

Discordances ontologiques et questions d'interopérabilité

Denis Berthier

Institut National des Télécommunications
Groupe des Ecoles des Télécommunications
9 rue Charles Fourier, 91011 Evry Cedex, France
<mailto:Denis.Berthier@int-evry.fr>
<http://www.carva.org/denis.berthier>

Résumé :

Au-delà de la recherche d'outils techniques pour les sciences sociales, celles-ci et l'informatique partagent deux *problèmes épistémologiques* majeurs : celui des ontologies (et de leurs discordances) et celui de l'intégration d'informations exprimées dans des ontologies différentes. Après avoir introduit les notions techniques d'ontologie et d'interopérabilité, nous montrons comment elles sont liées, comment elles peuvent être généralisées en préservant la nature de leurs liens, et en quoi elles sont pertinentes pour les sciences sociales.

Mots clés :

Epistémologie, ontologie, interopérabilité, intégration de schémas, homologie structurale

Summary :

Beyond the search for technical tools for the social sciences, the latter share with computer science two major *epistemological problems* : the problem of ontologies (and their discrepancies) and the problem of integrating information formulated in different ontologies. After introducing the technical notions ontology and interoperability, we show how they are interrelated, how they can be generalized so as to maintain the nature of their relationship, and how they are relevant for the social sciences.

Keywords :

Epistemology, ontology, interoperability, schema integration, structural homology

1. Introduction

L'informatique, on le sait depuis longtemps, en particulier depuis certains développements de l'intelligence artificielle (IA), n'est pas seulement un pourvoyeur d'outils techniques pour les sciences en général, et les sciences sociales en particulier. Elle soulève indirectement un certain nombre de questions qu'on considère habituellement comme portant sur la nature humaine. Dans notre livre « Le savoir et l'ordinateur », nous avons essayé de montrer qu'il était préférable d'aborder ces questions sous l'angle de la sémiotique et de la culture. Ici, nous concentrant sur un aspect particulier, nous allons montrer que les sciences sociales et l'informatique partagent deux *problèmes épistémologiques* majeurs : celui des ontologies (et de leurs discordances) et celui de l'intégration d'informations exprimées dans des ontologies différentes.

Deux questions majeures et étroitement liées parcourent en effet aujourd'hui tous les secteurs de l'informatique et des réseaux : la question des ontologies (et, ce qui vient aussitôt avec : le problème de leurs discordances), d'abord apparue sous ce nom en IA, et la question de l'interopérabilité des systèmes, d'abord apparue comme problème central du génie logiciel (c'est-à-dire de l'étude des méthodes de développement). Ces deux questions sont désormais omniprésentes en informatique. En particulier, elles concernent de très près les secteurs des bases de données, des systèmes d'information (avec les « composants métiers » ou « Business Objects ») et de l'Internet. Pour illustrer ce dernier cas, des enjeux économiques majeurs constituent une motivation puissante pour traiter le problème essentiel de la recherche d'informations pertinentes et de la synthèse d'informations en provenance de

sources multiples et hétérogènes (d'où les travaux sur le Web sémantique et le langage OWL), mais aussi de la spécification et de la recherche de services disponibles (d'où le secteur en pleine expansion des services Web et du langage WSDL). La multiplicité des ontologies développées dans les divers secteurs de l'informatique et des réseaux est telle qu'elles constituent à elles seules un témoin de la manière informatique de penser le monde – qui pourrait devenir un véritable terrain d'étude ethnologique.

Après avoir quelque peu précisé ces deux questions, ainsi que les liens qu'elles entretiennent, l'objet de cet article est essentiellement de montrer comment elles se situent et en quoi elles font sens bien au-delà de leur univers purement technique d'origine. En particulier, on repère comment la notion Lévi Straussienne d'homologie structurale se situe par rapport à celles d'intégration de schémas (dans les bases de données) et d'interopérabilité.

2. Ontologies

Le mot « ontologie » a été introduit en informatique par l'intelligence artificielle (IA). Dans tout cet article, nous adoptons le sens technique précis que lui a donné l'IA, en oubliant toutes ses connotations philosophiques : sommairement, c'est l'ensemble des concepts et relations d'un domaine et des axiomes logiques généraux qui les relie. Cette définition d'une ontologie est beaucoup plus générale qu'il ne semble au premier abord, car chacun des éléments suivants peut être considéré comme équivalent à une formule logique (cette équivalence est d'ailleurs explicitée de manière parfaitement rigoureuse dans le cas des « graphes conceptuels » de Sowa, 1984) :

- tout système de classification,
- tout système terminologique,
- tout schéma conceptuel d'une base de données,
- tout graphe ou ensemble de graphes de relations (jeux d'oppositions phonématiques, relations formelles entre actants d'un mythe, homologies structurales, etc.),
- tout modèle de « fiche » de recueil de données.

Précisons un peu ce concept.

2.1 Les ontologies et la modélisation

Dans la philosophie occidentale classique, l'ontologie est « tout simplement » la science de l'Être. En IA, le terme d'ontologie est utilisé dans un sens purement technique (heureusement plus facile à définir), conformément à la démarche scientifique de la modélisation ; il ne doit pas induire l'idée d'une quelconque naturalité des objets auxquels les diverses ontologies développées par l'IA se rapportent ; celles-ci, du fait qu'elles sont le produit d'une activité de modélisation, sont toujours liées à un objectif particulier, plus ou moins implicite.

