



**HAL**  
open science

## Ontologies et interprétation des textes sur le Web

Pierre Beust

► **To cite this version:**

Pierre Beust. Ontologies et interprétation des textes sur le Web. 16e Journées de Rochebrune (Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels) "Ontologies", 2010, 11 p. hal-01066831

**HAL Id: hal-01066831**

**<https://hal.science/hal-01066831>**

Submitted on 22 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Ontologies et interprétation des textes sur le Web.

Pierre Beust

GREYC CNRS UMR 6072 – ISLanD &  
Pôle ModesCoS de la MRSH  
Université de Caen - Basse Normandie  
14032 Caen Cedex, France  
pierre.beust@info.unicaen.fr

**Résumé :** L'interprétation des textes en IA est un domaine où l'utilisation et la conception d'ontologies est très fréquente. C'est le cas dans le projet dit du Web Sémantique. La subjectivité de l'utilisateur (voir d'un groupe d'utilisateurs) est particulièrement importante dans l'interprétation des textes or la question de la subjectivité est souvent duale de celle de la description ontologique. Les approches dites centrées-utilisateurs, comme celles du Web 2.0. sont sur ce point scientifiquement opposées à la démarche du Web sémantique. Nous verrons en quoi la théorie de l'enaction permet de trouver un cadre de réflexion scientifique intéressant pour la mise au point de systèmes centrés utilisateurs.

**Mots clés :** Terme, Concept, Représentation, Sens, Interprétation, Enaction

*Il ne faut pas se demander ce que sont des représentations ou ce qui se passe quand on se représente quelque chose, mais comment on sert du mot représentation.*

Ludwig Wittgenstein, *Investigations Philosophiques*, §370.

Cet article s'inscrit dans le domaine du Traitement Automatique des Langues (TAL) appliqué à la problématique de l'interprétation des documents textuels sur le Web. Depuis les début de l'Intelligence Artificielle (IA) et plus particulièrement depuis le *Knowledge Level* de A. Newel, la question des connaissances et de leur rationalité revêt une importance fondamentale en informatique. Dans bon nombre de développements (les systèmes experts par exemple) la dimension algorithmique des architectures devient même assez seconde par rapport à celle des connaissances manipulées. Le TAL en tant que branche issue de l'IA n'échappe pas à cette remarque. Les questions des ressources lexicales, des bases de concepts et de la représentation du monde sont presque toujours incontournables dans les architectures (en dehors des quelques rares systèmes dits endogènes qui travaillent sans ressources préalables<sup>1</sup>). Les développements informatiques pour instrumentaliser des tâches de veille documentaire sur Internet, de gestion de boîtes de courriers électroniques, de consultation de forums sont en plein essor, aussi bien au sein de sociétés de développement informatique privées qu'au sein de laboratoire de recherche publics. C'est un marché très concurrentiel en même temps qu'un champ de recherche actif où de la problématique de la description ontologique est bien d'actualité.

Dans cet article je développe l'idée que l'interprétation des textes n'est pas un domaine comme un autre (pas comme le diagnostic médical, la simulation de crises ou encore la vie

---

<sup>1</sup> Cf. par exemple (Déjean 98)

artificielle par exemple) relativement à la question de l'utilisation des ontologies. C'est un domaine où la subjectivité de l'utilisateur (voir d'un groupe d'utilisateurs) est particulièrement importante or la question de la subjectivité est souvent duale de celle de la description ontologique. Cela ne veut pas pour autant que les particularités subjectives des utilisateurs et les ontologies soient absolument incompatibles. C'est ce qu'on observe dans des démarches dites centrées utilisateurs à travers des ressources dites termino-ontologiques.

Dans une première partie de l'article je rappellerai quelques oppositions qui permettent de mieux caractériser les différences de fond entre la description de termes et la description de concepts. A la lumière de ce rappel, j'expliquerai dans une deuxième partie en quoi le projet du Web Sémantique est à mon avis un cadre d'utilisation et de développement d'ontologies scientifiquement mal posé. J'expliquerai alors pourquoi les approches centrées utilisateurs me semblent beaucoup plus pertinentes pour aborder la question de la signification des textes. Enfin un détour vers la théorie de l'enaction permettra de trouver un cadre de réflexion pour la mise au point de systèmes centrés utilisateurs.

## 1. Terme, concept

Quand il est question de mettre au point un système expert, il faut recueillir la connaissance d'un domaine et la formaliser. Les notions fondamentales de concept et de relation que l'on trouve dans la majeure partie des ontologies font alors très bien l'affaire. Quand on fait du TAL, le domaine c'est la langue. Une confusion est assez fréquemment faite quand on modélise les signes linguistiques par des concepts et les règles de fonctionnement de la langue par des relations entre ces concepts. Or termes et concepts ne sont pas du tout de même nature, de même que la langue n'est pas une base de connaissances, mais une activité sociale et culturelle.

