

ONTOLOGIE(S)

François RASTIER
C.N.R.S.

(Article paru dans la *Revue des sciences et technologies de l'information*,
série : *Revue d'Intelligence artificielle*, 2004, vol. 18, n°1, p. 15-40)

SOMMAIRE :

1. Introduction
2. De l'ontologie aux ontologies
 - 2.1. Ontologie et lexiques
 - 2.2. L'Être et les êtres
 - 2.3. Ontologies et hiérarchie
 - 2.4. Liens verticaux et horizontaux
3. De la hiérarchie ontologique aux relations sémantiques
 - 3.1. Les trois types de relations lexicales
 - 3.1.1. Les relations verticales et l'héritage
 - 3.1.2. Les relations horizontales, synonymie et antonymie
 - 3.1.3. L'articulation des relations verticales et horizontales
 - 3.2. Questions sur les relations sémantiques privilégiées
 - 3.2.1. Hyperonymie
 - 3.2.2. Improbables héritages
 - 3.2.3. Polysémie et synonymie
4. Éléments d'une dé-ontologie
 - 4.1. Philosophie du langage ou linguistique ?
 - 4.1.1. Lexique, langue et discours
 - 4.1.2. Le problème du signifiant
 - 4.1.3. Le problème des inégalités entre signes
 - 4.1.4. Le problème des contradictions
 - 4.1.5. La destruction du contexte
 - 4.1.6. L'oubli des textes
 - 4.2. À quoi servent les ontologies ?

RÉSUMÉ . Les réseaux sémantiques que l'on construit à présent sont nommés ontologies. Au-delà de la logique, ce choix repose sur un lien réaffirmé avec la tradition métaphysique. Cependant, les lexiques des langues ne sont pas structurés comme des ontologies. Les relations sémantiques sont en effet plus complexes et variables que ce que prévoient les constructeurs d'ontologies. Pour construire une " dé-ontologie ", il faut restituer la diversité des discours et des genres, qui rendent illusoire une ontologie unique ; insister sur le problème de la diversité sémiotique des textes, les corrélations complexes entre contenu et expression, l'incidence constituante du contexte. C'est là une condition pour remplir des tâches de caractérisation, notamment en linguistique de corpus.

ABSTRACT. The semantic networks that are built today are called ontologies. This choice rests not only on the tradition of Logic, but also on Metaphysics. Unfortunately, the vocabularies of natural languages are not shaped like ontologies : semantic relations are in fact much more complex and varied than ontologic relations. To build a " de-ontology ", we have to take into account the diversity of discourses and genres, because a unique ontology remains illusory ; we have to insist on the problem of semiotic heterogeneity of texts, on complex correlations between meanings and expressions, on the role of context. These are some requirements to fulfil characterization tasks, especially in corpus linguistics.

Ce n'est pas sans un certain saisissement que l'on voit alors la notion même de l'être tomber de son piédestal, dans la foule des notions imparfaites et sommaires.

Ferdinand Gonseth (1990, p. 61)

1. Introduction □

Le correcteur orthographique de Word n'admet pas le pluriel pour *ontologie*. Cependant, le florissant domaine des ontologies est en passe de devenir un champ disciplinaire autonome qui intéresse les ressources multilingues, le « web sémantique », etc. Nous questionnerons ici son statut et ses perspectives, du point de vue de la linguistique et de l'histoire des idées.

Techniquement, les ontologies sont des réseaux sémantiques comme on en connaissait voici vingt ou trente ans [1]. La nouveauté réside dans leur échelle sans précédent (par dizaine de milliers de « concepts ») et dans leur utilisation pour servir de base de connaissances interlangues.

La transformation des réseaux sémantiques en ontologies n'est pas simplement cosmétique : elle ancre ces formalismes de représentation sémantique par graphes dans la tradition philosophique du cognitivisme orthodoxe, qu'illustre notamment le psychologue George Miller, concepteur de WordNet et inspirateur d'EuroWordNet [2], les principales ontologies que l'on construit à présent.

WordNet est un dictionnaire électronique de l'anglo-américain, développé depuis 1985 et initialement conçu pour tester les déficits lexicaux dans des expériences de psychologie cognitive. Sa structure est celle d'un thésaurus. Il a été transposé à une dizaine de langues, du basque au bulgare. En outre, il sert d'interlangue (*ILI* ou *Inter Lingual Index*), et donc de représentation conceptuelle indépendante des langues, dans le projet EuroWordNet, développé depuis 1996. Chacune des langues décrites (italien, néerlandais, anglais, espagnol, et bientôt sans doute l'allemand, le français, l'estonien, le tchèque, etc.) développe son propre lexique à l'image de WordNet, en développant

- page 17 -

une ontologie « générale » commune [3]. Participant d'un mouvement général d'ontologisation (et de mondialisation), ces projets revêtent un intérêt exemplaire, tant par l'influence de leurs choix théoriques [4] que par les crédits affectés à une échelle sans précédent dans l'histoire des traitements automatiques du langage. S'ils s'imposent à l'attention, de même qu'à présent les ontologies construites pour le web sémantique, c'est plus généralement les conditions de construction des ontologies qui nous retiendront ici.

2. De l'ontologie aux ontologies □

Le choix du mot *ontologie* n'est pas sans conséquence. L'ontologie se définit en effet comme la « science de l'Être » : elle reste constitutivement métaphysique, car la métaphysique est la « science de l'Être en tant qu'Être » [5].

En revanche, le rôle des sciences reste précisément de rompre avec la métaphysique en définissant et en structurant de façon critique et réflexive des domaines d'objectivité. Ainsi, c'est à la linguistique d'étudier les lexiques des langues. L'ontologie, en tant que science des entités, ne peut prétendre décrire les langues, puisque les « concepts » qui sont censés représenter ces entités sont de fait universels. Mais le poids du préjugé « de bon sens » reste tel qu'on voudrait fonder la sémantique des langues sur la description préalable d'ontologies, d'ailleurs construites sans analyse linguistique et sans la collaboration de linguistes. Les conséquences de ce choix sont considérables.

Créant la conception occidentale de l'Être dans son poème *De la Nature*, Parménide (*circa* 544, *circa* 450 av. J.-C.) le décrit comme inengendré et sans trépas, immobile, un, au présent, continu, et

dépourvu de fin (fragment VIII, v. 3-6). Les prédicats majeurs de l'action sont ici niés : le transit temporel (« jamais il n'était ni ne sera », v. 5), le mouvement, la variation qualitative, et la finalité (*télos* est plutôt un but qu'une finitude). Sous diverses guises, on retrouvera donc jusqu'à nos jours ce type d'oppositions homologuées :

- page 18 -

l'être	le faire
la substance	les accidents
l'unicité	la multiplicité
l'intemporalité	la temporalité
l'invariance	la variation

Ce dispositif a de multiples conséquences. En privilégiant l'être par rapport au faire, on primera les noms, toujours plus nombreux dans les ontologies, et l'on déplorera que les verbes se laissent moins hiérarchiser. On négligera la variation temporelle : ainsi les ontologistes ne précisent jamais à quelle condition on devient une entité, ni à quelle condition on cesse de l'être – en effet, une entité est par nature intemporelle. On rapportera les variations à des accidents, inessentiels par définition, etc.

Plus généralement, on négligera la variation des situations, en rapport évidemment avec l'engagement historique et culturel de toute action. Par un paradoxe insistant, les ontologies sont en effet supposées indépendantes de toute pratique déterminée et l'on en conclut à leur utilité : elles seraient ainsi indéfiniment réutilisables ; mais pour la même raison, on peut tout aussi bien arguer qu'elles sont inadaptables aux applications.

2.1. Ontologie et lexiques □

On en reste généralement à l'idée, bien antérieure à la formation de la linguistique, que les langues sont des nomenclatures (et non des corpus de textes produits dans des pratiques différenciées) : leur lexique serait une représentation de choses qui, comme l'affirmait déjà Aristote, sont « les mêmes pour tout le monde ». Cette conception antique informait la scolastique, et, par le biais d'auteurs comme Stuart Mill, elle a été reformulée dans la philosophie du langage qui sert aujourd'hui de cadre conceptuel au cognitivisme orthodoxe.

Cependant, voici deux siècles, l'essor de la linguistique comparée a permis de constater que les lexiques des langues avaient des structures variables et ne « reflétaient » aucunement un même « monde ». Mais ces résultats, approfondis par des auteurs aussi divers que Sapir, Hjelmslev ou Coseriu, ne tiennent aucune place dans la réflexion actuelle sur les ontologies, qui s'appuie sur des linguistiques universelles (la première théorie chomskienne a constitué le cadre conceptuel de Miller), et non sur les linguistiques générales et comparées.

Ainsi, dans le domaine des ontologies, on ne s'avise guère de cette évidence : « Les relations fondamentales qui structurent le lexique sont tout à fait distinctes et indépendantes des relations que nous concevons entre les référents » [6]. En effet, les constructeurs d'ontologies adoptent une démarche qui consiste précisément à normer des relations entre des référents supposés indépendants des langues, pour en dériver

- page19 -

les relations internes au lexique des différentes langues. En outre, le choix de l'ontologie suppose non seulement une préconception de l'objet scientifique, mais encore l'impose à cet objet : extraire des informations consistera à réduire un texte à ce qui reste accessible à des modèles ne reconnaissant que des entités prédéfinies et des relations typées.

