

UN PROGRAMME DE RECHERCHE POUR L'ÉCONOMIE DE L'INFORMATION¹

Prof. Pierre Lévy, CRC, FRSC, Université d'Ottawa
plevy@uottawa.ca

1^{er} décembre 2006

¹ NDLR : il s'agit ici de la suite de l'article publié dans le numéro 27

Les Symétries d'IEML

Nature et fonctions des symétries

La Californie se trouve à l'Ouest de l'Amérique *et* à l'Est de l'Asie. La correspondance et l'opposition des positions dans l'espace (la gauche et la droite, le haut et le bas, l'Est et l'Ouest) ou dans la communication (le *je* et le *tu*) forment sans doute la base de notre intuition de la symétrie. Sur un plan opératoire, la distinction effective *et* l'identité virtuelle des éléments symétriques impliquent une possibilité de *substitution réciproque*.

Les symétries d'IEML ont deux fonctions interdépendantes: refléter la symétrie générale des points de vue cognitifs et faciliter les traitements d'information automatiques.

Premièrement, ces symétries organisent l'égalité, la réciprocité et l'équilibre entre les différents points de vue cognitifs possibles. Or il existe - par définition - une variété indéfinie de tels points de vue. C'est pourquoi les quelques *symétries fondamentales* d'IEML peuvent être combinées à volonté selon plusieurs niveaux d'articulation pour produire des *symétries composées* de plus en plus riches et complexes, capables d'accommoder la variété des points de vue.

Deuxièmement, les symétries d'IEML peuvent être utilisées par des logiciels à des fins de rangement et de manipulation des adresses sémantiques. On peut notamment élaborer des algorithmes capables de reconnaissance automatique des *motifs symboliques* structurés par les axes de symétrie propres à IEML. Loin de remplacer les anciennes, ces nouvelles méthodes constituent une couche « méta » qui *s'ajoute* aux méthodes de traitement de l'information déjà disponibles, tout en multipliant leurs effets. Je rappelle que les calculs autorisés par la structure symétrique du métalangage n'ont évidemment de pertinence que dans la mesure où la symétrie entre signifiants correspond à une symétrie conventionnelle entre signifiés.

Symétrie des chemins

Toutes les adresses sémantiques d'IEML ont la forme de *chemins* décrits par les

coordonnées de trois stations (Source → Destination / Traductrice). Les stations de départ, d'arrivée et de passage peuvent *échanger leurs rôles*, adressant ainsi des chemins symétriques.

Selon la symétrie principale, à deux faces, de la source et de la destination, le chemin $[X \rightarrow Y]$ est symétrique du chemin $[Y \rightarrow X]$.

Selon la symétrie secondaire, à trois faces, de la source, de la destination *et* de la traductrice $[X \rightarrow Y / Z]$ est symétrique de $[Y \rightarrow X / Z]$, $[X \rightarrow Z / Y]$, $[Z \rightarrow X / Y]$, $[Y \rightarrow Z / X]$ et $[Z \rightarrow Y / X]$.

A l'échelon des adresses fixes, ou concepts, chaque *station* d'un niveau donné (événements, relations, idées, phrases) est constituée par un *chemin* de niveau immédiatement inférieur, et ainsi de suite jusqu'aux stations indécomposables que sont les éléments. Quant aux graphes conceptuels, ce sont par définition des flux d'information entre concepts, décomposables en chemins identifiés par trois stations (source, destination, traductrice). L'ensemble de la structure symbolique du métalangage présente donc un schéma complexe de symétries fractales.

La symétrie des chemins articule l'ensemble du métalangage. Grâce au fait que chaque adresse sémantique - quelque soit son niveau d'articulation - comprend de manière régulière une adresse de source, une adresse de destination et une adresse de traductrice, cette symétrie est mise au service des opérations automatiques de sélection et de manipulation des expressions du métalangage.

Symétries des stations

Symétrie générale

Les stations d'un même niveau existent indépendamment des flux d'information qui les relient et aucune - *a priori* - n'a de statut privilégié, central ou périphérique. En ce sens, toutes les stations d'un niveau d'articulation donné sont symétriques entre elles, jusques et y compris le niveau primitif des éléments. A un niveau d'articulation donné, aucune station ne

constitue *a priori* un choix de *source* supérieur ou inférieur à n'importe quel autre et, une fois déterminée la station source, aucun choix de station *destination* n'est *a priori* plus probable qu'un autre. Cette symétrie générale s'accompagne de symétries particulières découlant de la structure des éléments, que l'on va maintenant examiner.

Symétrie cognitive

Les éléments, on l'a dit plus haut, sont regroupés en deux ensembles, ou *pôles* pragmatique et sémantique.

Le pôle *pragmatique* O comprend le virtuel U et l'actuel A, qui se répondent en une dialectique de l'action [O = (U, A)], tandis que le pôle *sémantique* M comprend le signe S (signifiant), l'être B (interprétant) et la chose T (référént) qui composent ensemble une dialectique de la représentation [M = (S, B, T)].

Ce n'est pas ici le lieu d'expliquer ou de justifier sur un plan philosophique et scientifique la polarité cognitive entre action et représentation, processus et entités, verbes et noms. Je veux souligner, en revanche, que cette symétrie se retrouve à tous les niveaux d'articulation d'IEML. La polarité d'un flux dépend en effet de la polarité de sa source : si la source est pragmatique le flux est pragmatique et si la source est sémantique, le flux est sémantique. Les stations de destination et de traduction qui composent ce flux au niveau d'articulation inférieur sont à leur tour pragmatique *ou* sémantique, et ainsi de suite.

Cette distribution polaire en cascade commande les *catégories grammaticales* des concepts IEML. Selon leur polarité, les *événements* sont des verbes (pragmatiques) ou des noms (sémantiques). Les catégories de *relations* sont déterminées à leur tour par les combinaisons de pôles qui les engendrent et il en est de même pour les catégories *d'idées* et de *phrases*. A titre d'exemple, les phrases IEML sont de type performatif (composées d'un verbe, d'un objet et d'un médium) ou propositionnel (composées d'un sujet, d'un attribut et d'une condition) selon que leur première idée est sémantique ou pragmatique².

La symétrie polaire fractale peut donc être exploitée pour la sélection d'adresses

² Pour plus de détails sur la grammaire d'IEML, voir, sur le site www.ieml.org... ainsi que l'article IEML, finalités et structure fondamentale...

sémantiques en fonction de leurs catégories grammaticales, et cela à tous les niveaux d'articulation. Comme dans le cas de la symétrie entre source et destination, la régularité de la symétrie entre pôles pragmatique et sémantique peut aussi être exploitée par des algorithmes d'analyse, de calculs de similarité, de rangement, etc.

Symétries complexes

Les pôles pragmatiques et sémantiques sont eux-mêmes articulés par des symétries internes.

Au niveau des cinq *éléments*, la symétrie interne au pôle pragmatique est binaire (virtuel U / actuel A) alors que la symétrie interne au pôle sémantique est ternaire (signe S / être B / chose T).

Au niveau des vingt-cinq *événements*, ces symétries binaires et ternaires sont combinées en symétries plus complexes...

- une symétrie à quatre faces : 2 x 2 (U→U, U→A, A →U, A→A),
- deux symétries à six faces : 2 x 3 (U→S, U→B, U →T, A→S, A→B, A→T et S→U, S→A, B→U, B→A, T→U, T→A),
- une symétrie à neuf faces : 3 x 3 (S→S, S→B, S→T, B→S, B→B, B→T, T→S, T→B, T→T).

Ces symétries plus complexes peuvent être observées sur le diagramme de l'alphabet IEML ci-dessus.

Au niveau des six cent vingt-cinq *relations*, on obtient des symétries à seize (4 x 4), vingt-quatre (4 x 6), trente-six (6 x 6 et 4 x 9), 54 (6 x 9) et quatre-vingt une (9 x 9) faces. Une grande partie de la difficulté de la création d'IEML a consisté à mettre en rapport les symétries formelles entre *signifiants* du langage avec des symétries parallèles entre *signifiés* (qui sont exprimés en langue naturelle), tout en fournissant un point d'appui pratique à la création d'idées et de phrases capables d'adresser un champ de signification *ouvert*. Pour une illustration de ces symétries à 16, 24, 26, 54 et 81 faces, on peut se reporter aux matrices de relations et d'idées du dictionnaire IEML en ligne (www.ieml.org).

Les *idées* interprétées en langues naturelles qui figurent en 2006 dans le dictionnaire IEML ont été créées en respectant autant que possible l'armature logico-conceptuelle proposée par les symétries complexes des éléments, événements et

relations. Il devrait en être de même pour les futures entrées du dictionnaire.

Comme dans le cas des autres symétries, les symétries complexes se reflètent dans les adresses sémantiques et peuvent être utilisées pour des opérations automatisées de sélection, d'analyse, de synthèse et de rangement.

Substantialité

Les cinq éléments d'IEML: virtuel, actuel, signe, être et chose sont placés sur une échelle qui va du moins substantiel au plus substantiel, ou du plus subtil au moins subtil.

Il faut d'abord noter que les processus ou les verbes (pôle pragmatique O) sont par définition moins « substantiels » que les entités ou les noms (pôle sémantique M). Au sein même du pôle pragmatique, le virtuel est évidemment moins substantiel que l'actuel. Dans le pôle sémantique, le signe est moins substantiel que la chose et l'être se trouve dans une position intermédiaire.