Je ne soutiens pas ici l'idée que lorsqu'on fait du TAL on doit nécessairement bannir les ontologies. Il y a des modélisations ontologiques des langues qui sont à fait pertinentes. Par exemple en syntaxe, les concepts de groupes verbaux et de groupes nominaux et les relations qu'ils entretiennent ne sont évidemment pas à mettre en cause. Ce que je précise c'est que la description de ce qui fait sens dans l'activité interprétative n'est pas de même nature qu'une description ontologique en terme de concepts et de relations.

La confusion terme/concept vient probablement de l'histoire du TAL en tant qu'héritage de l'Intelligence Artificielle ainsi que de quelques définitions assez floues et toujours actuelles de ce qu'est une ontologie. C'est le cas par exemple de la définition que l'on trouve sur Wikipedia<sup>2</sup> :

*« ... une ontologie est l'ensemble structuré des termes et concepts fondant le sens d'un champ d'informations ... L'ontologie constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que les relations entre ces concepts. »*

Dans cette définition on observe que les notions de termes et concepts initialement distingués se retrouvent au final absorbés dans celle de concept et même dans celle de donnée. Cette définition cache même, de manière un peu plus profonde, une autre confusion de taille celle entre représentation et sens. Le sens en langue n'est pourtant pas la représentation d'une donnée, c'est une activité d'intercompréhension (cf. Brassac & al. 96, Beust 07).

Cherchant à préciser l'utilisation des ontologies dans le domaine de la représentation de connaissance, Thomas R. Gruber (Gruber 93) donne plusieurs critères définitoires dont ceux

---

<sup>2</sup> [http://fr.wikipedia.org/wiki/Ontologie\\_\(informatique\)](http://fr.wikipedia.org/wiki/Ontologie_(informatique))

qu'il appelle la clarté, la cohérence et l'extensibilité. Rappelons les définitions de ces 3 critères :

- La clarté : La définition d'un concept doit faire passer le sens voulu du terme, de manière aussi objective que possible (indépendante du contexte). Une définition doit de plus être complète (c'est-à-dire définie par des conditions à la fois nécessaires et suffisantes) et documentée en langage naturel.
- La cohérence : Rien qui ne puisse être inféré de l'ontologie ne doit entrer en contradiction avec les définitions des concepts (y compris celles qui sont exprimées en langage naturel).
- L'extensibilité : Les extensions qui pourront être ajoutées à l'ontologie doivent être anticipées. Il doit être possible d'ajouter de nouveaux concepts sans avoir à toucher aux fondations de l'ontologie.

Concernant la clarté, si l'on s'accorde sur le fait que la définition d'un terme a rapport à son sens il est peu probable que l'on puisse la donner indépendamment d'un contexte. Il n'y a qu'à regarder un dictionnaire pour s'en convaincre. Le recours au contexte a valeur de règle plus que d'exception. D'autre part il est illusoire de penser que la définition d'un terme puisse être complète. A ce sujet, C. Kerbrat-Orecchioni montre que la définition d'un terme est liée à un contexte dont on ne peut fixer de limites :

*"Si je décris par exemple une consultation médicale hospitalière, le contexte ce sera le hic et nunc de la consultation, mais aussi l'hôpital particulier où elle a lieu et son fonctionnement, l'institution hospitalière en général, l'ensemble du système des soins tel qu'il fonctionne en France, et à la limite, la société française dans son entier." (Kerbrat-Orecchioni 96)*

Concernant la cohérence, la question de la contradiction entre deux définitions de termes est un non sens. Serait-ce une contradiction que deux termes aient le même sens ou que l'un en ait plusieurs. Si tel devait être le cas, il n'y aurait pas de synonymie ni de polysémie en langue. Il n'y aurait pas non plus de méronymie car une expression comme *boire un verre* apparaîtrait comme une contradiction. Au final on aurait du mal à dire autre chose que des énoncés du type *le chat mange la souris*.

Concernant l'extensibilité, il est bien évidemment impossible de prévoir quoi que ce soit dans la façon dont le matériau lexical va évoluer d'un point de vue sémiotique. Les langues, à l'inverse des ontologies, évoluent par leur usage dans une logique syncho-diachronique. Ces évolutions peuvent être très profondes. Une ontologie peut faire l'hypothèse d'un monde clos figé, ce qui n'est pas vrai des langues, en dehors justement des langues dites mortes.

Ces trois critères suffiraient déjà pour rejeter toute assimilation entre terme et concept. Pour compléter et enfoncer le clou on peut encore faire les remarques suivantes :

- Un terme est un signe socialement institué alors qu'un concept peut tout à fait être une construction d'un individu seul.
- La définition d'un concept c'est son contenu en soi (le plus souvent justifié de manière référentielle) alors que celle d'un terme est une différenciation avec d'autres termes. (Cf la valeur chez Saussure). La référence au monde est seconde dans la signification des termes alors qu'elle est au premier plan dans la définition d'un concept. Il apparaît que des critères sociaux et culturels sont plus justes pour les termes. Par exemple, (Rastier, 1987) rapporte le résultat d'une étude sur le contenu du mot *caviar* suite à une enquête au sein d'une population de collégiens. Il apparaît que le trait luxueux était le plus fréquemment cité tandis que d'autres traits ontologiques tels que

granuleux ou encore salé n'étaient jamais évoqués.

