste*, qui semble une lexicalisation privilégiée d'un concept, *image* participe donc sans doute à l'expression de divers concepts encore à identifier et décrire.

Émergence, définition et évolution des concepts. — En somme, relativement à une forme sémantique, une lexie peut : (i) la lexicaliser ; (ii) en lexicaliser une partie (on parle alors d'expression de contour) ; (iii) lexicaliser une partie du fond sur lequel se détache cette forme ; (iv) demeurer indépendante tant de cette forme que de son fond.

Comment alors les thèmes théoriques se lexicalisent-ils ? Leur lexicalisation s'accompagne-t-elle d'un appauvrissement de leur richesse contextuelle ? Ces questions engagent à une étude diachronique et comparative sur corpus numériques dans diverses disciplines ; au sein d'œuvres singulières comme celles de Guillaume (cf. Valette, 2003) et Deleuze (cf. Loiseau, 2003), elles engagent à retracer en diachronie l'apparition, l'évolution voire la disparition de concepts.

Si les concepts ne sont pas le reflet d'une ontologie préalable, leur émergence peut se décrire comme un processus d'évolution linguistique et les inégalités de statut que connaissent les lexies d'un texte théorique se laissent schématiser ainsi : (i) certaines concourent à l'établissement de fonds sémantiques (isotopies) ; (ii) d'autres participent à la construction de formes sémantiques, au premier rang desquelles les molécules sémiques, qu'elles lexicalisent partiellement (dans le cas où il existe une lexicalisation privilégiée, ce sont alors ses corrélats) ; (iii) enfin, d'autres encore lexicalisent de façon synthétique des formes textuelles (molécules sémiques). L'émergence d'un concept consiste à parcourir successivement ces trois stades.

Arrivé au troisième, il est encore susceptible d'être fixé par une *définition*¹⁵ qui sépare le terme de ses contextes effectifs pour lui associer un contexte canonique et éviter ainsi son évolution. Elle n'intéresse toutefois que les lexicalisations synthétiques et ne rend pas compte de la formation ni de la dissolution du concept.

Après ces quatre phases, la « déconceptualisation » ou dissolution met en jeu d'autres parcours. Soit le terme devient inusité et disparaît du domaine ; soit il cesse de lexicaliser une molécule sémique déterminée et s'étend à un nombre indéterminé d'acceptions : par exemple, le terme *système* en informatique a cessé de lexicaliser un concept déterminé. Ce processus s'achève quand le terme sort du domaine sémantique considéré : ainsi, la *matière première* des traités de métaphysique est devenue un terme général qui peut s'employer dans divers domaines dès lors qu'il s'agit de production.

¹⁴ Il peut paraître étrange que le mot *langue* ne soit pas considéré comme un concept mais seulement comme une notion : de fait, ce terme fréquemment utilisé n'est pas souvent problématisé, puisqu'il fait à l'évidence partie des acquis fondateurs de la discipline. Il tend donc à passer dans le fond sémantique, comme en témoigne sa répartition.

¹⁵ La définition, rappelons-le, n'est pas un moyen de connaissance, mais une simple configuration récurrente dans les discours théoriques, techniques et didactiques.