La substantialité constitue en quelque sorte une « masse cognitive » associée aux éléments : 1 pour le virtuel, 2 pour l'actuel, 3 pour le signe, 4 pour l'être et 5 pour la chose. Par convention, la substantialité est additive, ce qui signifie que la substantialité d'un flux ou chemin d'information est égale à la somme de la substantialité des stations qui décrivent ce chemin. De ce fait, il devient possible de ranger automatiquement les adresses de même niveau d'articulation par leur substantialité, de déterminer des ensembles d'adresses de même substantialité ou de mettre au point des requêtes complexes combinant le critère de substantialité avec d'autres critères. Le caractère additif de la substantialité implique que, à traductrice identique, deux concepts symétriques quant à leur source et leur destination ont nécessairement la même substantialité.

J'avais conclu le chapitre précédent en constatant que la structure d'IEML lui donnait la capacité virtuelle d'adresser un espace sémantique infini. Je conclus ce chapitre en constatant qu'il dispose en outre de traits formels (et notamment de symétries) qui donnent prise à des *calculs automatiques portant sur la signification* de ses expressions,

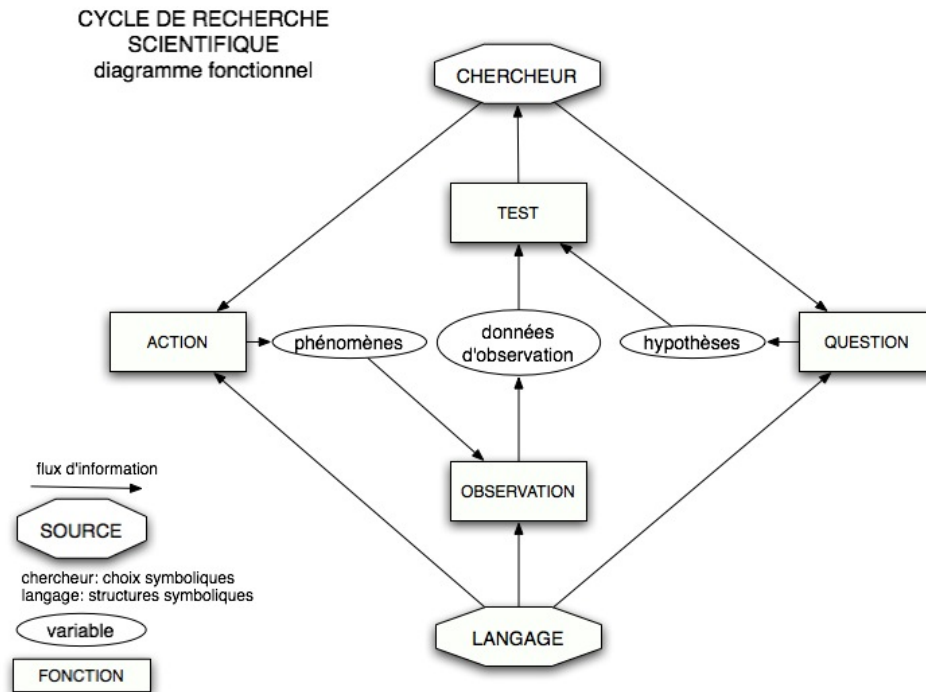
et cela de manière beaucoup plus régulière que les langues naturelles dans lesquelles sont rédigées les métalangages et ontologies actuellement disponibles. Les contraintes fonctionnelles et opérationnelles du problème général d'un métalangage de l'économie de l'information sont donc respectées par IEML. Il me reste maintenant à monter comment IEML respecte la contrainte épistémologique du problème du métalangage, c'est-à-dire, je le rappelle, la possibilité de formaliser et de tester des hypothèses scientifiques sur la mémoire numérique.

4) Les graphes conceptuels

Vers une recherche d'information scientifique

La recherche scientifique n'est pas moins libre, dans sa quête d'information, que celle du grand public. Ici comme là, les informations extraites de la mémoire collective ne prennent sens qu'en fonction d'une orientation intellectuelle et pratique singulière. Ici comme là, l'intention de recherche peut se tourner dans n'importe quelle direction. Ce ne sont pas ses objets qui distinguent *a priori* la recherche scientifique, mais ses exigences de transparence, de précision et de cohérence plus élevées. Elle ne creuse sa différence avec les autres types de recherche que par la complexité de son articulation symbolique : un raffinement qui vise à maximiser l'information extraite de la mémoire disponible.

Dans le schéma (ci-dessous) décrivant le cycle de la recherche scientifique, j'ai conservé exactement la structure du cycle de recherche d'information ordinaire. Les deux sources fondamentales d'information restent la communauté des chercheurs et les structures symboliques articulées par cette communauté. La différence entre les deux schémas vient de ce que les quatre grandes *fonctions* de la recherche d'information (production de documents, indexation des documents, formulation de requêtes et extraction d'information) et les trois *variables intermédiaires* (documents, métadonnées sur le document et métadonnées du chercheur) ont été rebaptisés en termes épistémologiques.



1) La création de documents est sophistiquée en protocole expérimental générateur de phénomènes - *l'action* du diagramme ci-dessous - dont les paramètres sont autant que possible contrôlés et reproductibles.

2) L'indexation des documents (qui sont les *phénomènes* de la mémoire numérique) se raffine en un processus *d'observation* et de description aboutissant à des métadonnées d'observation capables de confirmer provisoirement ou d'infirmer les hypothèses.

3) La requête se perfectionne en *questionnement* réflexif devant aboutir à la production *d'hypothèses*.

4) L'extraction d'information est maximisée par des *tests* qui mettent en regard les hypothèses et les données d'observation.

Il est clair qu'un processus de recherche scientifique a tout intérêt à adopter une articulation symbolique explicite et cohérente pour assurer la *coordination* des quatre fonctions qui viennent d'être énumérées. Si, au cours d'un cycle de recherche, (1) les protocoles de création ou de sélection de documents, (2) la description des documents produits et (3) les hypothèses portant sur ces descriptions sont formulés à l'aide de concepts disparates et incompatibles, on voit mal comment (4) les tests pourraient fournir une information pertinente. Le langage scientifique a donc notamment pour fonction

d'assurer l'unité opératoire de la recherche, de « boucler » correctement chaque cycle.

Mais la recherche scientifique ne se limite pas à un seul cycle de recherche. C'est un apprentissage ouvert, une exploration libre impliquant une succession de cycles. Au cours de cette succession temporelle, le langage utilisé pour formuler les hypothèses et décrire les données peut évidemment *évoluer*. De ce point de vue, un *méta-langage* scientifique capable d'explicitier et de mesurer les différences entre les versions successives du langage utilisé dans la recherche serait particulièrement utile. En permettant la mise en relation rigoureuse des différences successives d'articulation symbolique avec la quantité et la qualité de l'information extraite des tests, le métalangage - pièce essentielle du moteur de recherche sémantique - pourrait jouer un rôle important *d'aide à l'apprentissage*.

On a vu plus haut que le langage scientifique utilisé par un chercheur ou une équipe de chercheurs articule l'observation et le questionnement : l'équivalent - en plus sophistiqué - d'une ontologie ou d'un langage documentaire. Un tel langage scientifique reflète généralement l'appartenance disciplinaire et les choix théoriques des chercheurs et ne saurait être limité que par une contrainte minimale d'explicitation et de cohérence.

Aujourd'hui, cette liberté dans le choix des concepts est affectée d'une contrepartie négative : la fragmentation disciplinaire et la divergence des paradigmes dans les sciences humaines et sociales rendent difficile la collaboration transversale et l'accumulation raisonnée des connaissances. Un métalangage capable de traduire les langages scientifiques particuliers permettrait de conserver - voire même d'augmenter - la liberté des choix théoriques tout en assurant une collaboration

IEML se propose donc comme une sorte de métalangage commun pour d'explicitation des hypothèses et la description des données dans la communauté des chercheurs, tout particulièrement dans les sciences de l'homme et de la société. Il pose en principe l'équivalence *a priori* de toutes les hypothèses et de toutes les grilles d'interprétation des données choisies par les chercheurs, puisqu'elles s'exposent par construction à un *test* sur tout ou partie de la mémoire numérique.

Je vais maintenant exposer de quelle manière les graphes IEML peuvent traduire des hypothèses scientifiques et comment ces hypothèses peuvent être testées sur des données elles-mêmes traduites en graphes IEML.

Modèles, théories et données

Pour commencer, il me faut distinguer entre deux types d'hypothèses : les modèles et les théories.

Un modèle définit des variables entre lesquelles il établit *des types de relations*. Je distingue trois types de relations entre variables : relation d'interdépendance, relations d'appartenance et relations d'ordre. IEML traduit ces trois modèles de base par trois genres de graphes conceptuels: les *claviers* (ou matrices) pour les variables interdépendantes, les *arbres* pour les rapports d'appartenance entre variables et les *séries* pour les relations d'ordre. A partir de modèles de base appartenant à ces trois genres, il est possible de construire des modèles mixtes, plus complexes.