Un exemple simple suffit à illustrer les enjeux pratiques de ces choix. Les grandes ontologies unilingues (projet WordNet développé à Princeton) ou multilingues (Eurowordnet) [7], restent partagées en secteurs différents, pour les noms, les verbes et les adjectifs, les adverbes, sans

aucune autre justification que des préjugés ontologiques – coûteux pour la pensée comme pour le contribuable – sur la référence supposée de ces parties du discours. En outre, elles réduisent ainsi à quatre les parties du discours qu'elles représentent. Enfin, elles ne traitent que de mots [8]. La notion de morphème, cruciale en linguistique, reste ainsi ignorée : par exemple, *coureur*, *course* et *courir* seront situés dans trois sous-réseaux différents, alors que la sémantique du morphème *cour-* reste identique. Pour résoudre ce problème, on a imaginé dans EuroWordNet la relation XPOS_NEAR_SYNONYMY qui permet de relier *arrival* et *to arrive*. [9]

Dans les langues indo-européennes, les parties du discours diffèrent essentiellement par des affixes, qui sont des grammèmes liés. Or la notion de grammème lié reste inconnue de la tradition de la philosophie du langage, qui a toujours considéré le mot comme une unité. Par exemple, la théorie médiévale des *modi significandi* postulait que *albus*, *albedo* et *albet* désignaient la même blancheur sous divers modes de signifier : c'est une solution analogue qui a été choisie dans EuroWordNet.

La grammaire scolaire, d'ascendance scolastique, estimait, sans aucun fondement d'ailleurs, que les noms représentent des entités, les verbes des procès, et les adjectifs des qualités [10]. Elle cousine en cela avec la philosophie du langage, qui en fait des catégories référentielles. John Lyons, référence indiscutée, rappelle les fondements ontologiques de la distinction entre entités, qualités et actions : « On considère généralement que les noms dénotent des entités, les verbes et les adjectifs des propriétés du premier ordre » (1980, p. 74) ; « le recours à de telles catégories présuppose un certain nombre d'hypothèses ontologiques minimales sur ce qui existe dans le monde. Les nôtres, que nous tenons pour minimales et généralement admises, sont celles du réalisme naïf. La première (et également la plus fondamentale) est que

- page 20 -

le monde extérieur contient un certain nombre de personnes, d'animaux et d'objets physiques plus ou moins discrets » (p. 77). Pour justifier son *réalisme naïf*, Lyons se fonde sur Strawson : « Toutes les entités du premier ordre (les personnes, les animaux et les choses) ont pour caractéristiques communes d'avoir des propriétés perceptuelles relativement constantes dans des conditions normales ; de se trouver, à tout moment, dans ce qui psychologiquement du moins est un espace à trois dimensions ; et d'être observables par tous (Strawson, 1959, p. 39) » (p. 78) ; « par entités du deuxième ordre, on entendra les événements, les processus, les états de choses, etc., localisés dans le temps, dont on dit en français qu'ils ont lieu, et non qu'ils existent. Par entités du troisième ordre, on entendra les propositions qui sont en dehors du temps et de l'espace » (p. 78) ; « elles sont telles qu'on dit plus naturellement d'elles qu'elles sont 'vraies' par opposition à « réelles » (p. 79). Fait d'affirmations non argumentées, l'exposé en reste au bon sens [11], c'est-à-dire au préjugé hélas commun que toutes les langues représentent les mêmes entités. Pour décrire leur sens, il faut donc décrire ces entités, que l'on discrimine en fonction de propriétés spatio-temporelles. Les critères sont évidemment discutables : en quoi *coureur* ou *ciron* sont-ils situables dans le temps et dans l'espace ? pourquoi *course* et *courir* représenteraient-ils des entités d'un ordre différent ?

Les oppositions entre les ordres d'entités structurent explicitement WordNet et EuroWordNet. Elle président en effet à la partition de ces réseaux en types d'entités, correspondant à des noms, pour le premier ordre, des verbes pour le second, et des adjectifs pour le troisième. Cette projection de catégories ontologiques sur les parties du discours dispense de faire référence aux travaux linguistiques sur les structures des lexiques : l'objectif demeure de construire une représentation conceptuelle des référents, en postulant que les lexiques des langues en constituent une représentation.

La tradition philosophique n'est pas réfléchie pour autant, et, en perdant sa dimension réflexive, devient un dogme qui sous-tend des opinions très communes en philosophie du langage et en sémiotique philosophique : « L'étude du langage mène naturellement à l'ontologie », écrivait par exemple Auroux (1996, p. 128). Nous souhaitons questionner ce prétendu naturel, car l'ontologie est présupposée par toute la tradition logico-grammaticale au fondement des sciences du langage.

Elle a présidé à la définition comme à la classification des signes et reste au fondement de la sémiotique contemporaine qui en dérive. Eco évoque ainsi avec un lyrisme discret « ce quelque chose qui nous conduit à produire des signes », que « nous nous décidons à appeler l'Être » et qui est considéré comme « l'horizon, le bain amniotique, dans lequel se meut naturellement notre pensée » (1997, pp. 4-8, *passim*), ce « socle dur de l'Être » qui garantit les limites de l'interprétation (1997, p. 35) [12].

2.2. L'Être et les êtres □

- page 21 -

Cependant l'unité de l'être est-elle compatible avec la multiplicité des êtres ? En quoi l'ontologie permet-elle les ontologies ? Platon se trouvait devant deux conceptions qu'il estimait également insoutenables : celle d'Héraclite qui pose la fluence et la multiplicité absolues de toutes choses, et celle de Parménide qui pose le statisme et l'unité absolue de l'Être, ce qui amènera les éléatiques comme Zénon à nier le mouvement. Il propose alors une position de compromis qui admet une multiplicité des essences ou Idées. Comme les idées servent à catégoriser les perceptions, Platon définit dans le *Phèdre* l'homme comme celui qui a vu les Idées, sans quoi il ne pourrait unifier les différentes perceptions sous une idée commune : l'idée a une fonction d'unité et l'Être en est une radicalisation. Ainsi, le concept de *prototype*, en qui Eleanor Rosch elle-même voit incidemment une sorte de « Platonic idea » a-t-il été accueilli avec enthousiasme, car il reformulait une solution millénaire : l'équipement cognitif inné rendait compte des prototypes et se substituait à la vision antérieure des Idées pour résoudre le problème de la catégorisation [13].

Restait à instaurer une relation de subsomption entre types et individus, de manière à rapporter chaque individu à son type (ou Idée), ce qui fut l'objectif d'Aristote dans sa théorie des genres et des espèces. Les dix catégories d'Aristote correspondent à ce qu'on appelle à présent les *top concepts* (on en compte 63 dans EuroWordNet). Critiquant la transcendance des Idées, Aristote reproche à Platon d'avoir « séparé les idées » et affirme leur immanence : dès lors, les objets cessent d'être des apparences et jouissent d'un être propre, leur substance invariable qui en fait des entités.

Pour rendre compte des variations des objets, Aristote édifie alors une théorie des accidents, considérés comme des variations inessentiels : ainsi, connaître consiste à éliminer les accidents : « Il n'y a pas de science possible de l'accident » parce qu'en fait il n'y a de science que du nécessaire, alors que l'accident, variable dans le temps, ne l'est pas. Sans contenu conceptuel, « l'accident n'a guère d'existence que par le mot qui le désigne » (*Métaphysique*, E, II, 1026 b). Ainsi, « L'accident se rapproche beaucoup du non-être » (*ibid.*), et une ontologie digne de ce nom doit l'éliminer. On retrouvera dans la théorie des réseaux sémantiques le problème des accidents, à propos des voitures à trois roues et des oiseaux qui ne volent pas, pour lesquels il a fallu élaborer une théorie des exceptions (cf. notamment les travaux de Brachman).

- page 22 -

En bref, la connaissance stockée par les ontologies se constitue en deux étapes : séparer les qualités accidentelles des qualités essentielles (définitoires), puis catégoriser les entités en les plaçant dans un arbre ontologique, au juste niveau hiérarchique, entre leurs superordonnés et leur subordonnés. Ainsi, dans EuroWordNet, le « synset » (ou classe conceptuelle) [14] {*car, machine, motorcar*} est relié à des concepts plus généraux, ceux du « synset » hyperonyme : {*motor vehicle, automotive vehicle*}, et par ailleurs à des concepts plus spécifiques de « synsets » hyponymes, comme {*cab, taxi, hack, taxicab*}.