En IA, l'ingénieur « cognitif » (terme un peu passé de mode) se retrouve en situation de devoir modéliser les concepts et connaissances d'un secteur d'activité qui n'est pas le sien ; pour cela, les méthodologies les plus répandues (comme KADS) lui recommandent d'adopter le vocabulaire (l'ontologie) du domaine habituellement utilisé par les personnes qui ont à traiter le même type de problèmes que le système qu'il essaie de développer. On notera l'utilité de cette dernière distinction et précision (vocabulaire du domaine vs vocabulaire du problème) ; par exemple, dans le domaine pharmaceutique, le vocabulaire n'est pas le même selon le problème à résoudre : pour le traitement médical, les trois « partenaires en ontologie » que sont le médecin généraliste, le pharmacien et le patient utilisent le nom commercial ; pour la recherche, le chercheur utilise le nom de la molécule ; mais on sait qu'il n'y a pas de correspondance bijective : savarine (nom commercial) = chloroquine + proguanil (noms de molécules).

Malgré ces méthodologies assez rigoureuses (sur le papier, au moins), je pense qu'en IA, on se trouve confronté à certains problèmes analogues à ceux de l'ethnologue. En particulier, il peut être fort difficile de ne pas plaquer ses propres concepts et conceptions, et sa propre manière de concevoir les liens pouvant exister entre divers concepts (sa propre « grammaire conceptuelle »), sur ceux qui ont

vraiment cours chez les experts du domaine (respectivement, chez les membres de la société qu'on veut étudier).

2.2 Ontologie formelle et ontologies régionales

En fait, pour lever toute ambiguïté sur le terme d'ontologie, il faut en distinguer deux usages, selon qu'elle est qualifiée de formelle ou de régionale. Quand on ne précise pas, on parle en général d'ontologies régionales.

L'*ontologie formelle*, pour Husserl, est concernée par les primitives épistémologiques générales nécessaires à la description du « monde ». Dans le sens plus précis que nous lui attribuons en IA, elle est relative aux primitives épistémologiques générales nécessaires à la formulation des diverses ontologies régionales ; elle se matérialise concrètement dans les primitives générales sur lesquelles sont construits les langages de représentation de connaissances (par exemple : distinction classe relation, distinction classe instance, relation classe sous-classe, relation objet partie, relation objet attribut, etc.). Il résulte de plusieurs théorèmes de logique formelle que toutes les ontologies formelles usuelles en informatique (langages à types abstraits, langages d'objets avec attributs, langages de règles, graphes conceptuels, etc.) sont formellement équivalentes à la logique du premier ordre (ou à une partie) : autrement dit, ajouter des primitives épistémologiques (comme le typage ou la structuration des objets) n'augmente pas la puissance expressive théorique. Le choix d'une ontologie formelle plutôt que d'une autre est donc *a priori* sans incidence fondamentale sur les ontologies régionales ; mais il peut avoir une incidence pratique considérable sur la manière de formuler des connaissances particulières.

Une *ontologie régionale*, au sens technique de l'IA, consiste en un ensemble de concepts spécifiques, relatifs à un domaine particulier, éventuellement des objets distingués appartenant chacun à l'une des classes définies par ces concepts, et un ensemble de fonctions et de relations sur ces classes et objets individuels, ainsi que certains types d'axiomes « terminologiques » régissant les différents termes de ce vocabulaire. Formellement, une ontologie régionale n'est ainsi rien d'autre qu'une théorie logique des types d'objets de son domaine et de certaines de leurs relations ; comme la plupart des théories logiques, une ontologie régionale est en général incomplète.

2.3 Ontologies régionales et théories

Si une ontologie régionale ne se distingue formellement en rien d'une théorie logique, elle le fait cependant par l'usage auquel elle est destinée : permettre de conceptualiser des situations multiples ou variables d'un domaine. Une ontologie régionale vise à imposer une sémantique minimale des concepts de son domaine, mais pas une théorie particulière de ce domaine ; elle doit pouvoir être exploitée par des théories diverses portant sur son domaine (qui en sont, formellement, des spécialisations). Décider qu'un axiome (parmi lesquels il faut compter les relations taxonomiques) se situe au niveau d'une ontologie régionale plutôt qu'à celui d'une conception particulière de son domaine relève souvent de préjugés culturels. En ce sens, toute ontologie véhicule implicitement une idéologie plus ou moins marquée, plus ou moins consciente.

Typiquement, une ontologie régionale définira :

1°) une taxonomie (ou classification hiérarchique) des types d'objets de son domaine (relations d'hyponymie, du genre : un A est_une_sorte_de B) ; exemple : $\forall x [\text{chien}(x) \Rightarrow \text{mammifère}(x)]$

2°) des définitions ; exemple :

$$\forall x \forall z \{ \text{grand-père}(x, z) \\ \Leftrightarrow \exists y [\text{père}(x, y) \text{ et } (\text{père}(y, z) \text{ ou } \text{mère}(y, z))] \}.$$

3°) la structure interne des objets du domaine (leurs attributs) et des contraintes sur les valeurs possibles de chaque attribut ;

4°) certaines relations structurelles binaires entre objets (du type A est_une_partie_de B) ; exemple : *partie_de*(bras, corps) ;

5°) certaines contraintes (sous formes d'axiomes logiques) auxquelles ces différents termes doivent être soumis pour que l'on puisse légitimement leur attribuer le sens qu'ils visent (par exemple, une ontologie des relations familiales imposera que la date de naissance du père soit antérieure à celle du fils ; mais elle n'imposera pas que l'adresse soit commune).