Les modèles constituent des *grilles d'interprétation* sur lesquelles sont projetées les données d'observation (ou descriptions des documents). A proprement parler, un modèle n'est ni vrai ni faux. Un modèle peut seulement être *plus pertinent* qu'un autre, dans la mesure où il permet de « donner plus de

intellectuelle plus efficace qu'aujourd'hui au sein de la communauté des chercheurs. Les données produites par les uns pourraient être plus facilement utilisées par les autres. Les hypothèses en présence pourraient faire l'objet de comparaisons plus fines et mieux éclairées. En somme, une ouverture et une transparence de l'économie de l'information ne bénéficierait pas moins à la communauté des chercheurs travaillant sur la mémoire collective numérisée qu'au grand public.

sens » aux données d'observation. En termes opératoires, je dirais qu'un modèle est d'autant plus pertinent qu'il suggère plus clairement des relations régulières entre les variables.

Dans le cadre d'un modèle donné, une *théorie* énonce (ou prévoit) une certaine *règle de relations* entre variables. Si cette règle est validée par les données d'observation au cours d'un test, la théorie est provisoirement vraie. Si la règle est invalidée par les données d'observation, la théorie est fautive.

Il est toujours possible que d'identiques données d'observations valident des théories différentes, éventuellement basées sur des modèles différents. Ni l'épistémologie, ni l'expérience de la communauté scientifique rapportée par l'histoire des sciences ne s'opposent à ce que plusieurs théories différentes sur le même objet soient simultanément confirmées par des tests empiriques.

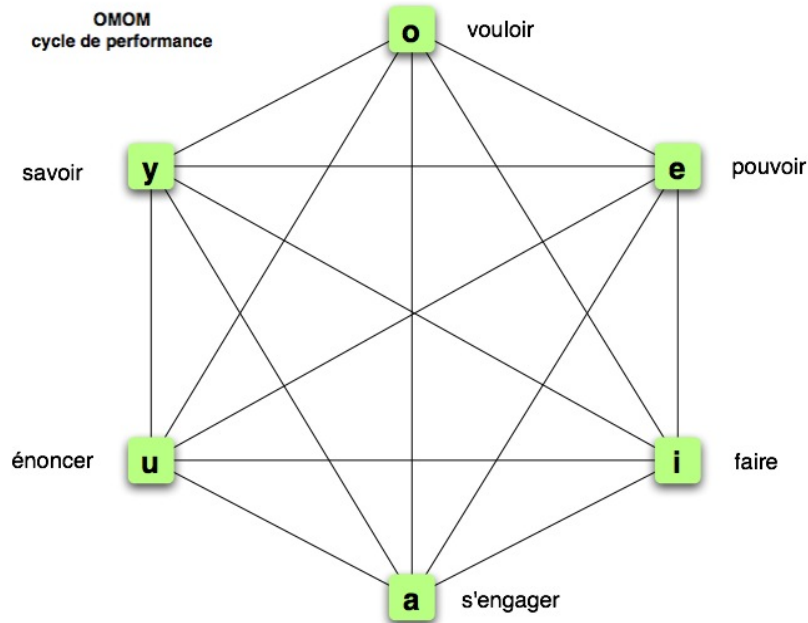
On considère traditionnellement que les tests qui offrent le plus d'information à la communauté des chercheurs sont les « découvertes », c'est-à-dire les premières validations de nouvelles théories (surtout si le modèle sous-jacent est original) ou les premières invalidations de théories antérieurement acceptées. Au fur et à mesure que les découvertes sont confirmées par des tests successifs, elles apportent de moins en moins d'information.

Les *données d'observation* sur lesquelles sont testées les hypothèses ne sont autres que les graphes conceptuels indexant les corpus de documents sélectionnés. Lorsque l'on parle des données, il faut bien distinguer les *documents*, correspondant pour l'épistémologie aux phénomènes « bruts » affectant les sens, et les *métadonnées* produites par l'indexation, qui correspondent au travail de *mesure* et de mise en forme des phénomènes permettant de tester les hypothèses (observation, description).

Les claviers : relations d'interdépendance

Les claviers sont des graphes IEMML qui représentent un flux d'information complexe d'un *ensemble* de concepts vers un

autre. Le nom « clavier » vient de ce que ces graphes peuvent être utilisés comme des claviers virtuels pour l'écriture d'autres graphes, et notamment de séries.



OMOM cycle de performance					
yy maîtriser un domaine de csce	yo établir des priorités	ye enseigner	yu expliquer	ya voter	yi innover
oy interroger	oo viser un objectif	oe diriger	ou analyser	oa convenir	oi choisir
ey apprendre	eo discerner	ee exercer une compétence	eu parler	ea posséder	ei emprunter
uy expliciter	uo synthétiser	ue prouver	uu communiquer	ua promettre	ui donner
ay mener une recherche	ao juger	ae créer	au délibérer	aa contracter	ai demander
iy computer	io conseiller	ie prêter	iu prendre	ia obtenir	ii produire

Dans l'exemple ci-dessus, les 36 flux d'information connectent six événements y, o, e, u, a et i *pris comme sources* aux mêmes six événements *pris comme destinations*. La représentation en graphe (avant) et la représentation en clavier (après) sont équivalentes. Chaque « touche » ou *clé* sémantique du clavier est un concept représentant un flux d'information (la première lettre représente la source et la seconde la destination).

Visuellement, un clavier IEML se présente comme une matrice de concepts, les rangées et les colonnes ayant respectivement des propriétés communes (sur l'exemple proposé, toutes les rangées ont la même source et toutes les colonnes ont la même destination).

Distances sémantiques entre patterns conceptuels

D'un point de vue géométrique, un clavier IEML représente un ensemble coordonné de *dimensions* (avec autant de dimensions que de concepts) définissant un sous-espace conceptuel particulier. Les données d'observation IEML sélectionnées et projetées sur ce sous-espace permettent

d'attribuer une *valeur* à chacune des dimensions - donc à chacun des concepts-clés - en fonction du nombre d'occurrence de ces concepts dans les données.

Dès lors, le test d'un modèle en forme de clavier sur un ensemble de données donne comme résultat *un vecteur multidimensionnel* ou « point » d'un sous-espace conceptuel. Puisque ce vecteur correspond à une certaine distribution de valeurs - un *pattern* - sur un clavier de variables conceptuelles, on peut appeler ce vecteur un *pattern conceptuel*. Le clavier se comporte donc, du point de vue de l'observation, comme une rétine conceptuelle.

Une branche particulière des mathématiques, l'algèbre linéaire, permet de calculer géométriquement des *distances* entre patterns d'un même espace conceptuel. Une fois les distances sémantiques calculées géométriquement, on peut leur appliquer divers programmes informatiques dans le but de *catégoriser* les patterns conceptuels et de repérer des *transformations régulières* dans leurs dynamiques temporelles. En intelligence artificielle, ce type de procédé a été effectivement utilisé dans la simulation de réseaux neuronaux et s'est montré

particulièrement efficace dans la reconnaissance de forme et les tâches de perception en général.

Il faut souligner que l'espace sémantique *en général* est absolument ouvert. Autrement dit : son nombre de dimensions est *a priori* illimité. En revanche, le sous-espace défini par un clavier conceptuel possède un nombre de dimensions limité. Il offre une sorte de fenêtre cognitive explicitement définie sur l'ouverture indéfinie de l'espace sémantique. A partir de cette matrice interprétative hypothétique, il est possible de repérer des patterns conceptuels, de mesurer des distances entre ces patterns, de tracer des dynamiques et de reconnaître des formes de dynamiques (fonctions sémantiques).

Indépendance et interdépendance

Lorsque l'on veut que les variables conceptuelles représentent les différentes *dimensions* dans lesquelles se déploie le phénomène décrit par les données d'observation, il est recommandé d'exprimer les hypothèses par des claviers. Dans ce cas, les variables sont dites *indépendantes* - au sens mathématique du terme - dans la mesure où l'on a besoin de la valeur de *toutes* les dimensions pour adresser un pattern conceptuel. Mais sur le plan de l'interprétation causale, en revanche, on peut parler de variables *interdépendantes* puisqu'elles représentent des dimensions coordonnées du *même espace*. Pour illustrer ce point, je vais d'abord évoquer la situation dans l'espace physique puis dresser un parallèle avec l'espace conceptuel.

Dans l'univers physique ordinaire (pourvu d'un système de coordonnées à quatre dimensions et d'une origine), si l'on veut déterminer la position d'un point dans l'espace-temps, on a besoin *a priori* de connaître les valeurs de x, de y, de z *et* de t. Les variables sont donc indépendantes. Mais dès que l'on repère la *trajectoire* d'un point ou les *rotations* d'une *figure*, les valeurs des variables se transforment de manière interdépendante : elles sont reliées par des fonctions.

Transportons-nous maintenant dans l'espace conceptuel. Dans l'exemple présenté plus haut, le clavier propose un système de coordonnées pour l'adressage des types d'actes. La valeur d'une clé est *a priori* indépendante de la valeur des autres clés et

l'adressage d'un pattern conceptuel demande la connaissance de la valeur de toutes les clés. On ne peut pas déduire la valeur d'une clé de la valeur des autres. Les *relations* précises *entre* variables conceptuelles sont fournies par les données d'observation : elles n'appartiennent pas au modèle. En revanche, le clavier explicite une hypothèse générale d'interdépendance ou de coordination entre les différents types d'actes représentés par les variables. Les clés du clavier ne prennent sens que les unes par rapport aux autres. Le clavier décrit donc un espace interdépendant dans lequel des trajectoires de patterns conceptuels ou des rotations de figures conceptuelles (rapports déterminés entre patterns) peuvent être décrites par des fonctions.