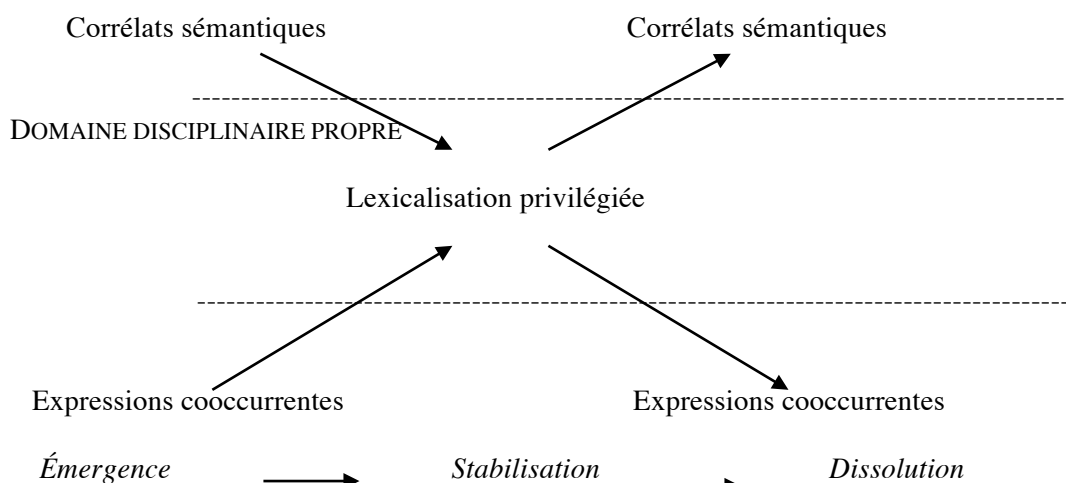


Figure 1 : L'évolution du concept

N.B. : Les flèches convergentes symbolisent une diminution du nombre des contextes, les divergentes une augmentation de ce nombre, qui conduit à une désémantisation banalisante.

À l'évolution sémantique des corrélats répond souvent une évolution morphologique qui accompagne la formation du terme. D'abord empruntée à un autre domaine ou à la langue supposée « générale », l'expression qui le lexicalise correspond à un simple homonyme du terme d'origine dès lors qu'elle revêt une signification propre à la discipline (ainsi des termes *étoile* ou *tonneau* en mathématiques). Ou bien encore, dans le cas de la formation d'un néologisme, elle combine des morphèmes courants dans des domaines scientifiques ou techniques : ainsi, Gustave Guillaume forme le terme *psychomécanique*¹⁶.

La banalisation s'accompagne ensuite d'une diminution de la densité sémique du concept, qui perd certains traits sémantiques : le concept ainsi désémantisé devient de ce fait compatible avec un grand nombre de contextes voire de domaines disciplinaires. En somme, dans la phase de formation du concept, on passe de corrélats non spécifiques et non disciplinaires à des corrélats spécifiques et disciplinaires. Puis dans la période de dissolution, on assiste soit à une disparition, soit à une banalisation qui revient à des corrélats sémantiques et des expressions associées non spécifiques à la discipline.

4. La dialectique, des personnages du mythe aux acteurs théoriques

Si les textes philosophiques semblent organisés par une narrativité abstraite, comment la dialectique des textes scientifiques se manifeste-t-elle ? Des mythes platoniciens aux expériences de pensée de la philosophie analytique, le discours philosophique abrite sans doute des formes incluses et résiduelles du mythe¹⁷. Si les personnages paraissent s'absenter du discours philosophique, ils ne tardent pas à y réapparaître sous la forme non anthropomorphe des concepts de la quête gnoséologique. Ainsi, en passant de l'épopée archaïque au dialogue philosophique, les personnages sont-ils subrepticement passés du récit à la narration, mais la dialectique n'a pas pour autant perdu tout contact avec le mythe. Même si l'on considère chaque personnage d'un dialogue philosophique comme le représentant d'une position abstraite du pensable, en analysant la progression du dialogue on s'aperçoit que des actes dialogiques comme *prendre un exemple*, *manier l'ironie*, *se faire l'avocat du diable*, *discréditer un adversaire* ne sont pas moins codifiés que les phases du combat singulier.

Si le passage du récit à la narration permet à la philosophie de se réfléchir en elle-même et d'y gagner sa réflexivité caractéristique, il reste à interroger le statut narratif des concepts. La dialectique dispose et met en scène des personnages comme la Vérité, l'Esprit, la Nécessité, l'Opinion droite, etc., mais leurs interactions peuvent-elles se regrouper en séquences stéréotypées comme des fonctions narratives ? Quels sont leurs

¹⁶ Cf. Valette, 2003, sur qui je m'appuie ici.

¹⁷ Ainsi l'apologue de la Terre Jumelle, chez Putnam, emprunte à la mythologie moderne de la science-fiction ; celui de la Chambre chinoise chez Searle rappelle l'énigme d'un roman policier.

modes de récursion et d'enchâssement ? Les réponses à ces questions ouvertes varient sans doute selon les époques et les écoles¹⁸.