3. La notion classique d'interopérabilité en informatique

Passons au concept technique d'interopérabilité, que nous étendrons ensuite progressivement, tout en établissant les rapports avec la question des ontologies.

3.1 Interopérabilité technique des systèmes informatiques

Il est généralement admis qu'il serait illusoire de viser à une architecture unique des systèmes informatiques, qui leur permettrait de communiquer simplement entre eux, ou à un langage universel qui les autoriserait à travailler directement sur des objets et données communs. Une telle uniformité serait synonyme d'une part du blocage de tout progrès et d'autre part de l'impossibilité d'optimiser des langages ou systèmes à des fins particulières. Elle imposerait en outre la réécriture, économiquement impensable, de tous les logiciels existants. Il en est de même pour les systèmes de communication et d'échange d'information. *L'hétérogénéité des systèmes informatiques et de communication s'impose donc comme une réalité incontournable.* Bien entendu, l'interopérabilité ne saurait se présenter comme une alternative à la normalisation, toujours indispensable pour éviter une multiplication désordonnée et inutile de systèmes incompatibles ; elle en est au contraire le complément, les systèmes qu'on veut voir interopérer étant en principe déjà normalisés.

L'interopérabilité des systèmes informatiques est donc fondamentalement ce qui pallie leur impossible et indésirable uniformité, mais qui concerne des systèmes déjà normalisés (au moins de facto). Elle est ce qui permet de réaliser un compromis entre d'un côté hétérogénéité et évolutivité et de l'autre uniformité et normalisation.

Sans entrer dans le détail des câbles et connexions physiques qui forment désormais un treillis de plus en plus dense (mais au maillage très irrégulier) autour de la planète, l'interopérabilité des systèmes informatiques et de communication est assurée aux plus bas niveaux physiques et informatiques par la normalisation des interfaces physiques. Au niveau juste au-dessus, elle s'appuie sur les standards informatiques d'échange de données, c'est-à-dire, par exemple, sur les protocoles de l'IETF (tcp/ip, smtp, http, etc.).

3.2 Les bases de données : exemple classique de discordance ontologique ; interopérabilité des données et « intégration de schémas »

De multiples bases de données ont été développées partout dans le monde, avec des modèles de données différents (hiérarchique, relationnel, etc.) et des langages de requête différents ; en outre, leurs contenus sont organisés selon des schémas conceptuels (c'est-à-dire des vocabulaires et des ensembles de structures de données – ce qui constitue des formes simples d'ontologies) qui sont liés à des applications particulières et ont toutes les raisons imaginables d'être incompatibles. L'interopérabilité des bases de données consiste à pouvoir retrouver et intégrer des informations en provenance de bases hétérogènes relativement à ces divers points, et qui peuvent en outre résider sur des machines différentes avec des systèmes d'exploitation et des systèmes de gestion de bases de données différents. Au-dessus de l'interopérabilité technique générale définie ci-dessus, elle recouvre deux aspects, l'un syntaxique, qui est traité par la normalisation des langages de requête (SQL), l'autre sémantique.

Pour l'aspect sémantique, qui nous concerne principalement ici, la méthode couramment utilisée est connue sous le nom d'« intégration de schémas » (figure 1) ; il s'agit de définir un schéma conceptuel global, externe à toutes les bases concernées, et d'encapsuler (enrober, ou « to wrap ») chacune de celles-ci dans un système de règles de transformation entre le schéma global et son schéma local. Mais ces techniques ont leurs limites, car un tel schéma global s'avère en pratique très difficile à

développer et à maintenir ; il est remis en cause chaque fois qu'une base doit être ajoutée à l'ensemble ou que la structure de l'une doit être modifiée. Tout en posant ainsi d'importantes restrictions à l'autonomie d'évolution de chaque base, cette solution reste fragile. En outre, un tel schéma global suppose que toutes les bases de données concernées soient cohérentes à un changement près de vocabulaire ; cela devient irréaliste dès qu'on prétend l'étendre à l'ensemble illimité de toutes les bases potentiellement consultables (par exemple à travers Internet).

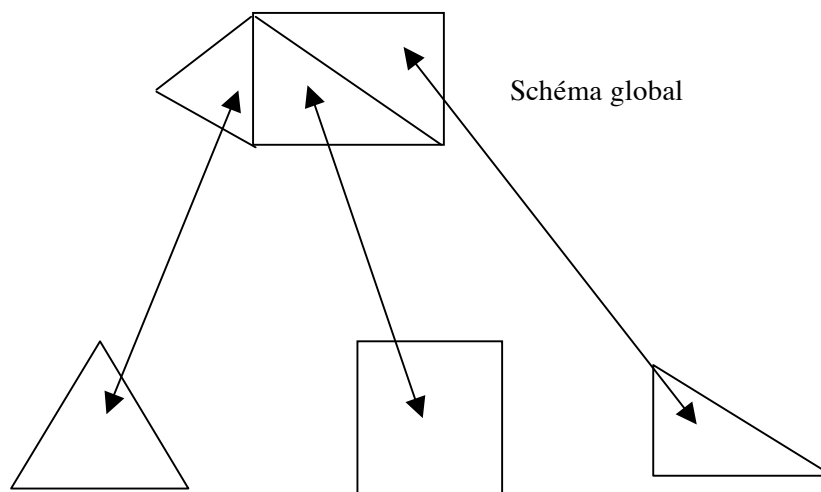


Figure 1 : l'intégration de schémas dans les bases de données et les règles de transformation entre schémas locaux et schéma global (flèches)