Claviers : modèles et théories

En tant que *modèle*, le clavier donné en exemple postule l'existence d'un espace des performances dont chaque variable particulière représente une dimension. Il prétend faciliter le repérage de *patterns* performatifs. Tester l'hypothèse revient à projeter sur ce clavier les métadonnées IEML de documents décrivant le déroulement d'actions complexes. On pourra alors, par exemple, catégoriser les documents en fonctions des proximités de leurs patterns de performance, étudier l'évolution de ces patterns au cours du temps, en inférer des régularités, ordonner les documents en fonction de leur valeur dans telle ou telle dimension du sous-espace conceptuel que l'on jugera prioritaire ou aider l'action et sa planification en fonction du repérage de forces ou de faiblesses dans la distribution des valeurs.

Si l'on utilise un clavier non plus comme modèle mais comme *théorie* (impliquant donc des règles de relations entre variables) la formulation de l'hypothèse consiste alors à expliciter un pattern conceptuel particulier ou une relation entre patterns. Le test consistera à vérifier que les relations entre patterns se trouvent bien dans un ensemble déterminé de données d'observation, ou à sélectionner les données qui valident l'hypothèse (et donc à trouver dans quels cas l'hypothèse est vérifiée).

En 2006, IEML propose déjà une trentaine de claviers conceptuels dans les domaines les plus variés (connaissances organisées, compétences, finalités morales, fonctions sémiotiques, rôles sociaux, fonctions

techniques...) et il est possible d'en construire autant que l'on voudra³.

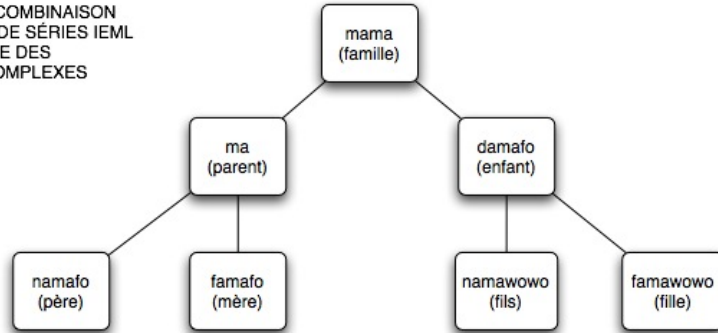
Les arbres : relations d'appartenance

Représentation de fonctions complexes

Les graphes conceptuels en forme *d'arbres* représentent des hiérarchies d'ensembles, de sous-ensembles et d'éléments. Dans cette approche, la « racine » première de l'arbre (le nom de l'ensemble) représente une *fonction* complexe, la hiérarchie descendante des noms de sous-ensembles représente des *sous-fonctions* et les « feuilles » terminales de l'arbre - c'est-à-dire les éléments des sous-ensembles - représentent les noms de *variables* conceptuelles.

³ Voir notamment les tableaux de relations et d'idées sur www.ieml.org. Je n'entre pas ici dans le détail des distinctions entre différents types de claviers. Le *CI lab. technical report n°2* (à la rubrique « Journal » du site www.ieml.org) précise les contraintes auxquelles doivent obéir les claviers *réguliers* et rien n'empêche de construire des claviers *irréguliers*.

EXEMPLE DE COMBINAISON
D'ARBRES ET DE SÉRIES IEML
POUR DÉCRIRE DES
RELATIONS COMPLEXES



Phrases performatives (verbe objet medium)

wawe damafo ma (les parents engendrent les enfants),
wuwe ma damafo (les enfants aiment les parents),
oeweh damafo ma (les parents éduquent les enfants).

Phrases propositionnelles (sujet attribut condition)

namawowo kamafo famawowo (le fils est le frère de la fille),
famawowo kamafo namawowo (la fille est la soeur du frère)
ak tutudu mama (les conflits sont nombreux dans la famille).

Dans l'exemple ci-dessus, la famille (nucléaire) est représentée comme une fonction générale dans laquelle les deux sous-fonctions principales sont occupées par les parents et les enfants. La fonction parentale comprend une variable « père » et une variable « mère » tandis que la fonction enfantine comprend une variable « fils » et une variable « fille ». Selon les besoins, on peut définir pour ces variables des valeurs numériques (quantité) ou alphabétiques (noms propres, par exemple).

Déduction automatique

Du point de vue du traitement informatique, puisqu'ils représentent des ensembles, les arbres autorisent l'usage des opérations ensemblistes (réunion, intersection, différence symétrique...), de l'algèbre de Boole et un certain degré d'automatisation des raisonnements déductifs, basé sur des héritages de propriétés du genre : « Socrate est un homme, tous les hommes sont mortels, donc Socrate est mortel ».

On voit dans l'exemple proposé ci-dessus que n'importe quel terme d'une phrase peut être remplacé automatiquement par les termes qu'il *contient* dans la hiérarchie décrite par l'arbre.

Par exemple, à partir de oeweh damafo ma (les parents éduquent les enfants), l'arbre permet de déduire automatiquement :

- oeweh damafo famafo (la mère éduque les enfants)

- oeweh damafo namafo (le père éduque les enfants)
- oeweh namawowo ma (les parents éduquent le fils)
- oeweh famawowo ma (les parents éduquent la fille)
- oeweh namawowo famafo (la mère éduque le fils)
- oeweh namawowo namafo (le père éduque le fils)
- oeweh famawowo famafo (la mère éduque la fille)
- oeweh famawowo namafo (le père éduque la fille).

Autre exemple, à partir de ak tutudu mama (les conflits sont nombreux dans la famille), l'arbre permet de déduire automatiquement :

- ak tutudu damafo (de nombreux conflits impliquent les enfants)
- ak tutudu famawowo (de nombreux conflits impliquent la fille)
- ak tutudu namawowo (de nombreux conflits impliquent le fils)
- ak tutudu ma (de nombreux conflits impliquent les parents)
- ak tutudu namafo (de nombreux conflits impliquent le père)
- ak tutudu famafo (de nombreux conflits impliquent la mère).

En somme, lorsqu'ils sont appliqués à des métadonnées prenant la forme de séries de phrases IEML, les arbres peuvent autoriser l'*inférence automatique* d'autres phrases. Je n'entre pas ici dans le détail des règles

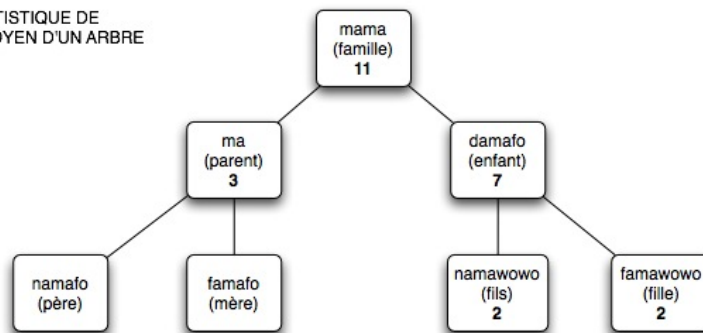
d'inférence automatique, qui supposent une précision de la nature logique des phrases (propositions universelles ou existentielles, par exemple). Mon but est seulement de suggérer le potentiel d'IEML en la matière. Les phrases automatiquement déduites peuvent à leur tour faire l'objet de sélections, d'analyse, de synthèse, de rangements par ordre de priorité

ou de proximité par rapport à une phrase de référence.

Induction automatique

Les arbres peuvent également servir à analyser la distribution des concepts présents dans les données.

ANALYSE STATISTIQUE DE SÉRIES AU MOYEN D'UN ARBRE



Phrases performatives (verbe objet medium)

wawe damafo ma (les parents engendrent les enfants),
wuwe ma damafo (les enfants aiment les parents),
oeweh damafo ma (les parents éduquent les enfants).

Phrases propositionnelles (sujet attribut condition)

namawowo kamafo famawowo (le fils est le frère de la fille),
famawowo kamafo namawowo (la fille est la soeur du frère),
ak tutudu mama (les conflits sont nombreux dans la famille).

Le schéma ci-dessus représente le même arbre et les mêmes séries de phrases IEML que le schéma précédent mais, cette fois-ci, l'arbre est utilisé comme un analyseur statistique des données en série. Chaque occurrence d'un terme IEML dans les séries est comptabilisée au niveau du noeud correspondant dans l'arbre. De plus, chaque noeud de l'arbre totalise les quantités comptabilisées dans les « feuilles » qui dépendent de lui.

Ce type d'analyse statistique peut servir de base à *l'automatisation de raisonnements inductifs*. Supposons qu'une règle fixe à 8 le seuil d'occurrences de concepts familiaux dans un ensemble de données pour déclencher la conclusion automatique que ces données concernent la famille. Comme, dans notre exemple, le nombre d'occurrences des concepts familiaux dépasse 8 (il est égal à 11), le modèle arborescent et la règle permettent ensemble d'inférer automatiquement que les données concernent la famille. Il est évidemment possible de raffiner les règles inductives autant que l'on voudra, en précisant, par exemple, le

rapport numérique entre occurrences de concepts liés aux parents et aux enfants, ou en ajoutant des règles logiques.