- Un concept est une représentation, un terme est un signifiant.
- Ce qui justifie un terme c'est son utilisation alors que la justification d'un concept répond à des prescriptions vériconditionnelles.

S'il m'a semblé ici nécessaire de faire ce rappel c'est bien parce qu'on peut constater que le projet de description ontologique des langues est bien toujours d'actualité. Effectivement dans un bon nombre d'approches sur les documents numériques, comme celles du Web Sémantique que j'évoque dans la partie suivante de l'article, on constate que l'idée dominante est que le texte porte de l'information et qui plus est de l'information représentable de manière ontologique. Du coup l'enjeu consiste à extraire cette information qui en tant que connaissance ontologique doit pouvoir s'énoncer et se représenter indépendamment du texte. Or il n'y a pas de moyens de dire le sens d'un texte en dehors d'un autre texte. Dit autrement, il n'y a pas de réalisations extra linguistiques du sens. Le sens n'est pas une connaissance, donc pas non plus une représentation, c'est une activité, une activité interprétative.

## 2. Le web sémantique

On attribue la paternité du projet du Web Sémantique à Tim Berners-Lee, initialement le créateur du langage HTML et le fondateur du W3C. En mai 2001, il définissait le projet de Web sémantique de la façon suivante :

*« The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in coopération »<sup>3</sup>.*

Pour Tim Berners-Lee, il s'agit d'arriver à un Web intelligent, où les informations ne seraient plus simplement présentes mais d'une certaine façon également « comprises » par les machines. Ainsi le Web sémantique permettrait de rendre des contenus de pages web interprétables non seulement par l'homme mais aussi par la machine pour autoriser des recherches d'informations plus précises et efficaces. Un des points de départ des recherches actuelles dans le cadre du web sémantique est le concept de métadonnées, comme données décrivant des données. Ici les contenus des pages web (textes, images, etc) sont vus comme des données qu'il faut annoter par des métadonnées pour en représenter le sens de manière computationnellement manipulable.

Le Web sémantique propose aujourd'hui essentiellement des méthodes, des langages et des techniques pour d'une part enrichir le contenu de des pages, d'autre part permettre l'exploitation de ces enrichissements sémantiques à l'aide de différentes ressources ontologiques. La représentation ontologique est donc au centre du projet et elle donne lieu à la création de plusieurs langages issus des technologies XML, par exemple RDF (Ressource Description Framework)<sup>4</sup>, OWL (Web Ontology Language)<sup>5</sup> basé sur RDF, ou encore DAML (DARPA Agent Markup Language)<sup>6</sup>. Ces langages sont conceptuellement issus des travaux

---

<sup>3</sup> <http://www.w3.org/2001/sw/EO/points>

<sup>4</sup> <http://www.w3.org/RDF/>

<sup>5</sup> <http://www.w3.org/2004/OWL/>

<sup>6</sup> <http://www.daml.org/>

en logique formelle développés dès les débuts de l'IA. Notamment OWL<sup>7</sup> est présenté comme basé sur les logiques de description, elles mêmes issues des réseaux sémantiques de Quillian (Quillian 68) et des *Frame* de Minsky (Minsky 75).

Dans la perspective des formalismes logiques, l'interprétation deviendrait un mécanisme d'inférence basé sur l'ontologie, une lecture analytique au sens de L.C. Paquin<sup>8</sup>, qui pour le coup pourrait s'affranchir des langues comme en témoigne le paragraphe suivant extrait d'un site officiel francophone sur le Web Sémantique<sup>9</sup> :

*« Les premiers efforts se sont naturellement attaqués aux multiples ombres relevant de l'ambiguïté des langages naturels ; ambiguïté à peu près surmontée par les hommes mais très handicapante pour les machines. Ainsi entre l'analogique humain et le langage binaire, s'édifient des référentiels hybrides servant d'espace de conciliation des deux ordres de contrainte. ... Une fois établies toutes ces formes de passerelles, les machines se chargeraient seules de s'accorder sur le contenu de toute interrogation et de sélectionner les meilleures réponses. On renoue ici avec les promesses théoriques de l'intelligence artificielle. »*

On constate notamment à la lecture de ce site que le Web Sémantique est un projet essentiellement technologique qui au passage laisse de côté un nombre important de questions scientifiques. Ainsi, les différences entre termes et concepts, entre connaissances et sens sont éludées mais, plus grave, la question de la nature de la signification est absente de la problématique du Web Sémantique. Il apparaît normal que le sens d'un texte puisse être porté par son auteur en annotation au texte. C'est pourtant une idée largement combattue par certains linguistes (Rastier, Courcil, Mauger par exemple) pour lesquels le sens est avant tout l'activité de l'interprétant, ce qui explique d'ailleurs une variabilité des interprétations d'un sujet à un autre qui va à l'encontre des principes du Web Sémantique.