3.3 L'interopérabilité des applications informatiques

Face à l'interopérabilité des données, se pose la question de l'interopérabilité des programmes informatiques. Il existe plusieurs environnements permettant un fonctionnement distribué des programmes orientés objets en environnement hétérogène : CORBA, COM+, et (dans une moindre mesure, puisque limité au monde du langage Java) l'environnement RMI de Java. La principale fonction de ces environnements est d'offrir des services d'interconnexion selon le modèle client-serveur. Leur principal atout économique est qu'ils peuvent encapsuler les logiciels existants dans des emballages (wrappers) permettant leur interopérabilité sans qu'il y ait besoin de les réécrire : ces emballages créent en effet un niveau de description de ces logiciels indépendant de leur implémentation, qui reste cachée. Leur principale limitation est la même que pour toutes les approches purement syntaxiques, exigeant de l'utilisateur la connaissance détaillée des diverses classes d'objets abstraits qu'il veut mettre en œuvre.

Les trois cas qui précèdent (niveau technique, bases de données et logiciels « orientés objets ») constituent ce qu'on peut appeler les notions classiques d'interopérabilité. Avant d'introduire une succession de niveaux plus abstraits d'interopérabilité, considérons le cas des « composants métiers ».

Dans de nombreux domaines spécialisés (télécommunications, santé, etc.), les techniques d'interopérabilité des programmes ont commencé par donner lieu au développement de vastes bibliothèques de classes d'objets, les « composants métiers » (ou « business objects », si l'on veut absolument faire *in*). Mais on a aussi dépassé ce stade technique, qui serait celui de bibliothèques de composants logiciels, pour atteindre à un niveau où ces technologies prennent un sens sociétal beaucoup plus important. Car on en vient à disposer de *bibliothèques de modèles abstraits des objets et processus qui permettent de décrire les organisations* (entreprises, administrations, etc.) actives dans ces domaines (et ce en des termes qui n'ont plus rien d'informatique). Dans ces secteurs

d'activité, les systèmes d'information ne seraient dès lors plus conçus au cas par cas, ou par paramétrage de progiciels généraux, mais par assemblage de ces briques de base. Les enjeux sociétaux sont considérables : la normalisation des ontologies avec lesquelles nous comprenons ces domaines d'activité débouche en effet sur la normalisation de notre manière de les penser, une boucle de rétroaction pouvant en outre s'installer puisque les systèmes d'information qui sont mis en place dans les organisations en suivant cette démarche ne peuvent que renforcer en retour cette manière de penser les organisations (voir Berthier 2002, chapitre 10).

4. L'interopérabilité « cognitive » des agents artificiels entre eux

Dans la pratique de l'IA (comme dans celle des bases de données et des systèmes d'information), on se heurte très souvent à un problème conceptuel majeur, celui de la *discordance ontologique* : des ontologies régionales développées séparément, même si elles se rapportent au même domaine, ou à des domaines ayant une intersection non vide, (et même dans des domaines qui peuvent donner l'illusion d'être simples, comme ceux de l'identification des individus ou des relations de parenté), n'ont aucune raison *a priori* d'utiliser le même vocabulaire ou d'avoir structuré leurs concepts de manière identique – et cela pour de multiples raisons : langues différentes, cultures différentes, points de vue différents sur une même culture, approches différentes du problème, etc. (Alors que le mot « oncle » permet de ne pas faire de distinction entre oncle paternel et maternel, l'arabe, qui a deux mots différents, exige de la faire.) Toute la question, comme pour les schémas des bases de données, est donc de pouvoir exploiter simultanément des ontologies régionales différentes, c'est-à-dire de travailler dans plusieurs « repères conceptuels » à la fois, mais, contrairement à la technique d'intégration de schémas, on ne cherchera pas ici à définir un repère unique qui les subsumerait tous et qui, de fait imposerait une certaine normalisation du vocabulaire, donc de la manière de penser leur contenu.

4.1 Interopérabilité « cognitive »

Or, il résulte des développements effectifs de l'IA qu'au-dessus du niveau de base de l'interopérabilité technique des systèmes informatiques, on peut développer un autre niveau, que nous proposons de désigner comme celui de l'interopérabilité « cognitive » (sans oublier les guillemets pour ne pas entraîner d'interprétations abusives) des agents artificiels entre eux. C'est le niveau où les logiciels d'IA, qu'on considère dès lors comme des « agents intelligents », peuvent « partager » leurs connaissances entre eux, au sens précis de communiquer à leur propos et de les exploiter en commun.