Différences hiérarchiques et ontologiques

On peut utiliser les arbres pour mesurer des *différences hiérarchiques* entre concepts. A partir de l'exemple ci-dessus, on peut calculer automatiquement...
- que la différence hiérarchique entre les concepts famafo (mère) et famawowo (fille) est nulle puisque les deux concepts sont de même niveau hiérarchique,
- que la différence hiérarchique entre famafo (mère) et ma (parent) est de un degré,
- que la différence hiérarchique entre famafo (mère) et mama (famille) est de deux degrés.

Un arbre permet également de mesurer des *différences ontologiques* entre deux concepts de même degré hiérarchique, en fonction du nombre d'embranchements franchis pour aller de l'un à l'autre. On peut voir sur notre exemple que la différence logique entre namafo (père) et famafo (mère) est d'un degré, tandis que la différence logique

entre namafo (père) et namawowo (fils) est de trois degrés.

Données arborescentes

Les arbres IEML ne servent pas seulement d'hypothèses permettant d'extraire des informations à partir des données, ils peuvent aussi servir à la description - ou indexation - des données d'observation. Un grand nombre de structures de données se laissent décrire comme des arbres, comme par exemple des hiérarchies de dossiers et de sous-dossiers ou des textes structurés en chapitres et sous-chapitres. Les arbres peuvent décrire des ontologies, des classifications, des analyses en partie et sous partie, des généalogies, etc.

Un sous-programme de recherche en mathématiques IEML pourrait porter sur des méthodes permettant de mesurer les similitudes ou analogies conceptuelles entre les arbres. Il serait alors possible de chercher dans une masse de données en forme de graphes arborescents, les arbres qui « ressemblent » le plus à un arbre hypothétique.

Les séries : relations d'ordre

Le troisième genre de graphes conceptuels est la série, ou séquence linéaire de phrases IEML. On vient d'en voir quelques exemples dans l'examen des arbres. IEML rend possible l'explicitation de la *raison* d'une série (succession temporelle, emboîtement logique ou autre).

Données sérielles

Les *données d'observation* sérielles peuvent représenter :

- des listes en général,
- des séries temporelles,
- des rangements par ordre de fréquence ou de valeurs associées aux variables.

Ces données peuvent être analysées ou développées logiquement (inférences automatiques) par des graphes conceptuels arborescents ou synthétisées en patterns conceptuels à partir de claviers. Elles peuvent aussi faire l'objet d'analyses de fréquence ou de cycles. En IEML, chaque lettre ou combinaison de lettres - jusqu'au niveau de la phrase - possède une signification distincte et représente une adresse sémantique unique. Il est donc plus aisé de réaliser des analyses automatiques pertinentes de fréquences sémantiques à partir de séries temporelles de

concepts qu'à partir de données en langues naturelles.

Hypothèses sérielles

Du côté des *hypothèses*, les séries sont impliquées de deux manières : a) les séries de référence, b) les critères de rangement. Après avoir brièvement évoqué les séries de référence, je vais principalement exposer dans la suite de ce chapitre les méthodes de rangement sériel de concepts qui sont inhérentes au métalangage de l'économie de l'information.

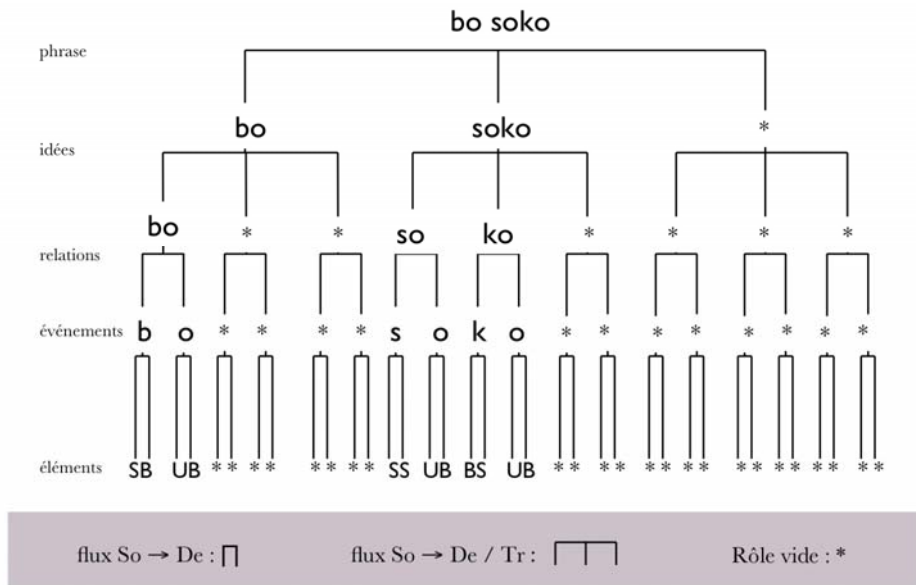
a) On peut adopter une série de concepts comme *modèle de référence* et chercher ensuite les séries qui « ressemblent » le plus à cette référence parmi les données. L'automatisation de ce type de recherche suppose la mise au point d'algorithmes spécialisés, qui pourraient, par exemple, s'inspirer de ceux qui sont utilisés en bio-informatique pour comparer les séquences génétiques.

b) Plutôt qu'une série particulière de phrases, on peut adopter comme hypothèse un *ordre de rangement sériel*, ou système de numérotation d'un ensemble de phrases sélectionnées. La grammaire d'IEML, qui ne fait qu'un avec son système d'adressage sémantique, permet en effet de classer et de *ranger automatiquement les données* sélectionnées selon un grand nombre de critères, au choix. Pour comprendre le fonctionnement de ces rangements automatiques, il est nécessaire de connaître la structure d'adresse d'une phrase IEML, que je vais maintenant résumer.

Niveaux d'articulation des concepts

Le diagramme ci-dessous met en évidence la structure de la phrase bo soko (signifiant : la langue de l'intelligence collective). On voit d'abord qu'il existe un format (une structure d'adresse) fixe que chaque phrase remplit plus ou moins en laissant des « places vides ». On remarque également que les cinq niveaux d'articulation de la phrase se traduisent automatiquement les uns dans les autres. Mais cette stricte équivalence syntaxique entre les niveaux n'empêche pas que chacun d'eux soit susceptible d'une interprétation distincte! Suivons par exemple les différents niveaux d'interprétation de soko, du bas vers le haut.

une phrase ieml



Au niveau des *éléments*, on a : signe (S), signe (S), virtuel (U), être (B), être (B), signe (S), virtuel (U), être (B).

Au niveau des *événements*, on obtient :

- S → S = s = pensée
- U → B = o = vouloir
- B → S = k = société
- U → B = o = vouloir

Au niveau des *relations*, on obtient :

- s → o = so = souci de la pensée
- k → o = ko = désir de lien social.

Au niveau des *idées*, on obtient so → ko = soko : intelligence collective.

Puisque la *phrase*, commençant par une consonne (appartenant donc au pôle sémantique) est propositionnelle, sa forme

grammaticale est du type : sujet - attribut - condition. Comme il n'y a pas d'idée condition,

- on obtient :
- sujet = bo = articulation linguistique
- attribut = soko = intelligence collective
- sujet + attribut = langue de l'intelligence collective.

La phrase bo soko contient les significations distinctes des *cinq* niveaux d'articulation des concepts. C'est pourquoi il est théoriquement nécessaire - et pratiquement faisable - de pouvoir choisir des critères de classement et de rangement à ces cinq niveaux. Le diagramme ci-dessous montre la structure d'une adresse de phrase IEML et les critères de sélection et de rangement des concepts qui lui sont liées.

Structure d'adresse d'une phrase IEML

niveaux d'
articulation

5	catégorie de phrase																																			
4	catégorie d'idée 1												catégorie d'idée 2												catégorie d'idée 3											
3	catégorie de relation 1				catégorie de relation 2				catégorie de relation 3				catégorie de relation 4				catégorie de relation 5				catégorie de relation 6				catégorie de relation 7				catégorie de relation 8				catégorie de relation 9			
2	évén 1		évén 2		évén 3		évén 4		évén 5		évén 6		évén 7		évén 8		évén 9		évén 10		évén 11		évén 12		évén 13		évén 14		évén 15		évén 16		évén 17		évén 18	
1	élé 1	élé 2	élé 3	élé 4	élé 5	élé 6	élé 7	élé 8	élé 9	élé 10	élé 11	élé 12	élé 13	élé 14	élé 15	élé 16	élé 17	élé 18	élé 19	élé 20	élé 21	élé 22	élé 23	élé 24	élé 25	élé 26	élé 27	élé 28	élé 29	élé 30	élé 31	élé 32	élé 33	élé 34	élé 35	élé 36

L'adresse d'une phrase IEML a 36 places distinctes

1. Le premier niveau d'articulation a 36 places
2. Le second niveau en a 18
3. Le troisième niveau en a 9
4. Le quatrième niveau en a 3
5. Le cinquième niveau n'en a qu'une

Critères de sélection et de rangement offerts par *chaque* place d'adresse

► Premier niveau d'articulation

2 catégories d'éléments (O, M) / 5 éléments de substantialité croissante (U1, A2, S3, B4, T5)

► Deuxième niveau

6 catégories d'événements / 10 degrés de substantialité / 5 sources / 5 destinations

► Troisième niveau

6 catégories de relations / 8 degrés de substantialité / 4 types de source / 4 types de destination

► Quatrième niveau

14 catégories d'idées / 6 degrés de substantialité / 2 types de sources / 2 types de destination / 2 types de traductrice

► Cinquième niveau

14 catégories d'idées / 6 degrés de substantialité / 2 types de sources / 2 types de destination / 2 types de traductrice

Il est possible de choisir, pour n'importe quelle place d'adresse, un ou plusieurs critères. On peut alors sélectionner les phrases répondant à ces critères, ou bien encore donner des numéros de priorité aux critères et ranger les phrases en conséquence. La combinaison de plusieurs critères permet de créer autant d'ordres linéaires ou de principes de rangement sériels que l'on veut⁴.