D'autres critiques que celles-ci existent à l'encontre du Web Sémantique, elles aussi sévères. En témoigne le rapport de l'Action Spécifique 32 CNRS/STIC de 2003 (Charlet, Laublet et al., 03). Il dresse un constat sur l'avancement du projet du Web Sémantique où des investissements financiers et humains sont colossaux. La conclusion du document parle d'obstacles *« particulièrement cruciaux pour les débuts même du Web sémantique »* (ibid. p.123). Ces obstacles concernent entre autres l'absence de consensus sur les langages et ontologies à utiliser mais surtout, c'est la possibilité même de l'utilisation généralisée de méta-données qui est mise en doute : *« la détermination et l'ajout, même de simples méta-données, n'est pas une activité naturelle pour la plupart des personnes. (...) Les expériences dans la construction d'ontologies sont (...) instructives et pourraient contribuer à lever quelques illusions.»* (ibid. p.124).

Les principes d'organisation et de représentation ontologique sur lesquels repose le Web Sémantique sont applicables à l'échelle d'un système d'information fermé comme une entreprise ou une bibliothèque. C'est un cadre limité dans lequel la description ontologique est appréhendable dans sa globalité et donc possible. C'est ce que H. Assadi (Assadi 98) appelle les ontologies régionales. Par contre il est beaucoup plus douteux que ces principes puissent être applicables au monde dans son ensemble (c'est ce qu'a montré l'échec du projet

---

<sup>7</sup> OWL est défini en trois sous-langages de plus en plus expressifs, chacun étant une extension du précédent : OWL Lite pour les utilisateurs ne nécessitant que des contraintes simples dans une classification hiérarchique, OWL DL (DL pour Description Logic) et OWL full.

<sup>8</sup> <http://www.ling.uqam.ca/sato/publications/bibliographie/Terms.htm>

<sup>9</sup> Extrait de <http://websemantique.org/ReperePointDepart>, page consultée le 10/09/08

CYC de Lenat<sup>10</sup>). On peut donc s'attendre à ce qu'elles ne fonctionnent pas ou mal à l'échelle du Web. Les raisons ne sont pas techniques mais bien plus sémiotiques et sociales, ce que les défenseurs du Web Sémantique ne veulent visiblement pas voir.

### 3. Les approches centrées utilisateurs

Là où le Web Sémantique cherche à rendre le plus possible partagées de vastes ontologies (vecteurs de dominance culturelle, et plus précisément dominance d'une culture nord américaine) qui normalisent et synthétisent une connaissance pensée comme objective et devant convenir à tous les utilisateurs<sup>11</sup>, une autre approche existe et elle procède de manière duale en partant du principe que c'est avant tout la subjectivité des utilisateurs qu'il faut prendre en compte. Ces approches sont dites centrées utilisateurs. Elles donnent une importance particulière aux spécificités des utilisateurs ou des petits groupes d'utilisateurs et mettent éventuellement à profit de manière cumulative ces spécificités pour le plus grand nombre.

Depuis (Beust 98) nous avons proposé un modèle de structuration de ressources lexicales personnalisées. C'est un modèle qui propose à son utilisateur de structurer un domaine terminologique en terme de regroupements et d'oppositions de termes avant tout motivés par les besoins de l'utilisateur, sa tâche, son point de vue et ses centres d'intérêt. Il amène à constituer des ressources dites termino-ontologiques (RTO) au sens de N. Aussenac-Gilles, c'est-à-dire des ressources qui distinguent bien les aspects terminologiques et conceptuels comme par exemple les thésaurus ou encore les bases de données lexicales.

Notre modèle de création de RTO personnalisées (appelé LUCIA dans ses dernières évolutions) a été mis en œuvre dans les recherches doctorales de V. Perlerin et T. Roy sur l'accès aux contenus de documents électroniques, sur l'analyse des emplois métaphorique et sur la cartographie de corpus. Ces recherches ont mis en évidence la valeur ajoutée des RTO LUCIA dans les tâches interprétatives des utilisateurs en recherche d'information à partir du Web. Notamment il découle des ressources propres à un utilisateur ou un petit groupe d'utilisateurs une certaine *légèreté*, dans la mesure où elles ne représentent que ce qui est pertinent du point de vue de l'utilisateur et restent ainsi de taille raisonnable (par exemple une centaine de termes) ce qui les rend moins complexes à construire, à maintenir et à enrichir. D'autre part, l'aspect centré-utilisateur est une réponse au constat que dressent Didier Bourigault et Nathalie Aussenac-Gilles à propos de la variabilité des terminologies :