2.3. Ontologies et hiérarchie □

Les rapprochements que nous venons de souligner témoignent d'une pérennité remarquable, comme si la stabilité de l'ontologie fondait celle de l'Être. Les principes de la construction ontologique restent invariables : il s'agit de subsumer la diversité des étants sous l'unité des concepts hiérarchiquement

supérieurs. D'où la permanence du modèle de classification des espèces, formulé par Aristote dans son ouvrage sur *Les parties des animaux* et développé par Linné. Ce modèle demeure au fondement de la théorie des réseaux sémantiques (cf. l'auteur, 1991, ch. IV), principal substrat figuratif de la « représentation des connaissances ». Un réseau sémantique est un graphe étiqueté connexe [15]. Soit l'exemple classique de Winograd (1977, p. 73) rappelé figure 1 :

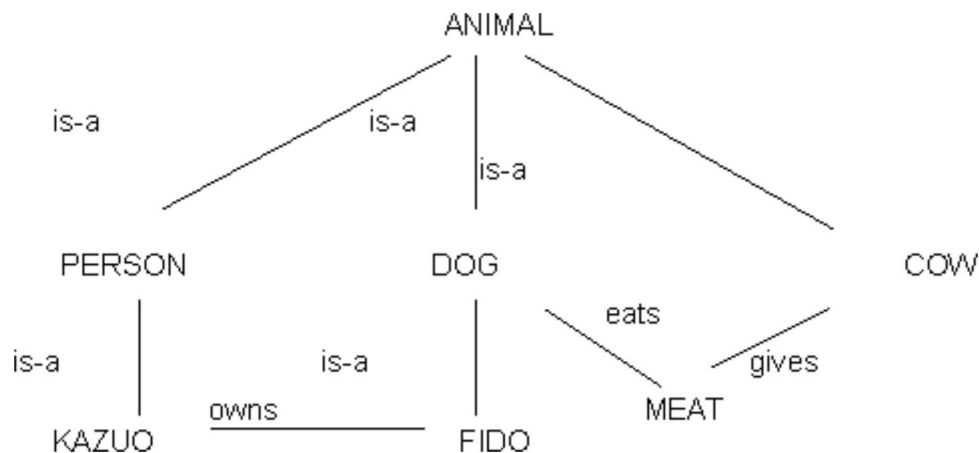


Figure 1. Un réseau sémantique

- page 23 -

Les nœuds du graphe sont des concepts, les arcs, ou liens (*links*) sont étiquetés par des relations sémantiques. La plus redondante est la relation *is-a* qui figure la hiérarchie ontologique entre la catégorie et la sous-catégorie ou l'individu ; ensuite vient la relation *part-of*, ou relation méronymique, qui figure la relation entre parties et tout. On note la polysémie débordante de *is-a*, dont Brachman (1982) a dénombré sept acceptions différentes : par exemple, la relation entre une classe et une sous-classe (ANIMAL et DOG) n'a aucunement le même statut logique que la relation entre une classe et un individu (DOG et FIDO) ; mais le logicisme fait souvent peu de cas de la logique.

Plus grave reste l'indifférence à l'égard des distinctions établies par les langues. Alors que leur lexique est en règle générale divisé par l'opposition /humain/ vs /animal/, qu'en français par exemple on distingue *bouche* de *gueule*, *jambe* de *patte*, etc., l'homme est ici un animal au même titre que le chien ou la vache, dans un univers ethnocentrique où les chiens sont nourris au bœuf. Ainsi, en concrétisant les préjugés du scientisme commun, la représentation conceptuelle, loin de représenter les langues, annule les distinctions sémantiques fondamentales qui structurent leur lexique.

La discordance entre la classification scientifique (l'homme est un animal) et la classification linguistique (l'homme et l'animal relèvent de dimensions sémantiques différentes, dont témoignent des classèmes distincts) pose un problème général : le lexique des langues ne reflète pas la conception scientifique du monde. Par exemple, en français, le soleil continue à *se coucher*, comme avant Galilée ; les *baies* restent de petits fruits généralement rouges, délicieux ou empoisonnés, sans souci de la botanique qui en fait des fruits à plusieurs noyaux, comme la tomate ou le melon.

La classification des espèces animales, d'Aristote à Linné jusqu'aux cognitivistes néo-darwiniens d'aujourd'hui, sert de pierre de touche aux théories de la classification et de la catégorisation ; ainsi, dans les discussions sur les réseaux sémantiques, on rencontre régulièrement Titi le canari, l'autruche qui ne vole pas et Clyde l'éléphant. Cependant, s'ils appartiennent à un même règne scientifique, rien ne permet d'affirmer que les animaux appartiennent à une catégorie unique. Le chat voisine dans la classe des animaux familiers avec le canari et le poisson rouge, l'autruche dans la catégorie des grosses bêtes africaines avec l'éléphant et la girafe ; bref, on ne peut trouver de discours (sinon peut-être celui de la psychologie cognitive) qui les juxtapose dans une même classe sémantique.

2.4. Liens verticaux et horizontaux □

Notons enfin une différence cruciale entre les liens soumis ici à un régime commun (cf. figure 1) : les liens verticaux sont des liens de catégorisation, les liens horizontaux sont des liens d'actance (figurés par des verbes, *owns*, *gives*, et *eats*). Les premiers sont définitoires et caractérisent la substance, puisqu'elle se définit par le genre et l'espèce, les seconds ne le sont pas : on ne peut dire au même titre que les

- page 24 -

chiens sont des animaux, qualité substantielle, et qu'ils se nourrissent de viande de bœuf, qualité accidentelle.

Les premiers relèvent d'un ordre ontologique, les seconds d'un ordre praxéologique. Malgré l'unité apparente du graphisme, ces ordres sont bien différents, voire inconciliables. En effet, l'ordre ontologique est caractérisé par les différences de niveaux d'abstraction qui permettent de passer des étants à l'Être, en l'occurrence des individus aux *top concepts*. En revanche, l'ordre praxéologique suppose une commune nature et une homogénéité des actants. Il est ici instauré, non au niveau des individus, mais des classes intermédiaires, celles des espèces : c'est le chien en tant qu'espèce (du moins le chien nord-américain) qui se nourrit de la viande que fournit l'espèce bovine [16].

Alors que l'ordre ontologique est celui de la permanence et de l'invariabilité, l'ordre praxéologique est celui de l'impermanence et de la variabilité. Le premier emprunte les voies de la catégorisation, et masque une subsomption de l'immanent (les étants) par le transcendant (l'Être) ; le second emprunte les voies du récit voire du mythe.

Au-delà se dessine une distinction fondamentale : les liens verticaux sont basés sur des formes de l'identité (ou de l'inclusion) et unissent les entités par subsomption, alors que les liens horizontaux opposent ces mêmes entités : dans le graphe, alors que le maître et le chien sont unis par leur nature animale commune, ils s'opposent par le lien de possession (*owns*).

Un des moyens d'oblitérer la différence entre relations verticales et horizontales consiste à ramener les liens casuels à des attributs substantiels, en antéposant le lien IS. C'est ainsi que procède Pustejovsky, dans *The generative lexicon*, où les « rôles de qualia » ou traits casuels se voient préfixés d'un IS : IS-FORMAL, IS-TELIC et IS-AGENT. Ce IS met sous la dépendance de l'Être les relations qui ne relèvent pas de l'ontologie classique. De la même façon, on affirmait jadis que tout verbe contenait implicitement la copule prédicative *être*, ce qui permettait de réduire les phrases actives à des phrases attributives : *il marche* devenait ainsi *il est marchant*, dans la grammaire de Port-Royal et toutes les grammaires philosophiques qui en dérivent.

Remarque. — Il est impossible ici d'évoquer la longue histoire des représentations arborées (cf. l'arbre de Porphyre). Massivement utilisés par la scolastique, ils ont structuré son discours, comme en témoignent les arborescences du plan des sommes. Ils procèdent de l'ontologie dans sa tradition métaphysique et sont à la base des ontologies actuelles.

- page 25 -

Une réflexion sémiotique sur cette forme graphique conduit à une critique des présupposés unifiants qu'elle concrétise : un format commun de représentation permet d'une part de représenter par les mêmes conventions graphiques des relations différentes, et d'autre part, d'unifier différents « niveaux de l'Être ».

3. De la hiérarchie ontologique aux relations sémantiques □

Les *relations sémantiques* sont cruciales dans les ontologies, puisqu'elles typent les liens entre les concepts (correspondant aux entités). Le mot *sémantiques* doit être ici entendu dans une acception

logique et non linguistique : il s'agit en fait de relations conceptuelles (bien que certaines soient empruntées à la tradition grammaticale, voire à la tradition rhétorique) : ce sont en effet des prédicats appliqués à des entités, le formalisme des réseaux sémantiques n'étant qu'une représentation graphique de la logique des prédicats du premier ordre.

3.1. Les trois types de relations lexicales □

Prenons l'exemple d'EuroWordNet, qui retient six sortes de relations conceptuelles : hyponymie-hyperonymie, holonymie-méronymie, synonymie, antonymie.

3.1.1. Les relations verticales et l'héritage □

Les premières spécifient les relations verticales de l'ontologie, et se résument à des formes de subsomption, de l'élément par la classe (relation *is-a*) ou de la partie par le tout (relation *has-a*). Ces relations sont déterminantes, car toute ontologie a une fonction unificatrice, l'Etre étant défini par son unité à soi.

Sous l'apparente unité d'une représentation graphique, les ontologies comme EuroWordNet juxtaposent de manière non-critique des relations qui relèvent de problématiques différentes : alors que les relations hyponymiques rendent compte de la référence, la liste des éléments pouvant passer pour l'extension du nom de la classe, les relations d'héritage relèvent de la problématique inférentielle [17]. La référence domine toutefois l'inférence, car elle permet la médiation entre l'universalité de l'Etre et la singularité des étants, alors que l'inférence se limite à l'héritage des propriétés essentielles [18].

- page 26 -

3.1.2. Les relations horizontales, synonymie et antonymie □

Les relations « horizontales », qui mettent en relation des entités de même niveau, restent plus intéressantes pour nous. Elles sont en effet sémiotiques et non conceptuelles (un concept reste indépendant de son expression, et donc il ne peut y avoir de relation de synonymie entre concepts) [19]. À proprement parler, seules des expressions peuvent être synonymes, dès lors qu'elles se réfèrent au même concept.

