

<https://laurentbloch.org/BlogLB/Terminologie-thesaurus-ontologie>



Terminologie, thésaurus, « ontologie »

- Zinformatiques - L'informatique : science et industrie - Systèmes d'information -

Date de mise en ligne : mardi 4 avril 2006

Copyright © Blog de Laurent Bloch - Tous droits réservés

Sommaire

- [Le langage travestit la pensée](#)
- [Nomenclatures et thésauri](#)
- [Exactions ontologiques](#)

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"

"<http://www.w3.org/TR/REC-html40/loose.dtd>">

Terminologie, thésaurus, « ontologie »

Ce texte est extrait de mon livre [Systèmes d'information, obstacles et succès](#), paru en mars 2005 aux Éditions Vuibert.

[Page d'accueil de ce livre](#)

Le langage travestit la pensée

(et la réalité encore plus !)

L'enthousiasme engendré par une percée intellectuelle telle que le modèle objet (par exemple) peut susciter une tendance à en surestimer la portée, à en prendre les auteurs pour des démiurges². La propension à verser dans cette erreur résulte souvent d'une mauvaise perception de la complexité des rapports entre les mots inventés par les humains et les choses dans la réalité : des naïfs peuvent croire que ces rapports seraient en fait une relation isomorphe, c'est-à-dire qu'à chaque mot correspondrait une chose dans le monde réel, et que la structure des relations entre les mots refléterait simplement celle des relations entre les choses réelles. Nous allons très brièvement et très sommairement expliquer ici pourquoi le point de vue qui considère le monde des significations comme un simple reflet du monde réel nous semble faux, puis envisager quelques occurrences de cette erreur dans le monde des Systèmes d'Information. Nous renvoyons le lecteur qui voudrait prendre connaissance d'une véritable analyse philosophique de la question à la lecture, par exemple, d'un livre de Karl Popper [1] qui lui consacre des développements détaillés et approfondis. Roger Penrose [2] donne aussi des arguments contre la thèse de l'isomorphisme entre les mots et les choses. Nous avons d'ailleurs déjà rencontré plus haut dans ce livre un point de vue de ce genre (la « théorie TDQM »), et sa réfutation par Isabelle Boydens [3].

Pour parler du monde qui nous entoure, nous employons un langage fait de signes, et nous avons appris à interpréter ces signes de telle sorte que nous pouvons grâce à eux échanger avec nos congénères quelques informations qui éventuellement pourront nous renseigner (ou nous induire en erreur) sur certains aspects de certains objets du monde réel. Des propos de notre interlocuteur, inférer ce qu'il a voulu nous dire, ce qu'il « avait derrière la tête » et qu'il voulait nous communiquer, est déjà un exercice où le risque d'erreur (de malentendu) est important ; pour comprendre ces propos, nous devons posséder toutes sortes de compétences relatives au contexte dans lequel a lieu l'échange de signes et à la nature particulière de cet interlocuteur et de nos relations avec lui. Alors *a fortiori* entre les signes du langage qu'il aura employés et un objet, ou un état, ou un événement du monde réel, la distance laissée à l'interprétation est considérable, et c'est là tout le problème de la *représentation*, qui n'a pas une réputation de facilité. Croire dans ces conditions qu'un langage humain (tel que le français ou le bambara, au hasard) puisse rendre compte « exactement et naturellement » ne serait-ce que d'une partie infime de l'univers, ce n'est pas seulement nourrir une illusion, c'est se méprendre sur la nature même du langage, du monde et de leurs

rapports. N'oublions pas que « le langage travestit la pensée. Et notamment de telle sorte que d'après la forme extérieure du vêtement l'on ne peut conclure à la forme de la pensée travestie ; pour la raison que la forme extérieure du vêtement vise à tout autre chose qu'à permettre de reconnaître la forme du corps. » [4]

Victor Klemperer, philologue allemand qui a miraculeusement échappé au génocide, a consacré à la langue nazie un livre [5] où il écrit (page 35 de l'édition française *Pocket*) : « On cite toujours cette phrase de Talleyrand, selon laquelle la langue serait là pour dissimuler les pensées du diplomate... Mais c'est exactement le contraire qui est vrai. Ce que quelqu'un veut délibérément dissimuler, aux autres ou à soi-même, et aussi ce qu'il porte en lui inconsciemment, la langue le met au jour ». Ce passage, loin de contredire la citation précédente de Wittgenstein, la complète et en achève le sens, qui réfute le réductionnisme linguistique : comment, si le langage travestit le propos du locuteur et si sa langue révèle ce qu'il dissimule, peut-on espérer établir une correspondance isomorphe entre discours et réalité ?