*(...) le constat de la variabilité des terminologies s'impose : étant donné un domaine d'activité, il n'y a pas une terminologie, qui représenterait le savoir du domaine, mais autant de ressources termino-ontologiques que d'applications dans lesquelles ces ressources sont utilisées (Bourigault & al., 2003).*

Les approches centrées utilisateurs dépassent largement la problématique du TAL. On en trouve plus généralement dans le domaine de la conception des logiciels et des systèmes d'information. C'est un même un domaine d'expertise qui donne lieu à des activités commerciales, comme en témoigne notamment la cible visée (*User Experience Design*) de la

---

<sup>10</sup> cf. (Lenat & al. 90)

<sup>11</sup> D'un point de vue très pratique, force est de constater que des ressources très généralistes et relativement perennes, valables pour tout type de traitement envisagé ainsi qu'à destination de tout utilisateur potentiel, ne sont pas facilement disponibles (sous forme électronique pour des traitements automatiques) et encore moins gratuites.

société de développement informatique Axance<sup>12</sup>. Il y a aussi un domaine où les approches centrées utilisateurs se développent massivement depuis quelques années (un peu près depuis la fin 2004) et de manière beaucoup plus visible pour tout à chacun, c'est le courant appelé Web 2.0. Un des principes du Web 2.0 est l'idée du Web collaboratif, c'est-à-dire l'idée que l'usage qu'ont certains utilisateurs du Web peut être cumulé de façon altruiste pour alimenter l'usage d'autres utilisateurs. C'est une démarche dite d'UGC : *User Generated Content* ou encore de *crowdsourcing*. La force d'une telle approche réside dans le nombre d'internautes concernés car bien qu'on estime à seulement 1% la part des utilisateurs altruistes qui sont générateurs de contenus<sup>13</sup> (les autres étant visiteurs/utilisateurs du contenu généré par les premiers), l'aspect cumulatif et la masse des utilisateurs fait qu'on arrive assez vite à des contenus très larges. Ainsi en s'appuyant sur le partage, la créativité, l'intelligence et le savoir-faire de tous on peut mettre en œuvre un service de recherche d'information aussi utilisable que ce que pourrait proposer un éditeur classique de site. C'est particulièrement ce que montrent des sites Web comme Delicious<sup>14</sup> ou encore Flickr<sup>15</sup>. Flickr est un site de partage et de recherche d'images. Delicious est un site de partage de favoris offrant un service de recherche d'information. Le point commun des deux sites réside dans le partage de *tags*, c'est-à-dire des étiquettes librement entrées (voir même spontanément entrées) par les utilisateurs, qui tiennent lieu d'entrées d'index pour les images ou les pages web. Ainsi plutôt que d'indexer de l'information relativement aux concepts d'une taxonomie (une classification hiérarchisée à la façon d'une ontologie) on indexe les contenus avec des termes choisis par les utilisateurs. C'est ce qu'on appelle en opposition à la taxonomie, la *folksonomie*<sup>16</sup>.

Pour bien mettre en évidence cette opposition, comparons par exemple le site de partage de favoris Delicious et l'annuaire de recherche d'information Yahoo<sup>17</sup> :

- Dans Yahoo l'indexation est faite par les catégories et sous catégories de l'annuaire, c'est-à-dire relativement à l'ontologie du Web que les auteurs de l'annuaire ont mis en place. Dans Delicious, l'indexation provient des tags choisis sans aucune contrainte ni justification par les utilisateurs quels qu'ils soient. Il en découle que les meilleurs index sont ceux qui reviennent dans le plus grand nombre de tags (c'est-à-dire les consensus dans la communauté des utilisateurs). Yahoo est donc onto-centré là où Delicious est socio-centré.
- D'un point de vue de l'évolution de l'indexation, Yahoo et Delicious sont encore diamétralement opposés : pour Yahoo il faut maintenir une ontologie sensée convenir à tous au risque de la voir devenir de moins en moins contrôlable (avec les problèmes de cohérences hiérarchiques et de recouvrements de catégories) alors que dans le cas de Delicious les index une fois agrégés donnent une valeur ajoutée qui indique l'évolution d'un courant dominant consensuel où il s'opère une modération naturelle des avis non consensuels par simple fréquence d'occurrence.