Cela suppose que ces agents communicants aient une compréhension commune :

1°) de chacun de leurs actes de communication. A cet effet, ils doivent partager la même pragmatique de la communication : par exemple, chaque interlocuteur doit savoir sans ambiguïté si on lui communique des faits, des questions, des requêtes de collaboration, etc. ; la théorie des actes de parole (issue de la philosophie analytique) est, dans des versions plus ou moins formalisées (Austin 1956, Searle 1969, Searle & Vanderveken 1985), à l'origine de tous les systèmes existants (KQML, FIPA ACL) ;

2°) de leur(s) divers langage(s) d'expression de ces savoirs (de leurs systèmes de représentation des connaissances) ? A cet effet, un langage (c'est-à-dire une syntaxe et une sémantique générale) doit être défini en tant que référence commune, pouvant servir de pivot pour passer de l'un à l'autre (figure 2) ;

3°) des contenus communiqués. Contrairement au problème de la traduction entre langages naturels, celui de la traduction entre ontologies régionales n'est pas toujours simplifié par la recherche d'une ontologie commune – qui équivaldrait, dans le cas des bases de données, à la recherche d'un schéma conceptuel commun, dont nous avons vu les difficultés. En effet, contrairement aux structures grammaticales fixes des langages, les ontologies régionales sont, comme les schémas conceptuels des bases de données et plus que les lexiques des langues naturelles, appelées à se développer, se modifier, se multiplier. Une approche moins ambitieuse, mais plus générale, qui relève des techniques

d'interopérabilité, consiste à établir des correspondances partielles entre ontologies (figure 3), sans chercher à construire une ontologie globale ; à cet effet, les concepts particuliers au sujet desquels les agents communiquent doivent se référer à des « ontologies » communes (c'est-à-dire, en gros, à des vocabulaires communs, donnant le même sens à leurs différents termes), ou à des ontologies qui sont au moins reliées entre elles par des règles précises (spécifiant par exemple que le « nom » d'une « personne » dans une ontologie correspond au « patronyme » d'un « individu » dans une autre ontologie).

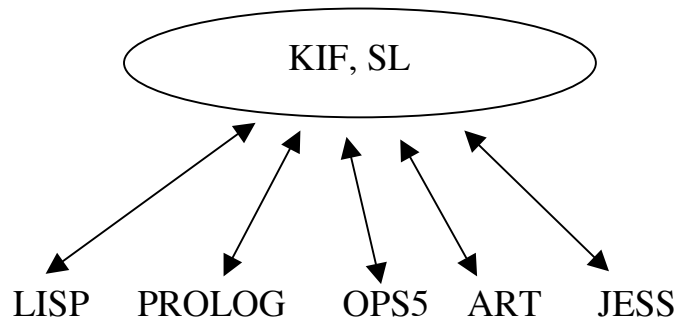


Figure 2 : le passage d'un langage de représentation des connaissances à un autre, via un langage pivot universel

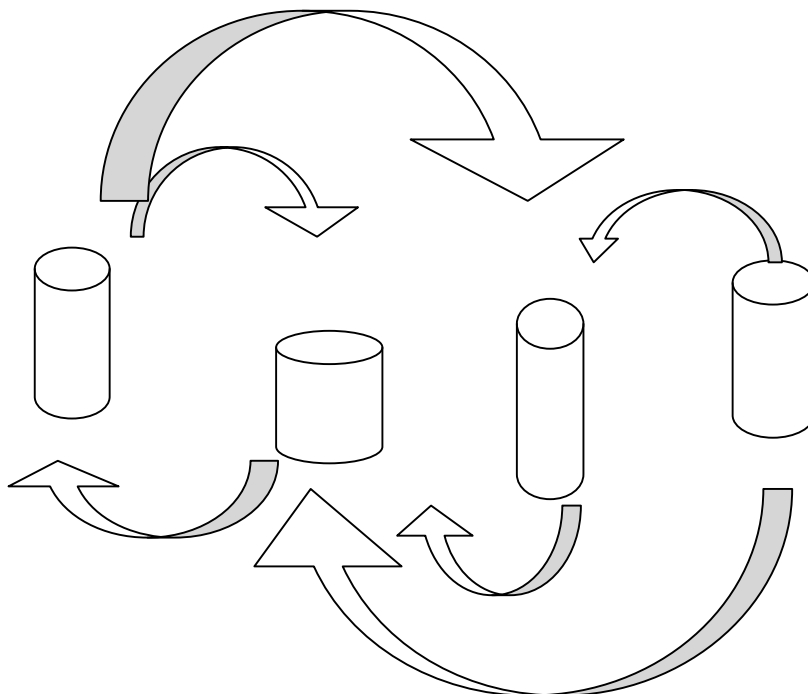


Figure 3 : correspondances partielles entre ontologies, à l'aide d'un langage uniforme de spécification des ontologies et de leurs correspondances (flèches) ; contrairement à l'intégration de schémas dans les bases de données, il n'y a pas ici d'ontologie globale.

Insistons : toutes ces technologies existent réellement et sont effectivement mises en œuvre. L'étude de ces trois capacités, que nous qualifions globalement d'interopérabilité « cognitive » entre agents artificiels, a fait l'objet d'un projet américain d'envergure, le Knowledge Sharing Effort (KSE), depuis le début des années 1990. A chacune des trois contraintes ci-dessus, KSE a associé un langage formel standardisé. Ces trois langages (respectivement KQML, KIF et ONTOLINGUA) et la philosophie générale de ce projet ont été repris et mis en pratique dans de nombreuses applications opérationnelles (pour un exemple d'application, voir la section 5).