Outre les difficultés liées à la labilité des langues humaines et à leur interprétation, il en est une autre qui s'oppose absolument au projet de décrire de façon exacte et complète une partie de la réalité, et William Kent [6] a attiré notre attention sur elle : « Il nous est impossible de tracer un cercle imaginaire autour d'un certain corpus d'information et de déclarer qu'il contient tout ce que nous savons au sujet d'une certaine chose, et que tout ce qui est contenu dans le cercle a trait uniquement à cette chose, et donc que cette information "représente" la chose. »

Nomenclatures et thésauri

Les inepties autour de l'aptitude supposée du modèle objet à mettre en correspondance les objets du monde, ceux de la pensée et ceux du langage ne sont que cuistrerie véniale et hybris modérée à côté des exactions de la branche de la psychologie cognitive qui élabore de soi-disant *ontologies*.

De longue date les spécialistes de la documentation élaborent des thésaurus, qui sont des vocabulaires contrôlés, normalisés et commentés. Leur utilité est grande pour indexer articles et documents : si chaque auteur donne pour son article les mots-clés qui lui passent par la tête avec des acceptations de son cru, les recherches documentaires risquent de devenir difficiles et peu fructueuses. Disposer d'un thésaurus qui explique quel mot utiliser dans quelle circonstance, et qui en donne les équivalents dans différentes langues, est éminemment précieux, même si le rêve d'un thésaurus complet et parfait est inaccessible, ne serait-ce qu'à cause des évolutions incessantes tant des sciences que des langues.

La démographie dispose ainsi du thésaurus POPIN, la biologie de *MeSH*. Tout ce que nous avons dit nous incite à la prudence à leur égard : de tels thésaurus ne sont utilisables que dans le cadre limité pour lequel ils ont été conçus, c'est-à-dire indexer des articles scientifiques. Et un thésaurus multilingue ne saurait prétendre abolir toutes les différences subtiles qui existent entre les langues et remplacer les dictionnaires linguistiques qui, pour chaque entrée, peuvent donner de nombreuses traductions. Il se trouve par exemple qu'en français le mot *grue* peut désigner un oiseau ou un appareil de levage. Le mot anglais *crane* possède les deux mêmes acceptations, ainsi d'ailleurs que son équivalent allemand *Kran*. Mais en français le mot *chemise* peut désigner soit un vêtement, soit une pièce de moteur à pistons : or ni l'anglais ni l'allemand ne possèdent ce couple d'acceptations pour un même mot.

Bref, selon que notre thésaurus sera ornithologique, mécanique, vestimentaire ou de travaux publics, il n'aura pas la même structure. Et encore ne s'agit-il ici que d'exemples simplissimes pris dans des langues très voisines.

Exactions ontologiques

Les travaux issus de cette entreprise terminologique laborieuse et méconnue mais utile à la science ont été détournés de leur objectif par des cognitivistes qui se sont dit un peu vite que, puisque l'on avait des mots, cela ferait bien l'affaire pour représenter des concepts, et que dès lors que l'on disposait de concepts, pourquoi ne pas les relier par des flèches qui matérialiseraient, par exemple, les formules de la logique du premier ordre, avec ses prédicts qui permettent le calcul logique. Les flèches seraient bien sûr instanciées dans des bases de données informatiques par des pointeurs ou des références, et l'on pourrait ainsi donner une structure formelle de la science et distinguer les énoncés vrais des faux à“ rien de moins. Et tant qu'à faire, on appellerait cela une ontologie, ou plutôt *des* ontologies, une par domaine scientifique, par exemple.