La synonymie revêt cependant dans WordNet et EuroWordNet une fonction cruciale, puisque les classes conceptuelles (*synsets*) y sont définies comme des classes d'équivalence synonymiques (*syn-* abrègé synonymie). Cette étrange synonymie affecte par exemple, dans le *synset* des moyens de paiement, le billet et la pièce de monnaie. L'antonymie s'exerce sur le fond de cette synonymie généralisée. Alors que les relations verticales rendaient compte de la référence et de l'inférence, l'antonymie relève de la problématique de la *différence*. En termes logiques, elle se spécifie en contraires et contradictoires. Comme la négation ne peut appartenir à une ontologie, les contradictoires n'y ont pas leur place [20]. La différence se résume alors à l'antonymie verbale, qui relie des contraires comme *cash* et *credit*.

3.1.3. L'articulation des relations verticales et horizontales □

Héritières des terminologies du positivisme logique, les ontologies sont des réseaux de concepts et n'ont donc pas de signifiants. Comment alors les articuler aux langues ? Ici intervient l'articulation de la synonymie et de l'hyperonymie : la première permet de réduire la diversité des signifiants (en fait des signes) ; la seconde permet de les subsumer sous des concepts. Ces deux subsomptions, des signifiants et des signifiés, permettent conjointement de réduire la diversité linguistique. On articule ainsi deux ordres de relations sémantiques : les relations hiérarchiques (onto)logiques, traitées par une logique des classes héritée de la syllogistique ; et les relations non hiérarchiques qui permettent de concevoir des graphes cycliques et non plus seulement arborés.

Articuler ainsi ces relations remplit deux objectifs : (i) réduire la diversité des signifiants par la

synonymie ; (ii) passer du signifiant au signifié [21], assimilé au concept. On passe ainsi des signes aux choses (par la référence) puis aux concepts (par la subsumption), étant entendu alors que les concepts subsument des choses et non des concepts individuels.

- page 27 -

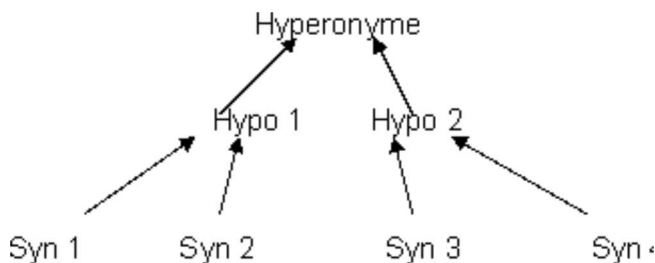


Figure 2. Des synonymes à l'hyperonyme

3.2. Questions sur les relations sémantiques privilégiées □

3.2.1. Hyperonymie □

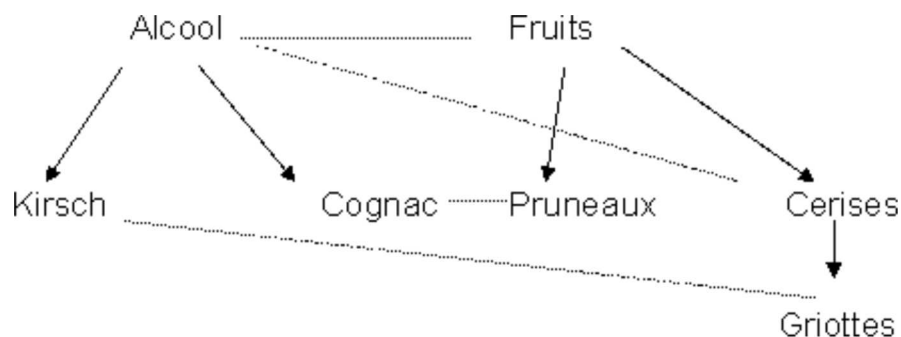
L'hyperonymie reste également privilégiée par la lexicographie (elle est présente dans toute définition de dictionnaire) et la lexicologie (dans les théories de l'archiséme ou de l'hyperonyme). Cette rencontre avec l'ontologie n'a rien d'étonnant, car les dictionnaires restent tributaires de la définition aristotélicienne par genre et espèce. Cependant, hormis des énoncés définitionnels du type « la chaise est un siège », on est souvent bien en peine de trouver le « genre » adéquat et le mot pour le dire : soit il n'y en a pas, soit il y en a trop, soit ils sont inutilisables, etc.

A fortiori, comme l'ontologie s'est constituée sur la définition des noms (supposés permanents) et non des verbes (qui marquent l'impermanence), on ne sait trop comment organiser les verbes en arborescences ; de même pour les adjectifs, les adverbes, et pire encore les autres grammèmes.

On n'en continue pas moins à structurer « taxinomiquement » le lexique, bien qu'il n'offre aucun repère coopératif pour une telle entreprise. Souvent, les hyperonymes n'appartiennent pas au même discours, ni aux mêmes genres, ni au même niveau de langue, ce pourquoi il n'est pas linguistiquement légitime d'unifier le lexique en un thésaurus structuré. Par exemple, on ne parle pas de *meubles* chez un antiquaire, mais de *voltaires*, *conversations*, *bergères* ou *chauffeuses*.

Même dans le « langage quotidien », qui devrait manifester uniment le « niveau de base », on rencontre des énoncés comme « De l'eau ou du beaujolais ? » qui « passe » tout aussi bien que « De la Badoit ou du vin ? ». Plus inquiétant, si l'on s'en tient aux documents attestés, le restaurant *Chez Henri*, rue Descartes, met à sa carte quatre desserts également recommandables : *Cerises à l'alcool*, *Griottes au kirsch*, *Pruneaux au cognac*, *Fruits à l'alcool* [22].

- page 28 -

Figure 3. *Troubles dans la classification*

De fait, les relations lexicales dépendent des types de discours, alors qu'en principe l'ontologie reste la même quel que soit le genre et le discours. Par exemple, dans le projet Safir sur la cogénération d'électricité, Grabar et Jeannin (2002) ont comparé les relations dans quatre corpus : – économie – écologie – réglementation – technique :

<i>relations/ corpus</i>	écologie	économie	réglementation	technique
hyperonymie	759	57	111	2141
méronymie	43	--	--	211
synonymie	494	389	35	255
antonymie	21	5	--	4

Tableau 1. *Poids relatifs de relations sémantiques en corpus*

Ces chiffres doivent être rapportés aux tailles des corpus : écologie : 266 873 occurrences ; économie : 422 589 occurrences ; réglementation : 322 048 occurrences ; technique : 198 007 occurrences. Ainsi, l'on compte 40 fois plus d'hyperonymes dans le corpus technique que dans le corpus économique. Alors que l'absence de méronymie dans l'économie et dans la réglementation reste

- page 29 -

inexplicable, la méronymie, comme les autres relations hiérarchiques, est florissante dans le corpus technique, mais rare voire absente dans les corpus théoriques : les corpus techniques sont en effet des corpus didactiques de montage et de démontage, ce qui se prête à l'expression de relations partie-tout.

3.2.2. *Improbables héritages* □

Une ontologie ne tient aucun compte de la diversité des discours, puisqu'elle a pour but d'affirmer l'unité de l'Être, mais cela affecte évidemment son utilité descriptive. En effet, il n'y a pas d'héritage dans les textes, sinon sous la forme d'anaphores : les hyperonymes n'ont pas de régime particulier, et tombent dans le cas des anaphores associatives. Souvent d'ailleurs, on reprend un terme par un hyperonyme : « En plat du jour, nous avons de l'églefin... Qui veut du poisson ? » (le patron du *Café du Midi*, Lyon). La substitution par un hyperonyme très élevé reste une figure de style : cf. *je vais promener l'animal* (au lieu de *Je vais promener le chien*). On objectera que les concepts de l'ontologie sont indépendants de leurs noms ; mais le rapport entre les usages (et les corpus qui les reflètent) avec les hiérarchies ontologiques qui sont censées en rendre compte reste fort problématique.

3.2.3. *Polysémie et synonymie* □

Comme les terminologies dont elles dérivent, les ontologies doivent réduire deux facteurs

complémentaires « d'imperfection linguistique », la polysémie et la synonymie : elles troublent toutes deux le rapport idéalement transparent qui, selon la conception instrumentale du langage, devrait exister entre le langage et la pensée. En outre, elles témoignent d'une multiplicité, alors que les ontologies sont des systèmes basés sur l'unicité : non seulement l'Être est unique, mais tous les nœuds d'une ontologie sont évidemment au singulier, comme les concepts l'ont toujours été.

Comme la sémantique lexicale est restée au centre de la sémantique, ces deux propriétés des signes ont donné lieu à une immense littérature. Notamment, on ne cesse de constater qu'un même signe peut avoir plusieurs acceptions ; des collectivités entières se sont structurées autour de cette question (*Word-Sense Disambiguation*) ; et bien des théories de la sémantique cognitive, comme celle des prototypes, ont précisément pour but de résoudre le problème de la polysémie.

La synonymie reste moins invasive, mais donne matière à divers produits comme des dictionnaires cumulatifs informatisés [23]. Elle inquiète, car elle semble perturber

- page 30 -

le statut ontologique du langage : si, comme le dit Aristote, l'Être se dit de plusieurs manières, sont-elles équivalentes ? Les théories de la référence sont là pour nous assurer que *l'étoile du matin* et *l'étoile du soir* sont bien le même astre (Frege) et autorisent bizarrement l'assimilation de ces deux objets linguistiques aussi différents que *le vainqueur d'Austerlitz* et *le vaincu de Waterloo*.