En outre, un récit philosophique peut fort bien ménager des séquences scientifiques et techniques. Une analyse détaillée des *Éléments d'Idéologie* de Destutt de Tracy (l'auteur, 1971) a pu établir que les développements scientifiques de la grammaire et de la logique sont surdéterminés par un récit de quête dont les trois épreuves successives se résumeraient grossièrement ainsi : l'Esprit humain, malgré les artifices et les sophistications qui ont fini par voiler l'origine, doit créer un langage conforme à la nature et donc parfait ; au moyen du langage, il constitue la connaissance en deux étapes, la représentation puis la mise en ordre des idées ; il instaure le règne des Lumières en communiquant la connaissance au moyen du langage et spécialement de l'écriture. On retrouve dans ce récit abstrait les trois épreuves canoniques que l'on trouve dans les récits épiques : qualifiante, principale et glorifiante¹⁹.

Existe-t-il de même une narrativité scientifique ? C'est le cas, certainement dans les sciences sociales. Par exemple, on peut lire dans les travaux néo-darwiniens sur l'origine du langage de petits récits anthropologiques qui expliquent les cultures en termes d'avantages adaptatifs : le langage servirait à faire la paix dans de chameilleuses hordes ancestrales, ou encore à médire des absents (*gossip hypothesis*), etc. Pichot emploie à leur propos l'expression de *romans anthropologiques* : c'est beaucoup d'honneur, mais il reste que dans les sciences sociales les hypothèses scientifiques revêtent souvent des formes narratives élémentaires, qui qualifient des acteurs et introduisent, avec une téléologie narrative, une sorte d'intelligibilité rudimentaire. La fin proclamée des « grands récits » n'aura donc aucunement empêché la multiplication des petits, même dans les lieux où l'on professe un scientisme éliminateur. Par exemple, la résorption des sciences sociales dans les sciences de la nature semble un tardif avatar, en un récit épistémologique de *naturalisation*, du retour à la Nature que prêchaient les moralistes antiques et les philosophes des Lumières.

5. La dialogique, des narrateurs du mythe aux « voix » du texte scientifique

Si les concepts héritent des personnages des récits, les agents de l'interlocution représentée héritent de leurs narrateurs. C'est l'évidence en philosophie : le dialogue, genre éponyme de la dialectique philosophique, met en scène des personnages qui, pour porter parfois le nom de personnes historiques, n'en sont pas moins littéraires. Ils se sont peu à peu absentés dans l'abstraction : soit ils ont revêtu des figures didactiques en résumant des positions traditionnelles (comme Hylas et Philonoo chez Berkeley) ; soit, dans le genre du traité, depuis les sommes médiévales, ils se sont réduits à de simples arguments, introduits par des expressions comme *sed contra*, qui signalent simplement que l'auteur n'adhère pas à la position qu'il décrit.

Mais qu'en est-il des positions représentées dans le texte scientifique ? Cette question mérite d'autant plus l'attention que leurs normes discursives, pour l'objectivation nécessaire à l'impression de vérité, bannissent précisément les marques de subjectivité. La confrontation des positions se représente alors soit sur un mode anthropomorphe par la mention de l'auteur et de ses collègues et/ou adversaires, soit par la mise en scène abstraite de paradigmes ou de théories.

Au-delà, des études sur la polyphonie du texte scientifique ont permis des observations intéressantes sur les modes de présentation de l'énonciateur. Pour éviter l'illusion ordinaire qui confondrait l'énonciateur représenté et l'auteur, il importe toutefois de distinguer l'auteur du narrateur et donc de préciser les spécificités de la narration théorique. Pour décrire la dialectique des textes théoriques, on peut s'appuyer sur la tradition de l'analyse structurale des récits, mais on ne dispose pas d'élaboration analogue pour la dialogique, car la narration proprement dite n'a pas été étudiée avec le même souci d'approfondissement : alors qu'il existe des théories des acteurs et fonctions du récit, rien de tel pour la narration²⁰. On peut

¹⁸ Par exemple, chez Hegel, qui théorisa l'épopée, ne peut-on reconnaître encore le caractère épique des aventures de l'Esprit, dont les trois étapes rappellent curieusement la triplication des épreuves dans l'épopée indo-européenne ?