Plus récemment, le problème a été repris au niveau d'instances de normalisation. L'association FIPA (Foundation for Intelligent Physical Agents) s'est donné pour objectif de standardiser la notion de plateforme multi-agents et les principes de communication entre agents. Ainsi, tous les agents de toutes les plateformes qui respecteront ces normes, aussi loin dans l'espace puissent-ils se trouver les uns des autres, devraient-ils pouvoir bénéficier de l'interopérabilité « cognitive », sous réserve que soient effectivement spécifiées les correspondances entre leurs diverses ontologies ; sur ce point, intelligence artificielle ou pas, il n'y a évidemment pas de miracle, et un travail important peut être exigé, en particulier quand on dispose d'une multitude de sources avec des ontologies hétérogènes – comme cela est souvent le cas, en sciences sociales, avec les observations de terrain. Le FIPA entend jouer le même rôle de normalisation en matière de systèmes multi-agents que l'OMG en matière de systèmes orientés objets ou que le W3C en matière de technologies Web. Depuis plusieurs années déjà, des plateformes conformes aux spécifications du FIPA sont disponibles et exploitées dans l'industrie (JADE, ZEUS, FIPA-OS, etc., pour ne citer que quelques exemples de logiciels « libres »).

4.2 Retour sur le concept d'interopérabilité ; extensions

Rappelons que le concept purement technique d'*interopérabilité* (litt. la capacité à opérer, ou à fonctionner, ensemble) tient aujourd'hui une place centrale dans les technologies de l'information et des réseaux. Notons déjà que ce terme donne à réfléchir : car l'interopérabilité n'est pas l'identité ; en pratique, c'est même quasiment le contraire, la question ne se posant que pour pallier le défaut d'identité (des machines, des systèmes d'exploitation, des langages de programmation, etc.). Il s'oppose donc à tous les discours identificatoires. Il véhicule cependant une plus grande exigence que le terme courant d'interface (lequel semble ne viser que la surface des choses).

Or, ce concept peut encore se généraliser considérablement. Grâce aux avancées réalisées dans le traitement informatique du langage naturel (terme consacré pour désigner les langages humains), on peut définir un type encore plus élevé d'interopérabilité, l'interopérabilité sémiotico-cognitive entre l'homme et les agents artificiels de l'IA ; de même, grâce aux technologies de la réalité virtuelle, on peut définir une interopérabilité sensori-motrice entre l'homme et l'ordinateur (Voir Berthier 2002 et 2004).

5. Applications concrètes : la question des ontologies et l'intégration d'informations hétérogènes sur Internet

Contrairement aux mythes savamment entretenus autour d'Internet, accéder à des informations ou connaissances *pertinentes* éparses sur le Web ne se limite pas à effectuer quelques clics de souris, car le véritable problème est de trouver les bons clics. Ensuite, assembler les informations ou connaissances recueillies à partir des diverses sources pour les exploiter conjointement pose de nouveaux types de difficultés : certaines techniques (différences de formats des données, etc.) ; d'autres conceptuelles, comme le problème des discordances ontologiques.

La question de la « lisibilité universelle » des documents mis sur le Web étant résolue par les techniques de base de l'Internet (comme le langage html et le protocole http), nous distinguons plusieurs niveaux techniques possibles d'accès effectif aux informations *pertinentes* sur le Web, et considérons les présupposés ontologiques correspondants :

- les *annuaires de recherche* (comme yahoo.com ou nomade.fr) repose sur : 1°) la construction « manuelle », par ses concepteurs, d'une *ontologie globale basée sur un classement hiérarchique*

rigide des différents domaines d'intérêt (supposés) des utilisateurs ; et 2°) l'association, elle aussi « manuelle » de sites appropriés à chaque catégorie terminale de l'arborescence ainsi définie. L'utilisateur descend progressivement dans cette arborescence et est *in fine* renvoyé à une liste de sites prédéterminés.

- Les *moteurs de recherche* (comme google.fr) permettent une recherche beaucoup plus fine, qui ne reste pas au niveau des sites, mais renvoie (en principe) les pages précises correspondant à une interrogation de l'utilisateur. Celle-ci doit être exprimée à l'aide d'un langage de requête, dont il doit donc connaître la syntaxe. La principale difficulté pour l'utilisateur est de deviner les mots (ou expressions) clés susceptibles de lui apporter les informations souhaitées. Pour pouvoir répondre efficacement aux requêtes des utilisateurs, les moteurs de recherche constituent au préalable une immense table croisant des mots clés et expressions avec toutes les pages accessibles du Réseau qui les contiennent. *L'ontologie implicite d'un moteur de recherche est définie par sa table de mots et expressions clés* ; elle est beaucoup plus souple que celle des annuaires et varie dynamiquement en fonction du contenu du Web ; elle manque pourtant encore trop de sémantique.

- L'utilisation dans un moteur des liens associatifs définis dans un réseau sémantique permet d'étendre automatiquement la requête de l'utilisateur avec des mots clés voisins de ceux qu'il a fournis. Mais la technologie d'indexation du Web, et donc aussi l'ontologie implicite, reste la même que dans les moteurs ordinaires.

Avec le niveau suivant, l'enjeu est de changer le statut de ces stocks d'informations *a priori* inertes que contient le Web. Car une possibilité d'accès adéquat est bien la première et la moindre des conditions pour qu'on puisse légitimement considérer ces milliards de « pages » comme *des savoirs pour l'homme*. Et justement, *la technologie des « agents d'information intelligents » annonce, de manière certes encore balbutiante, la possibilité pour l'homme et l'ordinateur de transformer conjointement les masses inimaginables d'informations éparpillées sur Internet (en des sens du terme « information » qui ne sont a priori pas les mêmes pour tous deux) en « savoirs » accessibles, compréhensibles et exploitables par tous deux*. Qu'il faille mettre en œuvre semblables techniques d'IA pour donner au Réseau sa pleine dimension de réservoir de savoirs démontre que *l'accès véritable à cette dimension d'Internet sera nécessairement simultané pour l'homme et pour l'ordinateur*.