---

<sup>12</sup> <http://www.axance.fr/>

<sup>13</sup> 2% des utilisateurs de Wikipedia sont à l'origine de 75% du contenu  
(<http://atelierlorient.wordpress.com/2008/05/14/initiation-aux-notions-du-web-20/>)

<sup>14</sup> <http://del.icio.us>

<sup>15</sup> <http://www.flickr.com/>

<sup>16</sup> <http://enssibal.enssib.fr/autres-sites/docforum/documents/23&24nov06SavResPar06InterBroudouxE.pdf>

<sup>17</sup> <http://fr.dir.yahoo.com/>



- La nature des index est très différente d'un site à l'autre. Ils sont conceptuels chez Yahoo, comme le montre d'ailleurs l'usage exclusif de groupes nominaux pour étiqueter les concepts (par exemple, actualité, société, arts, sciences ...). Ce sont des termes chez Delicious qui peuvent être contextuels et/ou temporaire. C'est le cas par exemple du tag « à lire » (dont on peut constater au passage que ce n'est pas un groupe nominal) qui n'est pas assimilable à un concept et qui pourtant indexe 2570 pages (expérience faites le 10/9/08).
- Les tags sont importants essentiellement pour ce qu'ils excluent et pour les leurs utilisations conjointes statistiquement significatives (les *related tags*) alors que les concepts sont justifiés par ce à quoi ils réfèrent. Les tags sont à ce titre assimilables à des signes au sens de Saussure qui se définissent par la valeur qu'ils occupent dans le système qu'ils forment avec les autres.

On constate aujourd'hui que le système de partage de favoris généraliste Delicious semble avoir réussi le pari de fédérer des millions d'utilisateurs à travers le monde. Il peut ainsi concurrencer un véritable moteur de recherche. Plus généralement, c'est vrai de tous les sites Web 2.0 basés la génération collaborative de contenus. Les étiquettes sont une idée des plus simple et le coup de force des sites Web 2.0 c'est l'énorme valeur organisationnelle de l'information que cette simplicité engendre. En comparaison, la grande complexité que représente l'utilisation de vastes ontologies standardisées explique peut-être en partie que le Web Sémantique n'ait pas la même croissance que le Web 2.0. Le Web Sémantique que l'on présente parfois comme le Web 3.0 serait de ce point de vue en train de perdre son pari.

La simplicité est donc vertueuse. Ce n'est pas parce qu'une tâche est complexe, ce qui est indiscutablement le cas de la veille sur Internet, que les outils qui instrumentalisent cette tâche doivent aussi être complexes. Les outils les plus simples du point de vue des concepts et de l'ergonomie sont bien souvent ceux qui rencontrent le plus grand succès auprès des utilisateurs novices ou même experts. Par exemple on peut remarquer que la popularité de Google et sa supériorité par rapport aux autres moteurs de recherche s'est accrue notamment parce que son interface utilisateur très simpliste tranchait par rapport à ses concurrents. Cette vertu de la simplicité se confirme aujourd'hui dans les nouveaux produits de Google. En témoigne encore le navigateur Chrome<sup>18</sup> sorti par Google le 2 sept. 2008 et qui est présenté de la manière suivante « *Google Chrome vous permet de naviguer sur Internet plus vite, plus facilement et en toute sécurité, à l'aide d'une interface très simple qui sait rester discrète* ». Par ailleurs les logiciels les meilleurs en terme de résultats techniques ne sont pas toujours les plus utilisés si leur appropriation par les utilisateurs n'est pas facilitée. Un exemple flagrant est celui des traitements de textes : tout le monde s'accorde pour dire que Latex est un excellent traitement de texte qui est notamment bien meilleur que Microsoft Word et pourtant ce dernier est de loin le plus utilisé.

La question ici au centre de la comparaison entre les systèmes centrés utilisateurs et les systèmes qui contraignent les utilisateurs est celle du couplage personne/système. Ce sont les usages qui sont prépondérants et de la simplicité d'usage découle la qualité du couplage personne/système. Une théorie, ou plutôt un cadre de réflexion, pour étudier l'émergence d'un couplage personne/système existe selon nous. Il n'est ni une théorie de l'information, ni une théorie de la formalisation logique : c'est l'enaction.

---

<sup>18</sup> <http://www.google.com/chrome>

## 4. L'énaction

La théorie de l'énaction (ou cognition incarnée) a été proposée dans les années 1980 par le biologiste, neurologue et philosophe chilien Francesco Varela (Varela 96). Varela propose le concept d'énaction pour appréhender l'action adaptative de tout organisme vivant à son environnement. En réaction au cognitivisme, l'énaction fait prédominer l'action et la boucle action-perception à la représentation. Ainsi la cognition, du point de vue de l'énaction, n'est pas la représentation d'un monde pré-donné mais l'avènement conjoint d'un monde et d'un être vivant à travers l'histoire des diverses actions et interactions accomplies dans le monde par cet être. C'est la négation de l'idée d'un matériau ontologique du monde préexistant à l'action.