¹⁹ Plus précisément, les *Éléments d'Idéologie* articulent trois récits enchâssés, gnoséologique, grammatical et idéographique, qui narrent la conquête de la connaissance, la création et l'élaboration des langues, enfin la création d'une idéographie. Le premier appartient à une isotopie mythique, le second à une isotopie scientifique, le troisième à une isotopie technique.

²⁰ L'histoire des formes littéraires modernes évolue vers une subjectivisation croissante, absorbant le récit dans la narration, les épreuves du héros dans les tribulations du narrateur. On pourrait retracer ce processus, sur le mode plaisant, du *Roland Furieux* au *Don Quichotte*, de Tristram Shandy à Jacques le Fataliste et jusqu'à Joyce et Faulkner. Il reste à étudier comment la topique du récit se voit redoublée par celle de la narration (adresse au lecteur, début par une confidence, etc.).

cependant spécifier des fonctions et des acteurs de la narration représentée, en distinguant deux paires d'acteurs.

a) La première assume des fonctions de communication. Ses acteurs sont le *Guide* et le *Régisseur*²¹. Le *Guide* assume une tâche didactique d'accompagnement du lecteur et utilise des figures de participation ; par exemple, il emploie le *nous* inclusif (« Nous avons vu plus haut... »). Ses actions figurent un parcours en commun du texte, d'où la fréquence de verbes de mouvement et de perception (*venir, voir, etc.*)²². Le *Régisseur*, pour sa part, trace les contours du propos : il utilise le *nous* exclusif de majesté académique ou le *je*, pour en préciser et surtout limiter les ambitions (on trouvera fréquemment des verbes comme *se limiter, se contenter, etc.*), et corrélativement des modalisations euphémiques, comme le conditionnel.

Alors que le *Guide* tient un propos thétique, caractérisé par des aspects accomplis, des temps perfectifs, le *Régisseur*, malgré sa position éminente, se cantonne souvent dans l'hypothétique, caractérisé par des aspects inaccomplis et des temps imperfectifs. Sur l'axe de la communication représentée, le *Régisseur* traite de la production, mais en anticipant l'interprétation par des figures d'*accommodatio*, tandis que le *Guide* se cantonne à la lecture plus qu'à l'interprétation.

Le *Guide* et le *Régisseur* sont dans le discours scientifique deux concrétisations par des acteurs de l'instance du Point de vue (cf. 2011, ch. 2).

b) Sur l'axe de la représentation, ou plus précisément celui de représentation dialogique de la mimésis, deux autres acteurs se partagent les tâches, le *Garant* (irénique) et le *Critique* (polémique). Ils concrétisent complémentairement dans le discours scientifique l'instance de la Garantie propre à tout objet culturel.

Le *Garant* s'engage en son nom ou au nom de l'institution qu'il représente, mais sur un mode objectiviste ; il introduit les définitions ; il adopte une position thétique qui lui fait à l'occasion privilégier le *nous* de majesté académique, le *on*, et surtout les formulations déclaratives sans pronoms de première personne ; enfin, les temps et modes thétiques, les aspects perfectifs.

Complément du garant, le *Critique* argumente contre des thèses ; soit celles qui viennent d'être présentées par le *Garant* (« on pourrait objecter que »), mais alors il joue le rôle d'adjuvant en prévenant par anticipation les critiques du lecteur ou des adversaires académiques ; soit surtout celles de ces adversaires, parfois exécutés sommairement en note et déjà diminués de n'être mentionnés que là, comme chez Bourdieu par exemple. Dans le premier cas, les modes sont hypothétiques, dans le second thétique – malgré un bon nombre d'interrogation rhétoriques. Les évaluations sont orientées péjorativement. En somme, nous aurions la disposition suivante :

Acteurs	<i>Guide</i>	<i>Régisseur</i>	<i>Garant</i>	<i>Critique</i>
<i>Pronoms typiques</i>	Nous (incl.)	Nous (excl.)	On	Je
<i>Temps</i>	Futur, PC	Conditionnel	Présent, PC	Présent
<i>Aspects</i>	Perfectifs	Imperfectifs	Perfectifs	Perfectifs
<i>Modalités</i>	Thétique	Hypothétique	Thétique	Thétique
<i>Évaluations</i>	Mélioratif	Péjoratif	Neutre	Péjoratif

Tableau 5 : Quelques manifestations morphosyntaxiques des acteurs de l'énonciation représentée dans les textes scientifiques.