En mettant en œuvre les technologies d'interopérabilité « cognitive », les systèmes d'agents d'information intelligents visent à aider les utilisateurs à surmonter leurs problèmes de surcharge informationnelle, en remplissant deux fonctions principales :

1°) par rapport à la recherche d'informations : ils cherchent les sources d'information pertinentes parmi toutes celles (généralement hétérogènes et distribuées dans l'espace) auxquelles ils ont accès ; à cette fin, ils raisonnent sur les requêtes de l'utilisateur, les décomposent, les traduisent, les affinent en fonction du contexte ;

2°) par rapport à l'exploitation des résultats : ils les filtrent, les analysent, les évaluent, les synthétisent, les présentent de manière lisible, etc.

En complément à ces techniques d'agents intelligents, et visant à faciliter leur développement, celles du Web sémantique visent à remplacer les champs « keywords » des pages html par des descriptions formalisées des ontologies sous jacentes. Les agents auraient ainsi un accès simplifié à la dimension sémantique, sans devoir passer par une analyse linguistique détaillée du texte de chaque page.

Même si la technologie est jeune, les applications existent bel et bien. Concernant le problème d'intégration de données en provenance de sources multiples hétérogènes, deux projets désormais bien connus, TSIMMIS et MOMIS (Bergamaschi & al. 1999), s'y attaquent, dans le cadre de données textuelles. Le mot « textuel » est pris ici au sens où il exclut les médias non textuels, mais pas au sens fort de textes quelconques en langage naturel ; les « textes » doivent être structurés en champs bien déterminés (comme par un schéma conceptuel dans une base de données, par des structures dans des langages à objets, ou encore par des balises précises dans une page Web en XML). Ces projets visent à

généraliser les travaux classiques sur la fusion de bases de données hétérogènes. Dans les deux projets, chaque source de données possède son propre langage de requête et son propre schéma conceptuel local (c'est-à-dire ses propres vocabulaire et modes de structuration des données) et l'on cherche à intégrer leurs données via un langage de requête commun dans un schéma conceptuel global.

6. *Intégration de schémas, interopérabilité et homologies structurales*

Au point où nous sommes arrivés, nous disposons de deux manières différentes de traiter des données d'origines multiples : l'intégration de schémas (figure 1), classique dans le monde des bases de données, et l'établissement de correspondances partielles entre ontologies (figure 3), basée sur les techniques d'interopérabilité. Cependant, il existe encore une autre méthode, beaucoup moins technique, qu'il nous semble utile de rappeler ici, car elle est plus connue en sciences sociales : c'est la méthode structuraliste, que nous aborderons sous la forme que lui a donnée Lévi-Strauss avec ses homologies structurales, dont nous proposons le schéma suivant (figure 4).

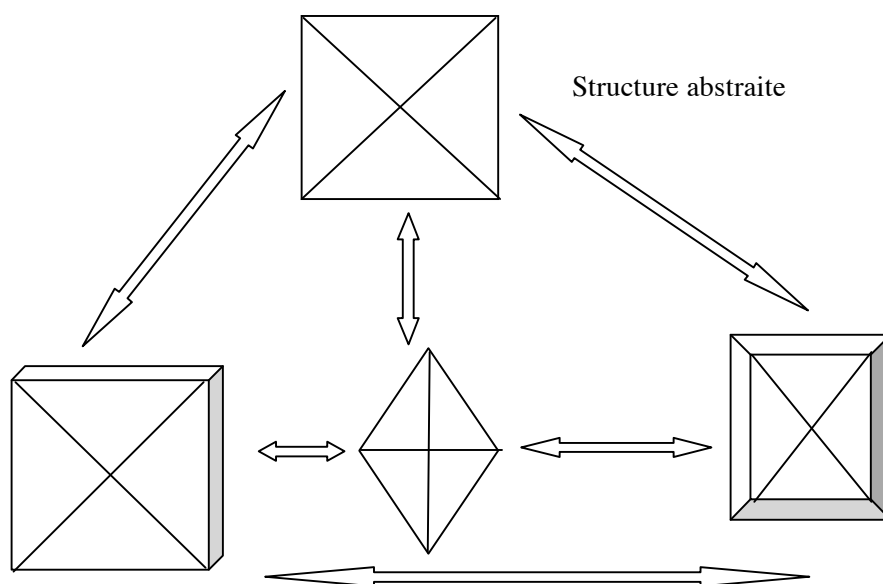


Figure 4 : différentes structures concrètes et la structure abstraite correspondant aux homologies structurales retenues par la modélisation (flèches)

La démarche de modélisation structuraliste (qui, à notre avis, imprègne encore toute la démarche scientifique actuelle, au moins dans la phase structurante initiale de définition de ses objets) consiste à simultanément (les deux aspects étant inséparables) : a) sélectionner les « différences significatives » qui vont être retenues dans chaque observation et b) établir les « homologies structurales » entre observations voisines (flèches horizontales de la figure 4) ; les homologies relient entre elles les « différences significatives » – qui sont elles-mêmes considérées comme significatives du fait qu'on trouve entre elles ces homologies. Elle aboutit donc à réunir sous une même structure abstraite (carré du haut de la figure 4) diverses observations, qui sont reliées entre elles par des « groupes de transformation ». Dans le cas simple des phonèmes, le carré du haut de la figure 4 représenterait un ensemble d'oppositions phonématiques, tandis que les trois dessins du bas représenteraient des

instances (des prononciations effectivement observées) de ces phonèmes. Dans le cas des mythes des origines, les dessins du bas représenteraient des mythes effectivement racontés, le carré du haut le mythe abstrait qui est considéré comme leur structure commune. Dans chaque cas, l'épaisseur « humaine » du son (toutes les caractéristiques vocales du locuteur) ou du récit (les détails « insignifiants » de l'histoire ou les répétitions d'un même épisode) est mise de côté pour faire apparaître ce qui est (considéré comme) significatif.