L'énaction est au départ une théorie du vivant mais cette théorie peut être réinvestie dans d'autres champs de recherche. Par exemple en systémique, il découlerait de l'énaction qu'un système autonome n'est jamais un système isolé car il est nécessairement couplé à un milieu ambiant et que dans ce milieu il maintient et fait évoluer continuellement son organisation en dépit des perturbations occasionnées par le milieu. Au sein des sciences cognitives l'énaction constitue un nouveau paradigme scientifique non plus inspiré par la métaphore de l'ordinateur comme dans le cognitivisme classique, mais inspiré par celle des organismes vivants. Ce paradigme unificateur des sciences cognitives qui vise la constitution d'une nouvelle communauté scientifique avec une réelle cohérence intellectuelle est en ce moment en pleine émergence. Ainsi en témoigne clairement la récente école thématique du CNRS organisée par l'ARCo (Association pour la Recherche Cognitive) du 29 mai au 3 juin 2006 à Ile d'Oléron (France) et intitulée « Constructivisme et énaction - Un nouveau paradigme pour les sciences cognitives »<sup>19</sup>.

Concernant le couplage personne/système, l'énaction conduit à examiner particulièrement la question des usages et des contournements d'usage dans l'interaction. Le contournement des usages par les utilisateurs n'est par principe jamais anticipé dans la conception des logiciels car seul ce pour quoi le logiciel est prévu est considéré. Pourtant l'informatique est depuis toujours un « lieu » de contournement d'usages par excellence car après tout les machines initialement conçues pour calculer servent aujourd'hui bien plus souvent à communiquer. Par ailleurs les logiciels qui par leurs fonctionnalités permettent de faire évoluer les usages pour lesquels ils sont prévus rencontrent en général un grand succès. C'est par exemple le cas des traitements de textes. Aujourd'hui, on écrit plus de la même façon avec les traitements de texte qu'avec un crayon et une feuille de papier car on tape souvent partiellement des bouts de texte que l'on complète, que l'on déplace, que l'on met en forme ou que l'on reformule en plusieurs étapes. Cette nouvelle façon de rédiger n'est pas explicitement prévue par le logiciel, elle est induite de ses fonctionnalités par les utilisateurs. Cela veut dire qu'elle s'inscrit dans des interactions et dans un couplage personne/système. Un autre exemple est celui des moteurs de recherche sur Internet que l'on peut tout à fait contourner pour vérifier si une expression se dit ou pas<sup>20</sup> ou encore si une traduction est bonne ou pas<sup>21</sup>. Cette idée que l'interaction personne/système amène à re-concevoir les logiciels est celle recherchée dans ce

---

<sup>19</sup> <http://liris.cnrs.fr/enaction/>

<sup>20</sup> Expérience faite sur Google le 22/8/06 : si on cherche l'expression "tirer avec des boulets rouges" on a 0 réponses alors que "tirer à boulet rouge" ramène 527 réponses. On sait alors quelle est la bonne expression.

<sup>21</sup> Expérience faite sur Google le 22/8/06 : si on veut retrouver que la bonne traduction en anglais de traitement automatique des langues est Natural Language Processing, il suffit de comparer différentes recherches : "natural language processing" (4 910 000 réponses) et "natural language computation" (279 réponses).

qu'on appelle les *interfaces énaclives* pour lesquelles il existe maintenant un réseau<sup>22</sup> de chercheurs et une conférence internationale<sup>23</sup> spécifiquement dédiée.

Le champ des applications numériques évolue avec les attentes actuelles des utilisateurs. La prise en compte de ces attentes et la façon dont utilisateurs s'approprient leurs outils au sein de sphères d'activités devenues numériques est une problématique informatique actuelle très forte qu'illustre bien le Web 2.0. En fait de développer des logiciels et des machines comme cela a été massivement le cas jusqu'à présent, il convient de proposer des environnements d'interaction avec les utilisateurs et entre les utilisateurs eux-mêmes. Dans cet objectif, une simplicité conceptuelle et ergonomique des outils logiciels proposés aux utilisateurs est un atout pour un meilleur couplage et finalement de bien meilleurs résultats. En permettant de théoriser le couplage personne/système on pourrait considérer l'énaclion comme une théorie qui sous-tend ou explique le Web 2.0 (à la manière par exemple de la théorie des ensembles qui explique les bases de données relationnelles).

## 5. Conclusions

Les sciences humaines et les sciences cognitives nous incitent à envisager le document électronique de manière indissociable à l'activité du ou des humains qui les produisent, recueillent, indexent et recherchent. Ainsi le champ de l'herméneutique nous montre à quel point les rapports intertextuels sont déterminants dans les interprétations des documents, et le champ de l'énaclion nous montre en quoi ces interprétations sont situées et subjectives. Les nouveaux outils du Web 2.0 qui privilégient les usages par rapport aux vastes ontologies l'ont bien compris.

Le courant du Web Sémantique porte l'idée que c'est l'ontologie (i.e. le matériau ontologique du monde, le réel ...) qui est commun dans l'intercompréhension entre les producteurs d'informations et ceux qui l'interprètent (quand bien même cette distinction serait valide). C'est faux. Ce sont les pratiques sociales et la culture qui sont au centre des mécanismes interprétatifs. Si le Web 2.0 prend actuellement de l'ampleur, ce n'est pas tant parce qu'il a choisi une autre voie que celle de l'ontologie, mais bien parce qu'il est en appui sur des réseaux sociaux et plus généralement une culture.