La polysémie comme la synonymie ne seraient-elles pas deux effets du dualisme qui sépare le signe de sa signification et donc deux artefacts de la tradition logico-grammaticale ? Faire varier les significations d'une expression ou les expressions d'un concept, cela suppose que ce concept et cette expression soient doués d'une existence hors contexte, essentielle par rapport aux variations en contexte : or, cette invariance demeure une des propriétés de l'Être, traditionnellement rapportée à une substance, alors que les variations sont considérées comme des accidents.

Selon qu'on considère le « concept » ou le signe (*i.e.* signifiant) comme un invariant ontologique, les variations corrélatives de son expression ou de son contenu apparaissent comme des accidents de cette substance. Pour la connaître dans son essence, il faudra donc en éliminer les accidents, c'est-à-dire assimiler entre eux les mots déclarés synonymes en les reliant à un unique concept (c'est le fondement des *synsets* de WordNet). Cette entreprise, dans les ontologies interlangues, conduit jusqu'à déclarer synonymes des mots de langues différentes.

Pour forger l'identité entre synonymes, le préjugé référentiel reste nécessaire : lui seul peut permettre de conclure par exemple que *ictère* vaut *jaunisse* ou *bicot*, *arabe* [24], sous le prétexte qu'ils renvoient aux mêmes référents et que leur substitution ne change donc par la valeur de vérité des propositions où ils se trouvent. Ainsi, l'on néglige les contextes d'emploi et les traits évaluatifs qui s'y propagent : par exemple, *espérance* semble synonyme d'*espoir*, mais revient plus fréquemment dans le discours religieux comme dans les passages des textes profanes qui font allusion à la religion.

L'existence même de la synonymie reste douteuse, depuis l'axiome de Prodicos de Céos (V^e a.c.), selon lequel deux synonymes ne sont jamais équivalents, jusqu'à la constatation de l'abbé Girard (1715) qu'il n'existe pas de synonymes parfaits. On pourrait même dire que la distinction des synonymes fut à la base de la dialectique, jusqu'au *differentiae verborum* de la seconde sophistique ; elle est au fondement de la réflexion sémantique des Lumières, comme à l'origine de la sémantique différentielle saussurienne. Allons plus loin : la critique de la synonymie est une

- page 31 -

condition pour constituer une sémantique linguistique – et une linguistique qui refuse la séparation des deux plans du langage imposée par le dualisme ontologique.

Le question de la polysémie n'est pas moins triviale que celle de la synonymie. De fait, pour

constituer ou constater la polysémie, on ne tient pas compte des contextes, puisqu'on juxtapose des acceptions qui n'ont pas la même histoire, ne se trouvent ni dans les mêmes discours, ni les mêmes genres, ni souvent dans les mêmes textes. En contexte, la polysémie est rarissime, ou prend plutôt la forme de l'équivoque, qui concerne tout autre chose : la plurivocité de parcours interprétatifs.

Ainsi, permises par le dualisme entre signe et pensée, la polysémie comme la synonymie sont suscitées par le refus de l'interprétation, qui se traduit en premier lieu par le suspens du contexte. Ce refus renvoie la signification à l'inconditionné plutôt qu'au multiple : Aristote, à bon droit, ne dit pas exactement que l'Être se dit en plusieurs sens, mais qu'il se dit diversement (*pollakos legomemon*) : *pollakos* est adverbial [25], et évoque ainsi une indéfinition plus qu'une pluralité. Il en va de même pour la synonymie : dans certains contextes et situations ou *rat* et *souris* pourront être jugés équivalents ; par exemple, dans un contexte didactique informel en classe primaire, on peut parfaitement définir un *rat* comme une *grosse souris*.

Si donc la polysémie et la synonymie sont des artefacts de la linguistique, cela tient au fait qu'elle néglige l'interprétation. Dès qu'on l'affaiblit, la polysémie et la synonymie explosent : cette indéfinition radicale touche aussi bien la langue (*love* peut être un verbe français ou un nom anglais) que la frontière des morphèmes : des informaticiens ont dénombré jadis 32.000 ambiguïtés dans la phrase *J'ai mal aux pieds* (par combinatoire : *j'aime halle*, etc.).

Le seul argument que l'on m'ait opposé pour justifier l'importance de la polysémie, c'est qu'on a besoin de la résoudre pour faciliter les traitements automatiques du langage : or, précisément, les systèmes informatiques sont dépourvus de capacités interprétatives, dans un état de l'art qui prime la représentation sur l'interprétation.

Bref, la polysémie ne relie pas les mêmes expressions à des idées différentes et la synonymie n'agit pas les mêmes idées sous des formes différentes. L'unité des deux plans du langage est telle que la différence des expressions entraîne celle des idées parce que des signes différents sont pris dans différents réseaux textuels et intertextuels. Par la sémosis, mise en relation des deux plans du langage, les deux « chaos », selon le mot de Saussure, « donnent un ordre » (*ELG*, I, paragraphe 11, pp. 51-52).

Ainsi, la synonymie comme la polysémie traduisent une angoisse ontologique sur la capacité représentationnelle du langage comme sur la stabilité du monde. Kant, qui pourtant marque une rupture sceptique à l'égard de la tradition ontologique, dans un des rares moments où il tient compte du langage, s'inquiète tant de la polysémie

- page 32 -

que de la synonymie, dès lors qu'elles ne seraient pas en quelque sorte garanties par l'ordre des phénomènes : la connaissance deviendrait impossible. Cette angoisse hante encore les théories cognitives, d'où par exemple le succès de la théorie des prototypes. Rien ne vient en effet fonder les représentations : celles qui sont induites par les pratiques sémiotiques appartiennent à l'ordre sémiotique, indépendamment de toute ressemblance (*homoïoma*) ou iconicité. C'est d'ailleurs, paradoxalement, cette absence de fondements dans la représentation d'êtres préexistants qui peut permettre de fonder une sémantique linguistique.

4. Éléments d'une dé-ontologie □

4.1. Philosophie du langage ou linguistique ? □

Le *langage* est comme on sait un concept philosophique, et la philosophie du langage ne se confond aucunement avec la linguistique, qui prend pour objet les *langues* et leur comparaison, et tient leur diversité pour le problème fondateur qu'elle tente de résoudre par la méthode comparative.

Depuis sa disciplinarisation voici deux siècles, elle s'est progressivement autonomisée à l'égard de la philosophie du langage. Cependant, avec l'essor du positivisme logique, qu'ont illustré Carnap, puis Chomsky (son élève), Montague et l'ensemble du paradigme formaliste, on a voulu faire de la philosophie formelle [26] l'organon de la linguistique. La théorie computationnelle de l'esprit a eu pour but de lui conférer une vraisemblance cognitive, si bien que la philosophie du langage est devenue une philosophie de l'esprit, comme l'a souligné Récanati. Elle a naturellement reçu un accueil chaleureux dans le milieu des traitements automatiques du langage.

Rappelons cependant quelques positions diversement argumentées ailleurs (l'auteur, 2002, 2003) et qui s'inscrivent dans une perspective déjà clairement articulée par Saussure [27] et des auteurs de la tradition saussurienne, comme Hjelmslev : abandon décisif de la référence, reconnaissance du statut linguistique (et non logique ou psychologique) des signifiés, indissolubilité du signe, solidarité des plans du langage, définition différentielle des signifiés et des signifiants, intégration de la linguistique à une sémiotique générale.

- page 33 -

Par exemple, les traits sémantiques qui constituent sans reste un signifié lexical (ou sémème) sont des « accidents », au sens où ils sont soit hérités du type lexical (relevant du discours et non de la langue) si le contexte n'y contrevient pas, soit propagés par prescription contextuelle. Ainsi, un signifié linguistique est un groupement stabilisé ou temporaire de traits qui seraient jugés accidentels dans une ontologie.

4.1.1. Lexique, langue et discours □

Nous n'irons pas plus avant dans le débat sur les relations sémantiques. Ce sont en général des relations logiques jadis abstraites à partir de relations linguistiques décontextualisées, et qui ont servi à construire la logique des classes ou la logique des prédicats du premier ordre. Mais depuis longtemps la logique a évolué ; elle s'est émancipée de son expression linguistique en se formalisant ; les logiques modales, les logiques plurivalentes sont apparues. Par contraste, la logique dont dépendent les ontologies reste fruste et passablement obsolète.

Par ailleurs, la sémantique a précisé son domaine et défini des relations non logiques comme les relations actantielles, et des relations textuelles complexes comme les métamorphoses (cf. l'auteur, 2001 b, ch. 1). Or, les traitements automatiques du langage ont à décrire, produire ou interpréter des textes (multimédia ou non), où le nombre et la nature des relations pertinentes n'est pas fixé *a priori* : elles s'établissent dans l'activité textuelle, mais leur nature et leur degré de complexité varient avec les discours, les genres, voire les moments du texte. Sauf à adopter un point de vue normatif, on ne peut projeter sur les textes un petit nombre de relations (onto)logiques. Projeter sur les textes les plus divers un même ordre ontologique conduirait à vérifier un préjugé structuré ; ce que décrivent les ontologies, c'est en effet un réseau de concepts supposé refléter rien de moins que l'ordre du monde. Au contraire, en fonction des applications, il faut construire et décrire les relations pertinentes.