Ordres de rangement des concepts

On peut décider, par exemple, de ranger les phrases en fonction de leur premier élément (niveau 1, place 1, 5 éléments). Appelons ce critère « critère n°1 ». On obtient ainsi cinq classes de phrases:

- 1) commençant par U (virtuel),
- 2) commençant par A (actuel),
- 3) commençant par S (signe),
- 4) commençant par B (être)

⁴ Les possibilités de rangement inhérentes à IEML sont liées à un adressage numérique qui double l'adressage alphabétique. On trouvera le détail de cet adressage numérique dans le document « IEML, finalités et structures »

5) commençant par T (chose).

Mais, à l'intérieur de chaque classe, les phrases ne sont pas rangées. Pour ce faire, on peut adopter un second critère de rangement. Par exemple, le caractère sémantique ou pragmatique du premier élément de la seconde idée de la phrase (niveau 1, place 13, catégories d'éléments). Appelons ce critère « critère n°2 ».

On obtient ainsi 10 classes de phrases, ordonnées de cette manière :

- 1) commençant par U, avec deuxième idée *pragmatique*,
- 2) commençant par U, avec deuxième idée *sémantique*,
- 3) commençant par A, avec deuxième idée *pragmatique*,
- ... et ainsi de suite jusqu'à :
- 10) commençant par T avec deuxième idée *sémantique*.

Ces dix classes sont maintenant distinguées et ordonnées, mais on peut encore vouloir faire des subdivisions à l'intérieur de chacune d'elles à partir d'un « critère n°3 » et ranger les phrases en conséquence, et ainsi de suite à partir d'autres critères.

Dans l'exemple qui vient d'être esquissé, les critères de rangement étaient pris au niveau des éléments. Mais on peut en prendre à tous les niveaux d'articulation :

phrases, idées, relations, événements ou éléments. Il est donc possible de ranger les concepts avec autant de précision que l'on voudra. Pour prendre un autre exemple, on peut sélectionner toutes les phrases de deux idées dont la première idée est *mama* (famille) et les ranger par ordre alphabétique IEML (wo, wa, wu, we, y, o, e, u, a, i, j, h, p, g, c, x, s, b, t, k, m, n, d, f, l).

Au niveau de la catégorie de phrase, les critères de rangement peuvent porter sur les variétés grammaticales auxquelles les phrases appartiennent (performatives ou propositionnelles), leur substantialité, leur taille (une, deux ou trois idées), ou les genres d'idées qui composent leur source, leur destination et leur éventuelle traductrice.

Au niveau des catégories d'idées, et pour chacune des trois places, les critères de rangement peuvent porter sur leur nature grammaticale, leur substantialité, leur taille (une, deux ou trois relations) ou sur les genres de relations qui composent leur source, leur destination et leur éventuelle traductrice.

Au niveau des catégories de relation, et pour chacune des neuf places possibles, on peut adopter des critères de rangement portant sur les types grammaticaux, la substantialité ou les types d'événements - source ou destination - qui les composent.

Au niveau de chacune des dix-huit places d'événements possibles, on peut adopter des critères de sélection ou de rangement portant sur leur identité, leur catégorie grammaticale (deux types de verbes voyelles et deux types de noms consonnes), leur substantialité, leur source ou leur destination.

Finalement, comme on l'a vu, pour chacun des 36 éléments possibles d'une phrase, on peut adopter des critères de rangement portant sur l'identité de l'élément ou sur la catégorie (pragmatique ou sémantique) à laquelle il appartient.

Je répète que toutes ces possibilités de sélection, de rangement et de numérotation des concepts ne sont pertinentes que parce que la syntaxe d'IEML est conventionnellement isomorphe à sa sémantique. C'est précisément cette isomorphie qui en fait un système optimal pour l'adressage sémantique automatisé.

Pour la présente discussion, un ordre de rangement des concepts - ou système de numérotation - peut être considéré comme une hypothèse (de type « modèle ») dont la

pertinence est testée par l'application sur un ensemble de données.

Différences sérielles

A chaque ordre de rangement (ensemble priorisé de critères) correspond l'attribution de numéros d'ordres aux concepts contenus dans les données d'observation. On peut considérer un principe de rangement comme une « règle virtuelle » permettant de mesurer des *différences sérielles* entre concepts. Chaque ordre de rangement (chaque règle virtuelle) engendre évidemment un système de différences distinct, reflétant la perspective cognitive du chercheur.

Je voulais montrer que le métalangage de l'économie de l'information permet au plus grand nombre possible d'hypothèses « sérielles » de s'exprimer et d'être testées. La chose me semble démontrée, au moins pour les séries de *concepts*. Le rangement linéaire automatique de graphes conceptuels (séries, arbres et claviers) devra faire l'objet de sous-programmes de recherche particuliers en mathématiques IEML.

Rôles cognitifs des graphes conceptuels

Avant de conclure la discussion sur le champ ouvert par le métalangage au test automatique des hypothèses scientifiques, je rappelle ma définition - minimaliste - de ce qui rend *scientifique* une hypothèse :
- elle est cohérente, explicite et partageable malgré sa complexité ;
- elle vise des tests maximisant l'extraction d'information à partir des données.

L'examen des graphes conceptuels du métalangage de l'économie de l'information nous a montré qu'ils pouvaient modéliser au moins trois catégories d'hypothèses : les claviers, les arbres et les séries. Chacun de ces *types* de graphe conceptuel joue un *rôle cognitif* distinct, mais cette différence de rôle n'empêche pas les graphes conceptuels de communiquer d'un type à l'autre puisqu'ils utilisent le même système d'adressage de leurs concepts.

Rôle des claviers

Les claviers décrivent les *relations d'indépendance et d'interdépendance* entre concepts. Chaque *concept-clé* d'un clavier représente une *dimension* distincte d'un *espace*

conceptuel. Un concept-clé joue le rôle d'un récepteur particulier dans une sorte de rétine intellectuelle élaborée pour produire automatiquement des *patterns conceptuels* par exposition aux données. A partir de ces patterns conceptuels précisément adressés, il devient possible de discerner des dynamiques complexes de patterns.

Rôle des arbres

Les arbres décrivent des relations ensemblistes : hiérarchies d'ensembles, de sous-ensembles et d'éléments. Les *concepts feuilles* d'un arbre sont entre eux dans des *relations d'appartenance*. Les combinaisons de leurs rapports se décrivent bien au moyen des opérateurs logico-ensemblistes classiques : et (réunion), ou (inclusion), sauf (différence ensembliste) et autres. A partir des hiérarchies de concepts établies par les arbres et des données auxquels ils sont exposés, des concepts peuvent être inférés automatiquement. Les règles d'inférence logique sont de deux types : déductives ou inductives. Les règles déductives décrivent la transmission de l'information de la hiérarchie conceptuelle vers les données. Elles ajoutent des séries de concepts aux données initiales. Les règles inductives décrivent la circulation de l'information des données d'observation vers les sommets de la hiérarchie conceptuelle. Elles abstraient l'information vers des concepts de niveau hiérarchique supérieur en passant par le déclenchement de seuils.

Rôle des séries

Les modèles de séries décrivent des *ordres de rangement linéaire*, des systèmes de numérotation de concepts. Parmi toutes les

manières de ranger les concepts selon le schème de la succession (avant/après), on peut souligner l'importance pratique des *séries temporelles* et des *ordres de priorités* (servant à coordonner l'action et à opérer des choix). L'application des ordres de rangement à des données d'observation variées produit automatiquement des séries co-ordonnées.

Instruments de mesure

Chaque type de graphe conceptuel fonctionne comme un instrument de mesure. En tant que *système de coordonnées* d'un espace conceptuel, les claviers autorisent le calcul géométrique de *distances sémantiques* entre les patterns conceptuels révélés par les données. Les arbres fournissent des *étalons logiques* permettant de mesurer les *différences de degré hiérarchique* entre concepts et des *différences ontologiques* entre concepts de même degré hiérarchique. Enfin, les ordres de rangement (*systèmes de numérotation* de concepts), lorsqu'ils sont appliqués aux données, permettent de mesurer des *intervalles sériels*, ou ordinaux, entre concepts. Toutes ces mesures de distances peuvent alimenter des fonctions automatiques de création d'hyperliens entre documents ou parties de documents.

Les graphes IEML peuvent être considérés comme des instruments scientifiques pour mesurer les rapports entre concepts dans la mesure où leur opération est entièrement explicite, invariable et reproductible. Le même graphe appliqué aux mêmes données d'observation donnera toujours exactement les mêmes mesures.