N.B. : Les critères utilisés sont simplement didactiques. D'autres variables auraient pu être choisies, car les prétendues « marques de l'énonciation » ne sont pas attachées à des catégories morphosyntaxiques particulières. Par exemple, l'usage des noms propres peut avoir un rôle critériel, puisque le *Guide* ne les utilise pas.

²¹ Comme les acteurs du récit, nous signalons les acteurs de la narration par des capitales à l'initiale.

²² En multipliant les figures de participation, le style *reader-friendly* aujourd'hui recommandé retrouve certains aspects de l'*accommodatio* rhétorique et confère un ton didactique vaguement convivial à un nombre croissant de textes scientifiques.

Cette première typologie pourrait être liée à des fonctions dialogiques dont on peut trouver des exemples parmi les figures non tropes comme la délibération ou l'argumentation, la concession ou... le dialogisme²³ qui prend dans le texte scientifique la forme de citations commentées. Malgré son caractère exploratoire, elle permet de premières conclusions. (i) Fondée sur la notion de métalangage, la notion de métadiscursivité reste ambiguë, car elle convient tant au Régisseur qu'au Garant. (ii) La distinction entre acteurs se confirme si l'on considère leur distribution selon les parties de l'article : le *nous* dans les notes est exclusif et renvoie à la position académique magistrale, alors que dans le texte il est souvent inclusif et renvoie par une figure de participation. La distinction entre acteurs permet alors de distribuer les acceptions selon les parties du texte : ainsi les notes sont-elles un lieu d'exercice du Régisseur (et non du Guide), ou du Critique et non du Garant. Kjersti Fløttum (2005) a clairement montré les variations pronominales dans l'emploi de *je*, *vous* et *on*, selon qu'il s'agit du rôle du Guide ou du Garant²⁴. Les stratégies diffèrent cependant selon les auteurs : entre deux articles scientifiques de même discipline, sur les trois pronoms *je*, *nous*, *on*, elle relève les scores respectifs de 12, 4, 14 vs 0, 27, 15 – le *on* étant la forme non marquée.

Les acteurs de la dialogique scientifique participent à la construction de l'éthos. Par exemple, les euphémisations et précautions modalisantes du Régisseur dérivent historiquement du *topos* de modestie jadis prescrit aux orateurs et qui participe de la *captatio*. En outre, dans le texte scientifique, il a été remotivé par la mimésis d'objectivation qui péjore les marques de la subjectivité et notamment le *je*. Par exemple, dans les articles de linguistique, les *je* sont trois fois plus fréquents dans les exemples que dans le corps du texte (25, 5 vs 8 %) alors que les *nous* sont trois fois plus fréquents dans le texte que dans les exemples (5 vs 17% ; cf. Poudat, 2006).

Bien entendu, la construction de l'éthos est un phénomène global qui implique tant les acteurs dialogiques que les acteurs dialectiques voire les thèmes. Enfin, les acteurs de l'énonciation représentée ne se distribuent pas indifféremment selon les instances théoriques (cf. *supra* tableau 13) : (i) les instances instituant ne disposent que du Garant et éventuellement du Critique pour l'épistémologie ; le plus souvent, elles ne sont pas associées à l'énonciation représentée, ce qui signale une forme d'implicite. (ii) Les instances instituées mettent à contribution le Garant et le Régisseur. (iii) Les instances auxiliaires multiplient les figures du Guide et du Régisseur.

6. Directions de recherche

De la forme locale à la textualité globale. — De fait, les descriptions linguistiques de textes scientifiques restent souvent très partielles, limitées à une gamme d'expressions, de termes, de « connecteurs argumentatifs », de pronoms, etc.