Relativement au domaine du traitement automatique des langues, il faut maintenant sortir d'une impasse scientifique qui chercherait une voie médiane entre une terminologie des concepts ou une conceptualisation des termes. D'une part, il faut redonner son importance à l'utilisateur car c'est lui en fonction de sa tâche, et plus particulièrement de son étendue (entre la recherche d'actualité sur le web et le classement d'un domaine), qui décidera de manipuler des termes ou d'utiliser des concepts. D'autre part, il n'est nullement prouvé que les utilisateurs soient en attente de machines qui sachent extraire et « manipuler » le sens des textes. Il n'est d'ailleurs pas moins prouvé que les textes soient des réservoirs de sens (☺). Les usages sur le Web montrent autre chose. Les utilisateurs ont besoin d'outils qui participent à leur compétence interprétative propres interactivement via un couplage personne/système. Ce sont les utilisateurs qui « projettent » du sens sur les textes relativement à leur pratiques sociales et à leur culture. Une machine peut tout à fait aider un utilisateur dans ce but sans nécessairement devoir inclure une représentation du monde, vaste et incontrôlable.

---

<sup>22</sup> <https://www.enactivenetwork.org/>

<sup>23</sup> <http://www.enactive2006.org/>

## Références

- Assadi H., 1998, *Construction d'ontologies à partir de textes techniques – application aux systèmes documentaires*, Thèse de doctorat en Informatique - Université Paris 6 (<http://www.lip6.fr/lip6/reports/1998/lip6.1998.048.ps.tar.gz>).
- Beust P., 1998, *Contribution à un modèle interactionniste du sens. Amorce d'une compétence interprétative pour les machines*, Thèse de doctorat en Informatique – Université de Caen Basse Normandie (<http://users.info.unicaen.fr/~beust/These/these.html>).
- Beust P., 2007, *L'approche interactionniste en Traitement Automatique des Langues, une discontinuité épistémologique ?* Journées de Rochebrune (Rencontres interdisciplinaires sur les systèmes complexes naturels et artificiels) "Catastrophes, discontinuités, ruptures, limites, frontières", Megève, France. ISSN 1242-5125 ENST S (Paris), p. 69-80.
- Bourigault D., Aussenac-Gilles N., 2003, Construction d'ontologies à partir de textes, Actes de *Traitement Automatique des Langues Naturelles (TALN)*, Tome 2, pp. 27-47.
- Brassac, C., Stewart, J., 1996. *Le sens dans les processus interlocutoires, un observé ou un co-construit ?* Cinquièmes Journées de Rochebrune "Du social au collectif ", 29 janvier - 3 février 1996.
- Charlet J., Laublet P., Reynaud C., *Web Sémantique*. Rapport de l'Action Spécifique 32 CNRS / STIC, 2003.
- Déjean H, 1998, *Concepts et algorithmes pour la découverte des structures formelles des langues*, Thèse de doctorat en Informatique – Université de Caen Basse Normandie.
- Kerbrat-Orecchioni C., 1996, *Texte et contexte*, Scolia, n°6.
- Lenat D., Guha R., 1990, *Building large knowledge based systems*, Reading, Addison Welsey.
- Minsky M., 1975, *A framework for representing knowledge*, In P. Winston ed., *The psychology of Computer Vision*, New-York, McGraw-Hill.
- Newel A., 1982, *The Knowledge Level*, Artificial Intelligencen Vol. 18, p. 87-127
- Perlerin V., 2004, *Sémantique légère pour le document*, Thèse de doctorat en Informatique – Université de Caen Basse Normandie (<http://www.revue-texto.net/1996-2007/Inedits/Perlerin/Perlerin.html>).
- Quilian M., 1968, *Semantic memory*, in M. Minsky : *Semantic Information Processing*, Cambridge, MIT Press, p. 216-270.
- Rastier F. (1987). *Sémantique interprétative*. PUF : Paris.
- Roy T., 2007, *Visualisations interactives pour l'aide personnalisée à l'interprétation d'ensembles documentaires*, Thèse de doctorat en Informatique – Université de Caen Basse Normandie (<http://roythibault.free.fr/these/index.html>).
- (de) Saussure F. (1915). *Cours de linguistique générale*. (1986) Mauro-Payot : Paris.
- Thomas R. Gruber - *Towards Principles for the Design of Ontologies Used for Knowledge Sharing* in Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer Academic Publishers, 1993
- Tim Berners-Lee, James Hendler, Ora Lassila, *The Semantic Web*, Scientific American, May 2001
- Varela F., 1996, *Invitation aux sciences cognitives*, Editions du Seuil.