Il y a là une confusion tout à fait dommageable entre les objets de la linguistique, ceux de la logique et ceux de la métaphysique. Les logiciens sérieux ont compris depuis longtemps que leur discipline devait se doter de ses propres formalismes justement parce que les langages humains étaient bien trop mobiles et fluides pour des opérations formelles. Comment formuler des propositions sûres avec les mots et les constructions grammaticales du langage humain, dès lors (pour ne prendre qu'une objection entre cent) que telle langue va exprimer telle signification par un procédé lexical, alors que telle autre le fera par un procédé syntaxique ? Ainsi, le français dira « j'ai oublié ma valise à la maison » là où le bambara dira « ma valise est restée à la maison », et d'ailleurs le verbe bambara qui signifie « oublier » n'admet que des compléments d'objet indirects, il n'existe aucun moyen direct d'oublier quelque-chose. Les ambiguïtés du langage humain ne sont d'ailleurs pas un simple problème de syntaxe, ce sont les limites mêmes au traitement de l'information par l'esprit humain. Dès la fin du XIXe siècle Gottlob Frege [7] avait compris les difficultés liées à l'usage du langage humain et avait entrepris d'élaborer un système formel pour les démonstrations logiques. Il suivait d'ailleurs en cela un chemin déjà ouvert par George Boole (1815 à“ 1864).

Certains cognitivistes contemporains semblent ne jamais avoir entendu parler de ces travaux, ou alors ne pas en avoir compris la portée. Il en résulte des confusions dramatiques, exprimées par exemple ainsi [8] : « les terminologies ne peuvent plus se contenter de recenser les termes et les organiser [sic] brièvement sous forme d'une hiérarchie. Elles doivent également proposer toute une gamme de relations qui reflètent au mieux les connaissances du domaine et répondent de manière adaptée aux besoins des applications. » On peut lire aussi, plus loin dans le même article qui fait une recension de travaux marquants de ce domaine, à propos d'un système formel à usage médical construit autour de la terminologie SNOMED : « Les règles du système formel utilisent et explicitent les relations, transversales ou non, qui peuvent exister entre les composants du nouveau terme [construit par le système, nda] mais aussi entre le nouveau terme et ses composants. Une règle de formation d'un diagnostic par combinaison d'un terme signifiant une région du corps (axe Topologie de la SNOMED) et d'un terme signifiant une pathologie (axe Morphologie) est par exemple instanciée dans la combinaison de termes élémentaires *glande apocrine* et *inflammation* (pathologie). La nouvelle notion *inflammation de la glande apocrine*, désignée également par le terme *hidrosadénite*, est une *inflammation* qui est *localisé-dans* la *glande apocrine*. Cette règle très productive est également à l'origine d'autres diagnostics... Le potentiel combinatoire des 5 880 termes de l'axe Morphologie de la SNOMED avec les 12 936 termes de l'axe Topographie, permettrait de créer 76 millions de concepts... » On croit rêver à“ et on espère ne jamais tomber entre les mains de médecins convertis à des systèmes formels de cet acabit. On remarque aussi que les opérateurs du système, écrits dans une police de caractères particulière, comme *localisé-dans*, en reçoivent une aura informatique spéciale qui semble autoriser les fautes d'orthographe.

François Rastier [9] a donné une critique salubre de cette insondable cuistrerie, en passant en revue un certain nombre de confusions et d'énormités repérées dans la littérature des « ontologies ». Empruntons-lui ce rappel aux définitions :

« L'ontologie se définit [...] comme la “science de l'Être” : elle reste constitutivement métaphysique, car la métaphysique est la “science de l'Être en tant qu'Être”.

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. »

En nommant « ontologies » des nomenclatures terminologiques ou des thésaurus, et en affirmant que les termes qui y figurent, arrangés selon des principes qui feraient recaler à ses examens tout étudiant en linguistique, sont ainsi érigés au rang de concepts, ces cognitivistes balayent d'un revers de la main vingt-cinq siècles de pensée philosophique et deux siècles de linguistique pour ne nous donner en échange qu'un positivisme caricatural.