La compréhension des rapports entre global et local est une des difficultés majeures que rencontre la linguistique des textes – et d'ailleurs l'ensemble des sciences de la culture. Comment donc lier les caractéristiques locales aux structures textuelles ? Dans la composante thématique, par exemple, il importe de relier les expressions locales, comme les termes, dont le contenu relève de la microsémantique, aux concepts, qui sont des molécules sémiques et à ce titre relèvent de la mésosémantique. Ensuite, il faut articuler la description des concepts aux sections (ou « séquences ») du texte scientifique. Enfin, reconstituer l'économie globale du texte, en tenant compte de la fonction du genre, mais aussi du projet propre de l'auteur. Cela engage un patient travail comparatif entre auteurs et disciplines qui suppose des programmes coordonnés (cf. Fløttum et Rastier, éd. 2003).

Les structures des univers sémantiques. — Un discours ne correspond pas seulement à des ensembles de genres, il est associé à un domaine sémantique spécifique qu'il configure et qui assure son unité thématique. Les structures de ces domaines et corrélativement les types de mimésis globale que ces discours induisent sont susceptibles de variations qui restent à caractériser. On peut utiliser la notion d'*univers sémantique* pour esquisser une typologie structurelle de ces domaines, et, en prolongeant la typologie des mimésis esquissée ci-dessus (fig. 1), proposer cette distinction :

²³ Cf. Fontanier, *Les figures du discours*, Paris, Flammarion, 1997, p. 375-377. Elle pourrait être étendue aux textes philosophiques et littéraires. Par exemple, le début de Jacques le Fataliste se caractérise par l'absence du Garant et de Guide, le Critique exerçant ses talents à l'encontre du lecteur.

²⁴ Elle emploie une autre terminologie en caractérisant « different manifestations of the self as writer, researcher and arguer ».

— L'univers sémantique des textes littéraires est stratifié en isotopies génériques entre lesquelles sont établies des relations métaphoriques, et, plus généralement, analogiques. Les structures dialectiques sont valuées et les acteurs des diverses isotopies sont susceptibles de se regrouper en agonistes²⁵.

— L'univers sémantique des textes philosophiques (du moins dans la tradition occidentale) se stratifie également en isotopies génériques, mais moins nombreuses et moins variées. Les métaphores et analogies ne sont pas systématiques. Enfin, seuls certains concepts hiérarchiquement supérieurs revêtent le statut d'agonistes (la Nature chez Spinoza, l'Esprit chez Hegel) et se subdivisent en acteurs²⁶.

— Enfin, les univers sémantiques des sciences ne comportent en principe qu'un domaine, ce pourquoi l'on estime qu'elles ont un objet. Les métaphores et analogies sont alors ponctuelles et purement didactiques. Enfin, les acteurs ne se regroupent pas en classes d'équivalence, en raison de la distinctivité principielle des concepts.

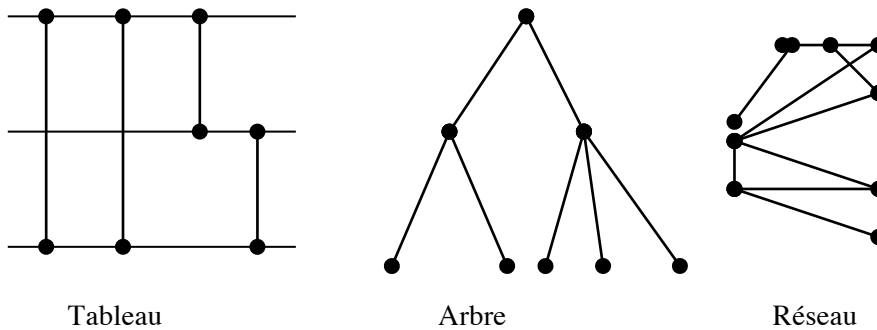


Tableau 6 : Typologie des formes globales des univers sémantiques

La structure tabulaire des univers littéraires dérive sans doute de ce que Lévi-Strauss appelait la « structure feuilletée » du mythe. La structure arborescente des univers philosophiques correspond, dans l'onto-théologie occidentale, à la forme définitoire de l'ontologie, fondée sur les relations de subsomption qui conduisent aux « genres suprêmes de l'Être », comme les catégories d'Aristote. Enfin, les discours scientifiques, obtenus par restriction puis autonomisation à partir du discours philosophique, structurent des domaines locaux de manière à interconnecter tous leurs concepts en réseau.