GRAPHES	Claviers	Arbres	Séries
CONCEPTUELS	Espaces conceptuels	Hiérarchies de concepts	Séquences de concepts
Mathématiques	géométrie	logique	Arithmétique
3 Types de Relations entre concepts décrite par le graphe	indépendance / interdépendance	appartenance (ensemble/sous- ensemble/élément)	ordre linéaire, système de numérotation
Modèles	concepts = dimensions coordonnées d'un espace conceptuel	concepts = feuilles hiérarchies de concepts, règles logiques	concepts = n° ordinaux critères de rangement, séquences temporelles
Il existe une qté infinie de modèles adressables			
Théories Application du modèle à des données	patterns conceptuels, dynamiques complexes de patterns conceptuels	concepts inférés déductivement inductivement	séries de concepts ou numérotés

Instruments de Mesure distances sémantiques

Métaphore neuronale cerveau droit

Métaphores neuronales et cognitives

Le programme de recherche ici défendu ne vise pas la simulation informatique d'une intelligence humaine individuelle mais l'augmentation de l'intelligence collective de communautés de chercheurs par l'automatisation partielle de certaines fonctions cognitives et l'exploitation optimale de la mémoire disponible. Les graphes conceptuels peuvent multiplier la puissance de traitement symbolique des chercheurs en leur permettant de manipuler de manière significative, cohérente, explicite, partageable et automatique des masses énormes de données numériques. L'usage des claviers correspond métaphoriquement à l'opération du *cerveau droit*, capable de *perception globale* et d'appréhension holistique des situations. Le maniement des arbres correspond à l'usage du *cerveau gauche*, *logique et analytique*. Quant aux séries et aux rangements linéaires, ils correspondent plutôt au *cervelet* spécialisé dans la *coordination motrice*, qui se déroule nécessairement dans le temps et aboutit à des résultats séquentiels, ainsi qu'au *lobe frontal*, impliqué dans les *choix délibérés* qui appellent forcément des ordres de priorités. Ces trois types de processus cognitifs peuvent travailler de concert parce qu'ils utilisent un identique système d'adressage des concepts. Ils sont automatisable parce que le système d'adressage des concepts donne prise à des opérations mathématiques.

Au commencement de cet exposé sur les graphes conceptuels, je demandais s'ils étaient capables de formaliser des hypothèses scientifiques en vue d'un test automatique sur la mémoire numérique. Au terme de cette revue des possibilités mathématiques offertes par IEML, il semble que la réponse doive être positive.

Un moteur de recherche ouvert basé sur IEML devrait donc en principe satisfaire non seulement aux contraintes fonctionnelle (liberté d'expression et de questionnement) et formelle (automatisation des recherches sémantiques au moyen d'une méthode générale) mais également à la contrainte

différences logiques & différences sérielles
différences
hiérarchiques

cerveau gauche cervelet / lobe frontal
épistémologique (permettre une recherche d'information de type scientifique, capable de tester des hypothèses complexes de manière reproductible). Ceux qui se consacrent à ce programme de recherche ont maintenant la tâche de confirmer cette conjecture par des essais en vraie grandeur.

Conclusion

Vers une réflexivité de la mémoire numérique

Ce n'est pas seulement la taille de son cerveau qui rend l'être humain capable d'intelligence réflexive. A ce compte, il serait surpassé par les éléphants et les baleines. Ce n'est même pas le rapport entre le poids de son corps et celui de son cerveau : les paléoanthropologues nous disent que l'homme de Neandertal avait proportionnellement plus de matière grise que l'homo sapiens qui l'a supplanté. La pensée humaine est autoréférentielle parce qu'elle s'enracine dans une *faculté de langage*, unique dans le règne animal, qui fait de notre espèce l'exploratrice par excellence de la complexité symbolique:

1) du côté de la *syntaxe*, nous pouvons combiner de façon régulière des symboles signifiants selon une hiérarchie ouverte de niveaux d'articulation ;

2) du côté de la *sémantique*, nous pouvons forger à l'aide de ces symboles une quantité potentiellement infinie de représentations distinctes, elles-mêmes organisables en une multitude d'ensembles et de sous-ensembles ;

3) du côté de la *pragmatique*, l'usage du langage nous permet de nouer et de dénouer une variété virtuellement infinie de liens sociaux et de relations humaines.

Cette puissance symbolique propre à notre espèce est la source de la culture - l'intelligence collective humaine - comme de son évolution cumulative et innovante. Au cours de l'histoire, plusieurs mutations dans les systèmes de signes et les médias ont permis d'augmenter cette intelligence collective : écriture, imprimerie, médias électroniques, ordinateurs. Dans cette ligne de pensée, de nombreux auteurs ont associé au

développement du cyberspace le franchissement d'un seuil d'intelligence collective.

On a souvent comparé l'Internet à un « cerveau global » en voie de constitution. Ce nouveau cerveau global organise l'échange d'information entre ordinateurs dans le même *réseau physique*, malgré la variété des processeurs et des systèmes d'exploitation. Il relie les documents numériques dans le même *réseau hypertextuel* malgré l'hétérogénéité de leurs modes de représentation, de leur langues, de leurs sujets et de leurs points de vue. Il rassemble enfin une nébuleuse de *communautés virtuelles*, qui produisent, explorent et reçoivent une foule de messages variés tout en alimentant une mémoire numérique commune. Le cerveau global en émergence, même s'il a déjà permis d'immenses avancées dans la communication et le traitement coopératif de l'information, ne nous a pas encore fait franchir de seuil décisif d'intelligence collective. Ce cerveau global est nécessaire, il représente une condition *sine qua non* du franchissement d'un seuil d'intelligence collective, mais il ne suffit pas. Un seuil irréversible ne sera franchi que par l'institution - conventionnelle - d'un système symbolique capable de faire accéder l'intelligence collective humaine à la conscience réflexive à *l'échelle du nouveau cerveau global*.

IEML - ou n'importe quel langage *ayant les mêmes caractéristiques* - peut jouer ce rôle de langue de l'intelligence collective contemporaine pour au moins quatre raisons.

1) IEML bâtit, pour commencer, *un pont entre langues naturelles*. N'importe quel graphe de mots écrit dans une langue naturelle au moyen d'un éditeur IEML peut être lu dans n'importe quelle autre langue naturelle supportée par le dictionnaire IEML.

2) IEML forme, ensuite, *un pont entre cultures, disciplines, domaines de connaissances, contextes, terminologies, ontologies, etc.* La structure indissolublement syntaxique et sémantique de ce métalangage permet en effet de déterminer automatiquement des relations entre concepts, des distances sémantiques entre documents et des synthèses comparables quelles que soient l'hétérogénéité des corpus considérés.

3) IEML construit *un pont entre humains et ordinateurs* en donnant aux manipulateurs automatiques de symboles les

moyens d'analyser et de computer la complexité sémantique et pragmatique humaine... lorsque cette complexité est exprimée en IEML. De même que la majorité des utilisateurs d'ordinateurs contemporains n'ont pas besoin d'entrer directement en contact avec le binaire ou même avec des langages de programmation, la majorité des utilisateurs humains d'IEML n'auront pas besoin d'apprendre le métalangage, qu'ils manipuleront au moyen d'interfaces iconiques, ou dans la langue naturelle de leur choix.

4) Enfin, IEML instaure *une réflexivité des opérations cognitives à l'échelle du cerveau global*. En inscrivant la mémoire numérique dans une économie de l'information ouverte, IEML lui donne la capacité de se penser elle-même de façon intrinsèque et distribuée. Nous avons vu, en effet, que le métalangage articule les hypothèses dans le même langage que les données sur lesquelles elles portent : ce sont des graphes conceptuels. Dans une phase particulière de la recherche, n'importe quel ensemble circonscrit de graphes IEML peut jouer le rôle de filtre réfléchissant, de point de vue cognitif ou d'outil intellectuel c'est-à-dire *d'hypothèse*, vis-à-vis d'un vaste ensemble hétérogène d'autres graphes, qui jouent le rôle de *données* d'observation. Mais, dans d'autres phases de la recherche, perspectives cognitives et données analysées peuvent échanger leurs rôles, tour à tour miroitantes et reflétées, selon une dynamique distribuée de cognition autoréférentielle. Les graphes conceptuels sont en relation P2P : sujets et objets de connaissance, outils et matières conceptuels peuvent symétriquement se substituer les uns aux autres. La mémoire numérique peut donc devenir un *milieu* de réflexion intellectuelle.

Perspectives de développement

La mise en oeuvre d'un moteur de recherche sémantique basé sur IEML, capable d'indexer et d'interroger une portion significative de la mémoire numérique, représente un effort de recherche scientifique et technique concerté de longue haleine. Parmi les tâches à accomplir, citons notamment :

- le développement du *vocabulaire* d'IEML⁵ ;

⁵ IEML comptait environ 2500 idées (ou « mots ») en 2006 et ce nombre devrait être

- la traduction de ce vocabulaire dans les principales *langues naturelles* utilisées sur le Web⁶ ;

- la construction et la mise au point *technique* des divers modules du moteur de recherche sémantique (à sources ouvertes), y compris les outils d'indexation automatique et les interfaces de navigation pour le grand public ;

- la formation de *spécialistes* en sciences et technique de l'information et gestion des connaissances qui soient capables de manier cette langue scientifique ;

- la formation de *chercheurs* en économie de l'information capables de perfectionner le métalangage et de développer son potentiel sur le plan mathématique comme dans le domaine des sciences humaines.

Je voudrais maintenant terminer cet article en évoquant deux domaines de développement de ce programme de recherche qui me semblent particulièrement prometteurs : l'informatique cognitive et les sciences humaines.