Bien que les techniques d'interopérabilité s'appliquent en principe à des sources de connaissance déjà formalisées, alors que la méthode structuraliste vise à constituer des modèles formels, cette dernière partage avec les techniques d'interopérabilité le souci d'établir des correspondances entre éléments « significatifs » issus de ses diverses sources de connaissance. Mais elle apporte aux techniques d'interopérabilité une restriction essentielle dans la mesure où elle vise, comme la méthode d'intégration de schémas dans les bases de données (et après avoir épuré ses données des éléments considérés comme secondaires), à développer un schéma global commun à toutes ses sources.

Mais, à vouloir pousser trop loin ce souci d'intégration des schémas locaux en schémas de plus en plus généraux, on en vient inévitablement à ce qui a été la tentation permanente de Lévi-Strauss, à savoir la réification de la structure. Comme le remarque Eco (1972, pp. 342-343) : « C'est à ce moment qu'un personnage a fait son entrée sur la scène de la réflexion structurale [...] : l'esprit humain. Le fait d'étudier un mythe signifie alors identifier un système de transformations d'un mythe à l'autre qui démontre comment en chacun d'eux certains chemins fondamentaux de la pensée étaient parcourus de nouveau, que les constructeurs des mythes le sachent ou pas. Quelle que soit la chose que les mythes prétendent raconter, ils ne faisaient que répéter – et ils répètent – la même histoire. Cette histoire est la narration des lois de l'esprit sur lesquelles ils se basent. Ce n'est pas l'homme qui pense les mythes, *les mythes se pensent dans les hommes*, ou mieux, ainsi que Lévi-Strauss le souligne, in *le Cru et le Cuit*, dans le jeu des transformations réciproques possibles, *les mythes se pensent entre eux* ».

Or, il nous semble que les idées issues des problèmes d'interopérabilité permettent de dépasser ce dilemme. Considérons un ensemble d'observations relatives à un même objet (mythes, phonèmes, totems, etc.). Plutôt que de rechercher systématiquement des « homologues structurales » globales et une structure abstraite globale qui se retrouve (modulo des « transformations » adéquates) dans toutes les observations disponibles, une généralisation de la méthode structuraliste, moins radicale, consisterait à se contenter, comme dans la figure 3, d'un ensemble d'« homologues structurales locales », c'est-à-dire partielles, qui établissent un système de correspondances entre diverses parties des observations disponibles.

7. Conclusion

Nous avons indiqué que la question des ontologies est partagée par l'informatique et les sciences sociales. Nous avons montré comment le problème des discordances ontologiques peut être abordé, par les techniques d'interopérabilité, d'une manière plus subtile que par la recherche d'une ontologie globale qui réunirait toutes les ontologies particulières ; par exemple, nous avons indiqué comment ces techniques sont déjà appliquées à la recherche et à la synthèse d'informations sur Internet. Enfin, nous avons comparé trois schèmes de synthèse d'information et montré comment le schème structuraliste n'est qu'une forme extrême d'un schème plus général basé sur l'idée d'« homologues structurales locales ». Notons pour conclure que les problèmes de discordances ontologique se posent chaque fois qu'on veut synthétiser des données éparses en provenance de sources indépendantes, cas fréquent en sciences sociales, et que les techniques d'interopérabilité, adaptées à chaque cas particulier, pourraient aider à les résoudre sans obliger à modifier les sources elles-mêmes.

Bibliographie

Berthier Denis, *Le savoir et l'ordinateur*, L'Harmattan, coll. « Impacts des Nouvelles Technologies », Paris, 2002.

Berthier Denis, La culture et la rationalité modernes, du structuralisme à l'IA symbolique, *Automates Intelligents*, n° 42, Paris, Juin 2003.

Berthier Denis, *Méditations sur le réel et le virtuel*, L'Harmattan, Paris, 2004.

Berthier Denis, Penser notre relation à la Machine (plutôt que nous penser comme des Machines), *Terminal*, à paraître.

Bergamaschi Sonia & Beneventano Domenico, Integration of Information from Multiple Sources of Textual Data, in (Klusch 1999).

Breton Philippe, *L'utopie de la communication*, La Découverte, Paris, 1992 ; 2^{ème} éd. 1997.

Eco Umberto : *La structure absente*, éd. française : Mercure de France, Paris, 1972.

Klusch Matthias, ed., *Intelligent Information Agents*, Springer, 1999.

Lévi-Strauss Claude, *Anthropologie structurale*, Plon, Paris, 1958.

Lévi-Strauss Claude, *Le cru et le cuit*, Plon, Paris, 1964.

Newell Allen, Physical Symbol Systems, *Cognitive Science*, Vol 4, pp. 135-183, 1980.

Sowa John, *Conceptual Structures: Information Processing in Mind and Machine*, Addison Wesley Pub., 1984.