L'ordre du lexique ne semble pas affaire de règles. Certes, les relations sémantiques en instaurent, en figeant et en qualifiant des parcours entre entités, mais vraisemblablement, le lexique n'a pas de structure globale. Les classes lexicales sont structurées (relativement à un discours voire à un genre), mais les domaines sémantiques qui correspondent aux discours ne sont pas uniformisés ni hiérarchisés entre eux. Bref, plutôt qu'à un arbre, le lexique d'un état de langue serait comparable à une savane arborée, où parmi une multiplicité de petites structures lexicales une discipline ou un domaine technique ont de loin en loin systématisé une terminologie.

Un dictionnaire « de langue » suppose une totalisation très utile en pratique mais illusoire en théorie. En effet, à la différence du lexique des morphèmes, le lexique des lexies ne fait pas partie de la langue. Quel que soit leur degré de figement, les lexies sont des syntagmes dotés d'une syntaxe interne, bref, déjà des phénomènes de discours susceptibles d'une évolution diachronique très rapide et qui peuvent être créés ou oubliés à tout moment. Une reconception du mot s'impose ainsi, pour le

- page 34 -

redéfinir comme un *passage* dans un texte et une mention dans un corpus (cf. l'auteur, 2002).

4.1.2. Le problème du signifiant □

En tant qu'idéalité structurée, une ontologie n'a pas d'expression (aucune forme métaphysique n'en revêt sans déchoir) et cela permet de la considérer comme une interlangue. Or les deux plans du langage, contenu et expression, sont indissociables, et les progrès de la linguistique de corpus permettent de confirmer de multiples corrélations entre contenu et expression. En d'autres termes, le sens d'un texte n'est pas déterminé par une ontologie, mais par ce rapport entre contenu et expression que l'on peut appeler la *sémiosis textuelle* et qui reste pour l'essentiel normé par son genre.

Dans la tradition ontologique, la division entre l'Être et les Êtants n'est vraisemblablement qu'une abstraction de la séparation entre le signifié et le signifiant. Or, les expressions sont liées à la matière : alors que le cognitivisme se veut résolument matérialiste, il est paradoxal qu'il ne traite pas véritablement l'aspect matériel du langage. Réduire, comme il l'entend, les concepts à leurs substrats neuronaux ne suffit point à résoudre le problème de l'expression.

On peut distinguer grossièrement quatre niveaux, selon un gradient qui va du variable à l'invariable, du contingent au nécessaire, du matériel à l'idéal.



Figure 4. *Des expressions aux catégories*

Le cognitivisme orthodoxe ne reconnaît que les deux niveaux supérieurs dans la mesure où il réduit les contenus à des concepts et ne traite pas des expressions. Par contraste, le saussurisme ne reconnaît que deux plans non isomorphes, celui du contenu et celui de l'expression (pour un développement, cf. l'auteur, 2003) ; aussi, l'analyse sémantique ne recourt point aux concepts et catégories *a priori* dès lors qu'elle a défini des procédures pour analyser les signifiés en les différenciant.

Décrire le lien indissoluble entre les deux plans du texte est une tâche essentielle pour la linguistique textuelle, mais aussi un gage de son applicabilité : en effet, tout ce qu'on appelle les indices d'expression sont extrêmement discriminants, car ils permettent de caractériser des textes dans leur intégralité, plan du signifié compris. Or ni les affixes, ni les ponctèmes, ni *a fortiori* les polices de caractères ou les étiquettes html ne figurent dans une ontologie. Cependant, l'expérience a montré la grande capacité discriminante de ces indices [28].

4.1.3. Le problème des inégalités entre signes □

Le postulat ontologique divise les signes selon qu'ils sont ou non des expressions référentielles (cf. e.g. Lyons, 1980, p. 80 sur les noms). La référence étant le moyen fondamental de l'ancrage ontologique, les expressions jugées non référentielles seront négligées. En structurant le lexique des « mots sémantiques », les ontologies démembrement le lexique effectif dont sont faits les textes. En effet, des catégories entières de signes, notamment les grammèmes liés, ne sont pas représentés

dans les ontologies. Ils sont en effet censés introduire des accidents : d'ailleurs leurs noms traditionnels, comme *flexions*, *cas* (latin *casus*, grec *ptosis*) désignent diverses « irrégularités » voire « déchéances » et illustrent le poids de l'ontologie sur la grammaire.

De fait, les nœuds d'une ontologie sont simplement des mots « sémantiques » lemmatisés : on écrit les noms, verbes et adjectifs au singulier car l'être est unique, on écrit les verbes à l'infinitif, car les formes temporelles ou modales sont censées introduire des accidents (puisque l'Être se définit par une permanence que le verbe retrouve dès lors qu'il est substantivé dans un infinitif).

4.1.4. Le problème des contradictions □

D'après Aristote, Platon excluait du nombre des Idées les négations et les privations [29]. En effet, une ontologie se doit d'être non contradictoire, car elle instaure une isonomie conforme au principe d'identité fondateur de l'ontologie. Les contradictions effacées, toutes les propositions qui composent une ontologie sont déclaratives. Cependant, les textes ne se privent pas d'être contradictoires : dans un discours, même scientifique, il existe plusieurs « points de vue » qui ne sont pas des descriptions équivalentes des mêmes choses ; leur multiplicité voire leur hétérogénéité revêtent une valeur constituante, ce pourquoi les théories du dialogisme ont rencontré un tel écho. Ainsi, pour rendre compte de paradiastoles

- page 36 -

comme *vos héros sont des assassins*, on est obligé d'avoir recours à plusieurs mondes ou univers, car il reste impossible d'unifier les références. Aussi par exemple, on ne peut construire une ontologie commune aux discours racistes et antiracistes, car ni les relations ni les « entités » ne se correspondent dans leurs écrits.

On pourrait certes assimiler l'ontologie à une doxa (ensemble d'opinions moyennes) érigée au rang de norme. Mais sauf dans les textes normatifs – et une ontologie est de fait un texte normatif –, une doxa ne règne jamais de manière homogène et sans contradictions. Plus radicalement, chaque texte crée des innovations sémantiques et dans cette mesure fait évoluer la doxa qu'il est censé exprimer.

4.1.5. La destruction du contexte □

Dans les textes, le contexte est constituant, car il témoigne de la détermination du global sur le local. En voici un exemple, qui intéresse la distinction entre identité et équivalence. Toutes les ontologies supposent que toutes les occurrences du même concept sont identiques, puisqu'elles visent la même référence. C'est ainsi que Platon traite le problème de la répétition dans le *Parménide* (147d) : « *Parménide* : Mais peux-tu prononcer le même nom plusieurs fois ou une seule fois ? *Aristote* : Plusieurs fois. *Parménide* : Est-ce qu'en prononçant un nom une fois, tu désignes l'objet auquel appartient ce nom, et qu'en le prononçant plusieurs fois, tu ne le désignes pas ? ou bien, soit que tu prononces une fois ou plusieurs fois le même nom, ne désignes-tu pas toujours de toute nécessité le même objet ? *Aristote* : Sans doute. »

À l'inverse de ce postulat philosophique incontesté dans notre tradition grammaticale, Saussure commente ainsi *Messieurs ! Messieurs !* : « Chaque fois que j'emploie le mot *Messieurs*, j'en renouvelle la matière ; c'est un nouvel acte phonique et un nouvel acte psychologique. Le lien entre les deux emplois du même mot ne repose ni sur l'identité matérielle, ni sur l'exacte similitude des sens, mais sur des éléments qu'il faudra rechercher et qui feront toucher de très près à la nature véritable des unités linguistiques [30] ». Saussure reformule ainsi un thème de la tradition rhétorique/herméneutique, repris aujourd'hui par l'ethnométhodologie : il reconnaît que chaque emploi est un hapax et que la linguistique peut caractériser les différences entre les occurrences d'une même acception.

4.1.6. L'oubli des textes □

Les traitements automatiques du langage ont affaire non à des mots mais à des textes. Et comme les

textes ne se réduisent pas à des mots, un traitement purement lexical est invalide. Or, les ontologies ne traitent que de mots, et encore pas tous. Par ailleurs, les textes relèvent de normes de genres et de discours, assez fortes pour que les « concepts » varient en fonction de ces normes. Alors que dans un discours il n'y a pas de polysémie, si l'on ne tient pas compte des différences entre discours, on

- page 37 -

crée de la polysémie : par exemple, un dictionnaire « de langue » est conduit à juxtaposer sous la même entrée l'*assiette* du cavalier (qui relève du discours de l'équitation) et l'*assiette* à dessert (qui relève du discours culinaire). Or ces deux acceptions ne se rencontrent pas dans les mêmes contextes ni en général dans les mêmes textes.

La purification du langage. — Les ontologies contemporaines sont issues du programme terminologique du Cercle de Vienne : pour le positivisme logique, l'objectif était de purger les langues de leurs défauts, projet qui s'inscrit dans la continuité des recherches sur les langues parfaites, recherches auxquelles nous devons notamment les langages formels (cf. l'auteur, 1995a).

Si l'on ne partage pas la théorie du métalangage que le positivisme logique héritait de la logistique russellienne, on peut décrire d'un point de vue sémiotique la construction d'ontologies comme le passage de textes linéaires à des réseaux graphiques. Les nœuds de ces réseaux sont remplis des mots prélevés dans les textes, écrits en capitales et/ou traduits en anglais pour leur conférer une aura conceptuelle. Les étiquettes des liens sont également des expressions en capitales et en anglais. Les relations textuelles, tant syntagmatiques que paradigmatiques, se transposent enfin dans les lignes qui figurent les arcs des graphes. Par ces trois types d'opérations, on construit ainsi un objet sémiotique hybride (linguistique, logique et graphique) qui revêt la forme familière d'une ontologie. Mais la critique des présupposés de cette transposition sémiotique reste à faire, car les modes d'interaction des sémiotiques hétérogènes qu'elle met en jeu ne sont pas problématisés.