L'encre des paragraphes précédents à peine sèche, il convient de préciser ceci : il existe quelques logiciels parfaitement raisonnables, et même recommandables, dont le seul défaut est de se prétendre ontologiques. Il s'agit tout bonnement de carnets de notes électroniques et hypertextuels, qui permettent de gérer et d'indexer des données peu structurées, ce qui est utile, et l'informatique est bien adaptée à un tel usage. Et bien sûr le slogan « ontologie » est plus éclatant que « bloc-note électronique », il faut bien le reconnaître. Se classent aussi dans cette catégorie des logiciels capables de comparer des sites Web consacrés à des thèmes voisins et d'établir entre eux des interliens vers des notions communes, c'est utile et intéressant, mais appeler cela le « Web sémantique » est pour le moins abusif : ici comme là, le péché vénial d'enflure verbale finit par confiner à l'imposture intellectuelle, ce qui est plus gênant.

Isabelle Boydens a écrit un livre (dérivé de sa thèse) [10]

où, entre autres analyses pénétrantes dont nos chapitres précédents se sont déjà fait l'écho, elle traque dans les bases de données et les systèmes d'information des erreurs du même ordre que celles que nous évoquions quelques lignes plus haut. Elle y démonte des théories qui laissent croire qu'entre les symboles d'un langage et les objets d'un univers réel qu'il peuvent désigner il pourrait y avoir une correspondance isomorphe, c'est-à-dire telle qu'à chaque symbole corresponde sans ambiguïté un objet et un seul de cet univers, et que l'univers puisse être totalement décrit par le langage. Ainsi munie d'une armure à l'épreuve des syllogismes creux et des impostures techno-scientifiques, elle entreprend une analyse du système d'information de la sécurité sociale belge sur laquelle tout candidat à la construction d'un SI administratif ou de gestion devrait se pencher.

Nous croyons en avoir assez dit pour que le lecteur, surtout s'il a pris soin de consulter les ouvrages que nous avons cités, soit convaincu de l'impossibilité d'un isomorphisme entre quelque univers réel que ce soit et un système d'information quel qu'il soit. Un SI n'est par nature qu'une représentation, une abstraction qui ne retient à propos de la réalité qu'elle représente que les informations qui importent à l'objectif poursuivi, exprimées avec la précision nécessaire, mais pas plus. Cette conclusion ne vise pas à abolir tout espoir de construire un système d'information pertinent et utile, mais elle veut souligner les difficultés de l'entreprise et fixer les bornes des ambitions qu'elle pourra se donner.

²La tentation de se croire égal aux dieux, d'excéder ses limites, porte le nom savant d'*hybris* ; nous avions eu l'occasion à la section ² d'en signaler une occurrence, qui consistait à imaginer que justement le modèle objet pouvait représenter naturellement le monde réel. Nous avions dit que cette hypothèse n'avait tout simplement aucun sens, parce que rien ne peut « représenter naturellement le monde réel ». © copyright Éditions Vuibert & Laurent Bloch 2004, 2005

- [1] Karl Popper. *La connaissance objective*. Aubier à€ Flammarion (pour la traduction française de Jean-Jacques Rosat), Paris, [1979] 1991.
- [2] Roger Penrose. *L'esprit, l'ordinateur et les lois de la physique*. Interéditions (pour la traduction française de Françoise Balibar et Claudine Tiercelin), Paris, 1992.
- [3] Isabelle Boydens. *Informatique, normes et temps*. Bruylant, Bruxelles, 1999.
- [4] Ludwig Wittgenstein. *Tractatus logico-philosophicus*. Gallimard (pour la traduction française), Paris, [1918].
- [5] Victor Klemperer. *LTI, la langue du IIIe Reich*. Reclam Verlag [Albin Michel pour la traduction française], Leipzig [Paris], 1975 [1996].
- [6] William Kent. *Data and Reality. Basic Assumption in Data Processing Reconsidered*. Elsevier North Holland, New York, 1981.
- [7] Gottlob Frege. *Écrits logiques et philosophiques*. Le Seuil, Paris, 1971 [1879-1925].
- [8] Natalia Grabar Thierry Hamon. « Les relations dans les terminologies structurées : de la théorie à la pratique ». *Revue d'intelligence artificielle*, 18, 1/2004.
- [9] François Rastier. « Ontologie(s) ». *Revue d'intelligence artificielle*, 18, 1/2004. article publié en ligne ici : <http://www.revue-texto.net/inedits/...>
- [10] Isabelle Boydens. *Informatique, normes et temps*. Bruylant, Bruxelles, 1999.