<i>Discours</i>	Littéraire	Philosophique	Scientifique
<i>Dialectique</i>	agonistes	acteurs et agonistes	acteurs
<i>Thématique</i>			
Isotopie(s) générique(s)	multiples	limitées	unique
Structure	tabulaire	arborescente	en réseau

Tableau 7 : Quelques critères de caractérisation des univers sémantiques selon les discours

Comme plus haut pour la mimésis, le discours philosophique occupe ici encore une position intermédiaire entre le discours littéraire et le discours scientifique. Cela s'explique sans doute par des raisons historiques : la philosophie occidentale s'est formée au sixième siècle av. J.-C. à partir du discours mythique dont la

²⁵ 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 ; mais ils restent cependant susceptibles de relations métaphoriques.

²⁶ Ainsi Loiseau (2003) peut-il distinguer sous le concept de Nature chez Spinoza trois thèmes qui en fait correspondent à trois acteurs, *Nature-Dieu* sur l'isotopie ontologique, *Nature-État* sur l'isotopie politique, et *Nature-Loi* sur l'isotopie juridique.

littérature est issue, mais en réaction contre lui. À leur tour, les discours scientifiques ont pris peu à peu leur autonomie à l'égard du discours philosophique dont ils sont issus et qu'ils ne cessent de récuser — bien qu'il reste rémanent, comme en témoignent par exemple les brouillons scientifiques.

Une dimension polémique demeure toutefois, chaque discours cherchant à se démarquer de celui dont il est issu : ainsi, des scientifiques affectent-ils de garder leurs distances avec la philosophie en affichant aujourd'hui une nette indifférence à l'égard de l'épistémologie philosophique et *a fortiori* de la littérature, en refusant par exemple les métaphores et comparaisons²⁷, voire en adoptant un style des plus inélégants, qui ne doit rien au principe de plaisir, et donc tout, espèrent-ils, au principe de réalité.

Cependant, la pensée théorique peut conserver une unité que masque son élaboration disciplinaire. Construisant le texte à partir de notes diverses, le travail génétique du savant consiste notamment à unifier par les normes d'un genre des fragments préexistants : toutefois, les brouillons scientifiques restent souvent, chez Saussure et chez Guillaume par exemple, dans une indistinction de genre comparable à celle des brouillons littéraires, voire peut-être identique, si l'on admet que la méditation artistique et la méditation scientifique connaissent un régime commun, celui d'une pensée qui n'est aucunement désincarnée et autarcique comme le prétend la théorie instrumentale du langage : elle doit son unité contradictoire au fait qu'elle se tient certes dans la langue, mais encore en-deçà d'un genre strictement défini, et reste dans la pensée sans devenir encore de la science. Cela concorde avec cette observation de Valéry : « Science et art sont presque indiscernables dans la période de l'observation et de la méditation pour se séparer dans l'expression – se rapprocher dans l'ordonnance – se diviser définitivement dans les résultats » (*Cahiers*, II, 1974, p. 926). Allant plus loin encore, Cassirer suggérait une origine mythique de la science : « le point de départ authentique de tout devenir de la science, l'élément immédiat qui est à son origine, se trouve moins dans la sphère sensible que dans l'intuition mythique. [...]. On ne peut comprendre intégralement le développement de la science – considéré en un sens idéal et non temporel – qu'à la condition de montrer comment elle procède de la sphère de l'immédiateté mythique et s'élabore à partir d'elle [...]. » (Cassirer, 1925, p. 11). C'est pourquoi peut-être les brouillons scientifiques sont si fascinants, car une théorie peut se périmier, alors que la dynamique d'une pensée reste inachevée et donc inépuisable.