4.2. À quoi servent les ontologies ? □

Malgré des coûts élevés, la construction d'ontologies ne connaît guère de difficultés de financement. En effet, une ontologie ressemble fort à un organigramme, où les *top concepts* seraient à la même place que les décideurs : on peut y voir un effet spéculaire, car nos organisations étatiques sont structurées comme des ontologies — en grec antique, le commandant suprême a d'ailleurs le même nom que le principe ontologique ou *top concept (arché)*. Aussi, les décideurs semblent nourrir une secrète tendresse pour ces formalismes qui expriment si bien la conception administrative de l'univers. Des voix insistantes et autorisées s'élèvent pour que la France participe au projet EuroWordNet, alors même qu'elle ne dispose pas d'un corpus de référence digne de ce nom. Or, EuroWordNet n'est pas construit à partir de corpus, mais voudrait imposer sa norme à leur étude.

Les « ressources » linguistiques utiles ne résultent pas de la projection d'ontologies sur des textes, mais dépassent par des études de corpus les ontologies préconçues (cf. par exemple le dictionnaire anglais COBUILD, construit par Sinclair

- page 38 -

et ses collègues). En outre, au lieu de concrétiser les préjugés des décideurs, les ressources doivent tenir compte des tâches, et plus généralement de la demande sociale. Or, les ontologies se présentent comme universellement utiles parce qu'elles ne sont construites en vue d'aucune application déterminée. Cependant, chaque application *détermine un régime de pertinence*, et donc sélectionne dans le domaine d'application des « entités » et des « qualités » propres ; en d'autres termes, les qualités définitoires d'une « entité » varient avec les « points de vue » qui la configurent : pour être adéquate, une ontologie se devrait donc d'être *ad hoc*, ce qui contredit le programme ontologique.

Enfin, à présent que les corpus se multiplient, s'étendent, et que les données deviennent envahissantes, de nouvelles applications apparaissent et les tâches évoluent vers la recherche thématique, la discrimination de sites, etc. : elles ne consistent pas à « comprendre » et à représenter, mais à identifier, caractériser et interpréter. Pour cela, les ontologies n'ont pas fait la preuve d'utilité pratique, même si elles concrétisent des préjugés théoriques fort anciens et fort répandus. On n'évitera pas la question : « À quoi sert effectivement ce type de ressources ? ».. Rien ne confirme que les résultats obtenus soient supérieurs aux petits glossaires interlangue du commerce, formes électroniques de dictionnaires bilingues. Ce ne serait pas la première fois qu'une formalisation fascinante en théorie se révèle oiseuse en pratique.

Si le problème à résoudre est celui de constituer des dictionnaires électroniques, il n'est nullement besoin pour cela de mobiliser l'ontologie. Pour leur donner une dimension multilingue, il en va de même *a fortiori*, car chaque couple de langues pose des problèmes spécifiques qui échappent à toute ontologie : notamment celui des rapports entre lexique et syntaxe (car des traits lexicaux d'une langue peuvent être rendus par des traits syntaxiques d'une autre), et celui des rapports entre discours et genres (car on ne traduit pas seulement de langue à langue, mais de discours à discours et de genre à genre).

Même dans l'objectif, d'ailleurs illusoire, d'une traduction mot à mot, une interlangue ou langage pivot « conceptuel » ne fait que compliquer le problème, car les mots ne sont pas des unités linguistiques. Aussi la traduction a pour but de transposer des morphologies textuelles : notamment des fonds (représentables par des isotopies) et des formes sémantiques (représentables par des molécules sémiques). Pour cela, des treillis de sèmes semblent préférables à des hiérarchies de « concepts » (cf. notamment Janssen, 2002, qui applique des principes de sémantique interprétative).

Pour une praxéologie linguistique. — Sortir de la métaphysique « de bon sens » concrétisée par les ontologies n'est pas seulement un agenda scientifique mais aussi un agenda technique, car les préjugés ontologiques compliquent singulièrement la mise en œuvre des applications.

Toute application relève d'une pratique qui dépend étroitement de sa situation historique et culturelle. Sans développer ici le programme d'une praxéologie

- page 39 -

sémiotique [31], rappelons qu'elle intéresse le caractère doublement situé de toute application. D'une part, les textes étudiés par la linguistique de corpus et les TAL dans leur ensemble sont bien entendu configurés par leur situation de production et d'interprétation initiale, et leur modélisation ou représentation ne peut s'en abstraire, notamment en raison de la diversité radicale des discours et des genres. D'autre part, l'application elle-même, en les transposant dans une autre pratique, recontextualise décisivement toutes les réécritures qu'elle opère, que ce soit dans le cadre traditionnel de la « représentation des connaissances », dans celui de l'extraction d'informations ou dans tout autre traitement relevant des TAL. Cet aspect doublement situé a été depuis l'antiquité problématisé par la philologie ; il est naturellement remis en valeur par la philologie numérique qui se développe aujourd'hui et qui est en passe de refonder les TAL sur d'autres bases que la philosophie du langage.

Remerciements : J'ai plaisir à remercier Natalia Grabar, Sylvianne Rémi, Monique Slodzian.

□

NOTES

[1]. Cf. l'auteur, 1991, ch. 4.

[2]. George Miller est universellement connu pour avoir donné à la psychologie cognitive la mission de vérifier les théories chomskiennes. Décoré par le président Bush en 1991, il se flatte d'avoir reçu de la CIA 3 millions de dollars pour le projet WordNet. En janvier 2002, Wordnet comptait 139.000 entrées. À la même date, Miller se

proposait d'y ajouter, entre autres, *troop movement*, *bacillus anthracis* et *Intifada*, comme de modifier les définitions de *lira*, *franc* et *mark*, qui disparaissaient de l'ontologie avec l'arrivée de l'euro.

[3]. Dans le milieu des recherches cognitives, on donne habituellement le titre de concepts à des mots anglo-américains écrits en capitales. On pourrait s'étonner cependant qu'un dictionnaire de l'anglo-américain, sous réserve d'aménagements mineurs, devienne l'ontologie d'un système multilingue. Mais le cognitivisme orthodoxe reste sans doute un des facteurs efficaces de la mondialisation.

[4]. Des congrès mondiaux se réunissent à présent, le premier à Mysore, en janvier 2002.

[5]. Cf. Aristote, *Métaphysique*, E, I, 1026 a : « s'il existe une substance immobile, cette substance est supérieure à toutes les autres ; et il y a une Philosophie première, qui est universelle par cela seule qu'elle est première ». L'immobilité était un privilège divin ; les ontologies en ont hérité le préjugé que les entités sont invariables, identiques à elles-mêmes et non contradictoires.

[6]. Jean-Pierre Dalbéra, 2002, p. 98.

[7]. Cf. Slodzian (1999). Notre exposé doit beaucoup à cette étude.

[8]. Rappelons que les mots, à la différence des lexies et des morphèmes, ne sont pas des unités linguistiques (leur délimitation repose sur des conventions graphiques contingentes). En revanche, ils sont considérés comme des unités par la philosophie du langage et la psychologie cognitive qui s'en inspire – et a inspiré WordNet.

[9]. POS abrège *parts of speech*.

[10]. Pour s'en convaincre, on cite toujours des verbes d'action et des noms d'objet, plutôt que des noms d'action et des verbes d'état.

[11]. Un bon sens d'ailleurs fort approximatif ; que sont par exemple « des propriétés perceptuelles relativement constantes dans des conditions normales » ?

[12]. *Socle dur*, *bain ammiotique*, *horizon*, les métaphores ne manquent pas pour décrire l'Être, pourtant considéré comme le fondement de tout sens littéral : par un paradoxe qui devrait inquiéter les tenants de l'ontologisme, ils ne peuvent décrire le sens littéral qu'ainsi. Déjà, en 1975, Eco faisait allusion au « soubassement de la sémiotique » qui « ne peut être exclu du discours de la sémiotique sans que n'en résultent des lacunes embarrassantes dans la théorie tout entière » (p. 33-35). Philosophiquement, Eco, qui a consacré sa thèse à Saint Thomas d'Aquin, se rattache à la tradition néo-thomiste.

[13]. Pour une discussion, cf. l'auteur, 1991, p. 188 sq.

[14]. Tiré de WordNet 1.5.

[15]. Pour un développement et une discussion, cf. l'auteur, 1991, ch. 4.

[16]. Notons une incohérence : Fido, en tant qu'individu, ne peut manger de la viande générique, mais bien celle de ruminants individués eux aussi. En outre, *eats*, pour se conformer à l'invariabilité ontologique, revêt ici une valeur itérative de présent d'habitude et non singulative de présent ponctuel.

[17]. On en reste à la syllogistique : si les éléphants sont gris, et si Clyde est un éléphant, alors Clyde est gris, voilà le genre de raisonnement qui est à la base de réseaux comme le *KL-one* de Brachman.

[18]. Malgré la pauvreté conceptuelle de ceux-ci, les occurrences héritent des types, et non l'inverse : en effet, les types sont posés comme des archétypes, prototypes ou idées platoniciennes, et non comme des abstractions obtenues à partir de classes d'occurrences. Ainsi, malgré leur richesse accidentelle, les occurrences héritent de la pauvreté essentielle des types.