Bibliographie

- Biber, D., (2000) « Investigating language use through corpus-based analyses of association patterns », in M. Barlow & S. Kemmer (éds) : *Usage-based models of language*, Stanford, CSLI Publications.
- Bouquet, S. (1997) *Introduction à la lecture de Saussure*, Paris, Payot.
- Bronckart, J.-P., et coll. (1985) *Le fonctionnement des discours. Un modèle psychologique et une méthode d'analyse*, Paris, Delachaux & Niestlé.
- Cassirer, E. (1973-1975 [1925]) *Philosophie des formes symboliques*, Paris, Minit, 3 vol.
- Fløttum, K. (2005) Individual linguistic variation and possibilities for academic selfhood, in *Akademisk prosa*, Université de Bergen, Department of Romance Studies, Skrifter fra KIAP, 3.
- Fløttum, K. et Rastier, F. (éds) (2003) *Academic Discourse— Multidisciplinary Approaches*, Oslo, Novus.
- Forest, D. et Meunier, J.-G. (2000) La classification mathématique des textes : un outil d'assistance à la lecture et à l'analyse de textes philosophiques, in Rahman, M. & Chappelier, J.-C. (éd.). *Actes des 5es Journées internationales d'Analyse statistique des Données Textuelles*, 9-11 mars 2000, EPFL, Lausanne, Volume 1, pp. 325-329.
- Grabar, N. et Jeannin, B. (2002) Contribution de différents outils à la construction d'une terminologie pour la recherche d'information, *Actes Ingénierie de la Connaissance (IC)*, Rouen, 28-30 mai 2002, n.p.
- Loiseau, S. (2003) Philosophical discourse from autonomy to engagement: Deleuze commentator of Spinoza, in Fløttum et Rastier (éds), pp. 36-54.
- Loiseau, S. et Rastier, F. (2010) Linguistique de corpus philosophique — l'exemple de Deleuze, in Patrice Maniglier, éd., *Le tournant philosophique des années soixante*, Paris, PUF.

²⁷ « J'évite autant que possible, parlant des choses de la langue, qui ont une rigoureuse spécificité, de faire appel à des comparaisons » (Guillaume, cité par Valette, 2003). Dans leur livre sur les impostures scientifiques, Sokal et Bricmont érigent encore la métaphore en indice certain de l'imposture, comme si les métaphores computationnelles, neuronales et néo-darwiniennes qu'ils prennent au pied de la lettre pouvaient, en réifiant l'imaginaire scientifique, passer pour des thèses scientifiques.

- Petitot, J. (1985) *Morphogenèse du sens*, Paris, PUF.
- Poudat, C. (2006) *Etude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres*. Thèse de doctorat, Université d'Orléans, <http://www.texto-revue.net>.
- Rastier, F. (2001a) L'action et le sens. — Pour une sémiotique des cultures, *Journal des Anthropologues*, 85-86, pp. 183-219.
- Rastier, F. (2001b) *Arts et sciences du texte*, Paris, PUF.
- Rastier, F. (2004a) Doxa et lexique en corpus - Pour une sémantique des « idéologies », *Actes des Journées scientifiques en linguistique 2002-2003 du CIRLLEP*, Reims, Presses Universitaires de Reims.
- Rastier, F. (2004b) Ontologie(s), *Revue de l'Intelligence Artificielle*, Numéro spécial Informatique et terminologies, 18, pp. 16-39.
- Rastier, F. (2011) *La mesure et le grain – Sémantique de corpus*, Paris, Champion.
- Rastier, F., (1971) *Idéologie et théorie des signes*, La Haye, Mouton.
- Rastier, F., (1989) *Sens et textualité*, Paris, Hachette. Rééd. [http : //www.revue-texto.net](http://www.revue-texto.net)
- Rastier, F., éd. (1995b) *L'analyse thématique des données textuelles – L'exemple des sentiments*, Paris, Didier.
- Thom, R. (1980) *The Role of Mathematics in Present-Day Science*, Hanovre, Union internationale de philosophie des sciences.
- Valéry, P. (1974) *Cahiers*, Paris, Gallimard, t. II.
- Valette, M. (2003) Conceptualisation and Evolution of Concepts. The example of French Linguist Gustave Guillaume, in Kjersti Floettum et François Rastier, éd., *Academic Discourse— Multidisciplinary Approaches*, Oslo, Novus.