[19]. Cela explique l'absence de la polysémie, puisque la polysémie est la propriété d'une expression : un concept ne peut donc être polysémique.

[20]. Dans WordNet, l'antonymie est considérée comme une relation établie entre des mots, mais non entre des concepts.

[21]. En fait, le concept ou le référent représenté par ce concept, selon qu'on se trouve dans un paradigme de la désignation directe ou de la désignation indirecte.

[22]. La clientèle quotidienne des membres du Collège international de Philosophie n'aura pas permis que ce restaurateur respecte les principes cognitifs les plus élémentaires.

Plus inquiétant encore, dans *La cuisine de A à Z, Soupes et potages*, Livre de Poche, UGE, Paris, 1975, p. 122, la recette de la soupe à l'oignon (express) commence ainsi : « Ingrédients (pour 4 personnes) : Une grande boîte de soupe à l'oignon. 1°) Préparer la soupe à l'oignon en suivant les conseils donnés sur la boîte ».

[23]. Voici la liste des synonymes de *bleu*, dans le dictionnaire élaboré naguère par l'Elsap (Cnrs) : *aniline, ardoise, azur, azuré, azuréen, azurin, azurite, beige, bizut, blême, blessure, bleuâtre, bleu-bite, bleuët, bleuisse, bleussaille, bleuté, câble, cache-poussière, céleste, cerne, cérulé, céruléen, cérulescent, châtaigne, ciel, cobalt, combinaison, commençant, conscrit, contusion, coquard, cotte, coup, cru, cyanose, débutant, Dieu, ecchymose, écolier, émeraude, gendarme, guède, hématome, indigo, indophénol, induline, interdit, lapis-lazuli, lilas, livide, meurtrissure, myosotis, noir, non-initié, nouveau, novice, œil au beurre noir, pâle, pantalon, pastel, pers, pervenche, pierrot, pneu, pneumatique, recrue, résorcine, roquefort, rosaniline, roussette, salopette, saphir, smalt, soldat, souquenille, stupéfait, surtout, tache, télégramme, torgnole, tournesol, tuméfaction, vert, vert de peur, violet*. L'absurdité imperturbable de cette énumération suffirait à ruiner l'idée que la synonymie reste un problème linguistique digne d'intérêt.

[24]. Un dérivé du dictionnaire de l'Elsap, consultable en ligne, donnait *bicot* et *bougnoule* pour synonyme d'*arabe*, ce qui a valu la fermeture temporaire du site qui l'abritait. Chacun sait à présent que *islamiste* n'est pas synonyme de *musulman*.

[25]. Pour un commentaire, cf. Aubenque, 1962, p. 173.

[26]. C'est le titre de l'ouvrage posthume de Montague, *Formal Philosophy*.

[27]. « ETRE. Rien n'est, du moins rien n'est absolument (dans le domaine linguistique). (...) aucun objet n'étant naturellement délimité ou donné, aucun objet n'étant avec évidence.

SUBSTANCE LINGUISTIQUE. – Nous tendons perpétuellement à convertir par la pensée en substance les actions diverses que nécessite le langage. (...) Il n'y a point à admettre de substance fondamentale, recevant ensuite des attributs.

TERME (cf. ETRE). – Il n'y a aucun terme définissable et valable hors d'un point de vue précis, par suite de l'absence totale d'êtres linguistiques donnés en soi. » (ELG, p. 81).

[28]. Sur tout cela, nous nous permettons de renvoyer à l'auteur, 2003. Nous nous appuyons sur l'expérience du projet européen *Princip*, consacré à la détection automatique de sites racistes.

[29]. Il excluait également les relations, les choses artificielles, et celles où il y avait de l'antérieur et du postérieur. Ce refus présage la distinction entre les nœuds de l'ontologie, qui représentent des entités, et les liens qui représentent des relations ; le caractère intemporel des ontologies, qui ne connaissent ni avant ni après ; enfin, la difficulté des cognitivistes à rendre compte des entités artificielles parmi les prototypes (cf. Rosch et le prototype des Volkswagen, l'auteur, 1991, p. 189).

[30]. 1972, p. 152.

[31]. On trouvera au besoin quelques propositions chez l'auteur, 1999, 2001 b, 2001c.

BIBLIOGRAPHIE

Aristote, *Organon*, trad. J. Tricot, Paris, Vrin, 1984.

Aristote, *Métaphysique*, Paris, Presses Pocket 1991.

Auroux S., *La philosophie du langage*, Paris, PUF 1996.

- Benveniste E., « Catégories de pensée et catégories de langue », in *Problèmes de linguistique générale*, I, Paris, Gallimard, 1966, p. 63-74.
- Cassin B., « Quand lire, c'est faire », introduction à B. Cassin (éd.), *Parménide, Sur la Nature ou sur l'Étant*, Paris, Seuil, 1998, p. 9-70.
- Aubenque P., *Le problème de l'Être chez Aristote*, Paris, Puf, 1962.
- Brachman R., « What ISA Is and Isn't », *Proceedings CSCI : SCIEO*, Saskatoon, University of Saskatchewan, 1982, p. 212-221.
- Dalbéra J.-P., « Le corpus entre données, analyse et théories », *Corpus*, 1, 2002, p. 89-104.
- Eco U., *Trattato di semiotica generale*, Milan, Bompiani 1975.
- Eco U., *Kant e l'ornitorinco*, Milan, Bompiani, 1997.
- Fellbaum C., *Wordnet, an electronical lexical database*, Cambridge (Mass.), MIT Press, éd. 1998.
- Fløttum K. et Rastier F., *Academic Discourse— Multidisciplinary Approaches*, Oslo, Novus, (éds) (2003).
- Gonseth, F., *Le problème de la connaissance et la philosophie ouverte*, Lausanne, L'âge d'homme, 1990.
- Grabar N. et Jeannin B., « 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.
- Janssen M., SIMuLLDA, a Multilingual Lexicon Database Application using a Structured Interlingua, Thèse, Université d'Utrecht, 2002.
- Lyons J., *Sémantique linguistique*, Larousse, 1980.
- Malrieu D. et Rastier F., « Genres et variations morphosyntaxiques », *Traitements automatiques du langage*, 42, 2, 2001, p. 547-577.
- Miller G. .A., Johnson-Laird, P. N., *Language and Perception*, Cambridge (Mass.), MIT Press, 1976.
- Miller G. A, Avant-propos à Christiane Fellbaum, *WordNet*, Cambridge (Mass.), MIT Press, 1998, p. 15-22.
- Parménide, *Sur la Nature ou sur l'Étant*, tr. B. Cassin, Paris, Seuil, 1998.
- Pustejovsky J., "The Generative Lexicon", *Computational Linguistics*, 17, 4, 1991.
- Rastier F., *Sémantique et recherches cognitives*, Paris, PUF, 1991.
- Rastier F., « Le terme : entre ontologie et linguistique », *La banque des mots*, 7, 1995a, p. 35-65.
- Rastier F., La sémantique des thèmes – ou le voyage sentimental, in *L'analyse thématique des données textuelles — L'exemple des sentiments*, Paris, Didier, 1995 b, p. 223-249.
- Rastier F., « Action et récit », *Raisons Pratiques*, 10, 1999, p. 173-198.
- Rastier F., « L'Être naquit dans le langage — Un aspect de la mimésis philosophique », *Methodos*, I, 1, 2001a, p. 103-132.
- Rastier F., *Arts et sciences du texte*, Paris, PUF, 2001b.

Rastier François, L'action et le sens. — Pour une sémiotique des cultures, *Journal des Anthropologues*, 85-86, 2001c, p. 183-219.

Rastier F., Saussure, « l'Inde et la critique de l'ontologie », *Revue de sémantique et de pragmatique*, 11, 2002, p. 123-146.

Rastier F., Le silence de Saussure ou l'ontologie refusée, in Bouquet, éd., *Saussure*, Paris, L'Herne, 2003, p. 23-51.

Rastier F. (sous presse) « Enjeux épistémologiques de la linguistique de corpus », in Williams G., éd., *Actes des deuxièmes journées de linguistique de corpus de Lorient*, Rennes, PUR.

Rastier F. et coll., *Semantics for Descriptions*, Chicago, Chicago University Press [avec la collaboration de Marc Cavazza et Anne Abeillé] 2002.

Robin L., (*La théorie platonicienne des Idées et des Nombres d'après Aristote*, Hachette, Paris, 1908).

Saussure F. de, *Cours de Linguistique générale*, éd. Tullio de Mauro, Paris, Payot 1972.

Saussure F. de, *Ecrits de linguistique générale*, Paris, Gallimard 2002.

Slodzian M., « WordNet et EuroWordNet – Questions impertinentes sur leur pertinence linguistique », *Sémiotiques*, 17, 1999, p. 51-70.

Strawson P. F., (1959) *Individuals*, Londres, Methuen [tr. fr. *Les individus*, Paris, Seuil, 1979].

Vossen P., « Introduction to EuroWordNet », dans *Computers and the Humanities*, vol. 32, n°2-3, Kluwer Academic Publishers, 1998, p. 73-89.

Vous pouvez adresser vos commentaires et suggestions à : Lpe2@ext.jussieu.fr

© *Texto!* décembre 2004 pour l'édition électronique.

Référence bibliographique : RASTIER, François. Ontologie(s). *Texto !* décembre 2004 [en ligne]. Disponible sur : <http://www.revue-texto.net/Inedits/Rastier/Rastier_Ontologies.html>. (Consultée le ...).