La formalisation de la logique et de l'arithmétique a permis *l'automatisation des calculs* arithmétiques et logiques et, en fin de compte, la naissance de l'informatique classique. Les machines à calculer organisent un parallélisme, une sorte de miroir formel, entre des transformations physiques mesurables et contrôlables (sur leur face matérielle) et des calculs arithmétiques et logiques (sur leur face logicielle). Grâce à la formalisation de la sémantique et de la pragmatique proposée par IEML, on peut imaginer la naissance d'une *informatique cognitive*, capable de mettre en parallèle les calculs mathématiques automatisés par l'informatique classique avec les opérations conceptuelles (sémantiques et pragmatiques) explorées par les sciences de l'homme.

Cela ne rendra pas obsolètes les résultats des recherches en intelligence artificielle ou en théorie des jeux. Bien au contraire, l'informatique cognitive valorisera

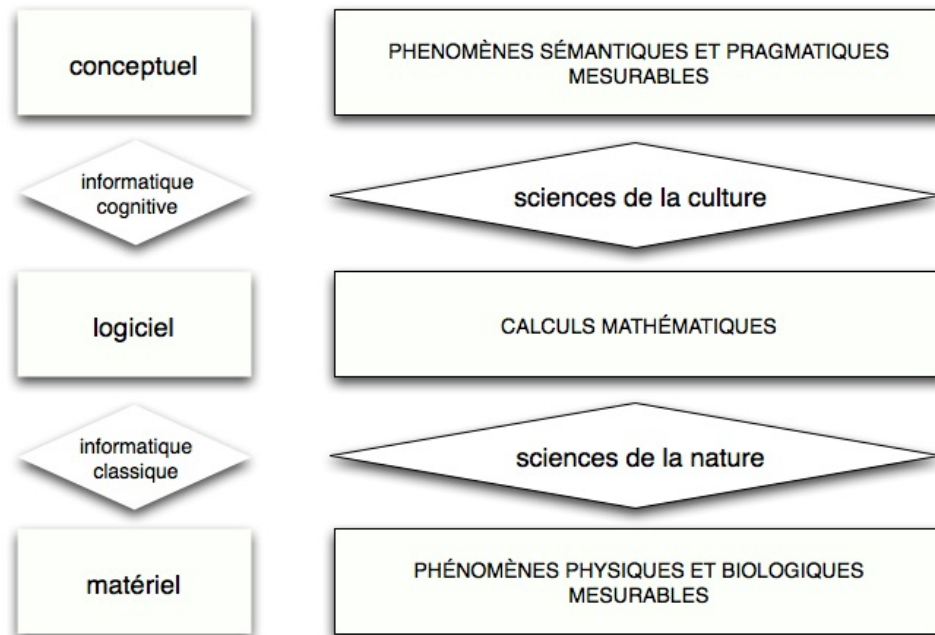
les acquis de ces disciplines en les enrichissant d'un contenu conceptuel.

Le diagramme ci-dessous montre que l'informatique classique joue le rôle d'interface entre le niveau matériel et le niveau logiciel, tandis que l'informatique cognitive opère la liaison réglée entre le niveau logiciel et le niveau conceptuel.

multiplié par 5 en 2009 pour atteindre plus de 12 000 idées.

⁶ En 2006, le vocabulaire IEML est seulement interprété en français et en anglais. Il devrait également être traduit en espagnol et portugais avant 2008.

LES TROIS NIVEAUX



Le diagramme suggère également une mutation dans les sciences de la culture. L'invention de la presse à caractères mobiles par Gutenberg, la fabrication de nouveaux instruments d'observation (téléscope et microscope), comme le choix épistémologique d'un espace unique, infini et mathématisable, ont favorisé une révolution scientifique dans les sciences de la nature aux seizième et dix-septième siècles. Dans un proche avenir, il est possible que les développements ultérieurs du cyberspace, et notamment la construction d'un moteur de recherche sémantique ouvert et les avancées d'une informatique cognitive, voient l'avènement d'une révolution scientifique dans les sciences humaines comparable à ce que fut la révolution de la science expérimentale au XVII^e siècle. Les acteurs de cette mutation épistémologique auront probablement tendance à considérer les phénomènes sociaux et culturels humains comme des processus cognitifs à l'échelle collective.

Je rappelle pour finir comment et à quelles conditions les processus de cognition collective (ou d'économie de l'information signifiante) peuvent devenir observables, navigables et modélisables par les sciences de l'homme et de la société.

1) La mémoire numérique - comme l'économie de l'information qui l'exploite et l'alimente - reflète une part croissante de la

culture, de la communication et des transactions humaines.

2) Parce qu'elle est codée en binaire, interconnectée et accessible par un réseau ubiquitaire, cette mémoire peut faire l'objet d'observations et de traitements reproductibles.

3) L'économie de l'information (cycles d'offre, de demande, de production et d'extraction) qui exploite et cultive la mémoire numérique ne porte pas seulement sur des symboles physiques - une dynamique des signifiants - mais aussi et surtout sur les concepts qui correspondent à ces symboles - la dynamique des signifiés.

4) L'économie de l'information - et la mémoire numérique qui en forme le capital - ne peuvent donc être pris pour objet de recherche scientifique que si l'on dispose d'un système symbolique mathématisable (un langage scientifique) capable de refléter la complexité et le caractère virtuellement infini de leur *nature conceptuelle*.

Ma thèse est qu'IEML permet d'exprimer de manière formelle, calculable, reproductible et mutuellement compatible les données, théories et modèles des diverses sciences de l'homme. Dès lors, via une indexation adéquate des données numériques, IEML pourrait permettre une observation scientifique non seulement quantitative mais aussi *qualitative* - conceptuelle : sémantique et pragmatique - de l'économie de l'information dans toute son ampleur et sa complexité. Pour

utiliser une métaphore, je pourrais décrire le métalangage de l'économie de l'information comme le « code génétique » (ou code mémétique) de la culture humaine. Je précise tout de suite que, si le code me semble déchiffré, l'ensemble du génome reste à inventorier, pour ne rien dire des écologies d'idées et des évolutions culturelles à explorer... La cartographie dynamique de l'espace cognitif humain sera nécessairement une entreprise collective de longue haleine. Au-delà de la fragmentation disciplinaire et théorique des sciences humaines contemporaines, le cœur de la révolution de l'économie de l'information sera la découverte-exploration constructive d'un univers cognitif multidimensionnel, fractal, unique et infini.

Bibliographie

- Anger R. (Ed.), foreword by D. Dennett, Darwinizing Culture: The Status of Memetics as a Science, Oxford U. P. 2000
- Aristote, De l'Interprétation, (en particulier le début du Chapitre 1), in *Organon II*, trad. Tricot, Vrin, Paris, 1977
- Ascott, Roy, (Ed. Edward A Shanken), *Telematic Embrace*, University of California Press, 2003
- Ashby, R., *Introduction to Cybernetics*, Methuen, London, 1956
- Atlan, H., *Les Etincelles de hasard, T. 1 : Connaissance spermatique*, Seuil, 1999
- Atlan, H., *Les Etincelles de hasard, T. 2 : Atheisme de l'écriture*, Seuil, 2003
- Auroux, S. *La Révolution technologique de la grammatisation*, Mardaga, Liège, 1994
- Austin, J. L., *How to Do Things With Words*, Oxford U. P. 1962
- Authier, M., Lévy, P. (préface de Michel Serres), *Les Arbres de connaissances*, La Découverte, Paris, 1992
- Balpe, J. P., *Contextes de l'art numérique*, Hermès science, Paris, 2000
- Barabasi, Albert Laszlo, *Linked, the New Science of Networks*, Perseus publishing, Cambridge, Mass, 2002
- Bardini, T., *Bootstrapping. Douglas Engelbart, Coevolution and the Origins of Personal Computing*, Stanford UP, 2000.
- Bateson, G. *Steps to an Ecology of Mind*, 2 vol., Chandler, NY, 1972
- Berners Lee, T. Hendler, J. and Lassila, O. *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American, Mai 2001
- Berners Lee, Tim, *Weaving the Web*, Harper, San Fransisco, 1999
- Bickerton, D., *Language and Human Behavior*, University of Washington Press, Seattle, 1995
- Bloom, H., *Global Brain, the Evolution of Mass Mind from the Big Bang to the 21st century*. Wiley & sons, NY, 2000
- Brooks, R.A. & Steels, L. (eds.), *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1995
- Buchanan, Mark *Nexus: Small Worlds and the Groundbreaking Science of Networks*, Norton and Cie, 2002.
- Bush, Vannevar, *As We May Think*; in *The Atlantic Monthly*; July, 1945; Vol. 176, No. 1; p. 101-108 (reproduced in Packer and Jordan, 2001)
- Carruthers, Mary, *The Craft of Thought, Meditation, rhetoric and the making of images*, 400, 1200, Cambridge UP, 2000
- Cassirer, E., *Philosophie des formes symboliques*, Minuit, Paris, 1972
- Castells, M., *The Information Age, Economy, Society and Culture*, (3 vol.) Blackwell, Oxford, 1998
- Centre pour la recherche et l'innovation dans l'enseignement, *Knowledge Management in the Learning Society*, OCDE, Paris, 1999
- De Libera, A. *La Querelle des universaux, De Platon à la fin du Moyen-Age*, Seuil, Paris, 1996
- Deacon, T. W., *The Symbolic Species: The Coevolution of Langage and the Brain*, Norton, NY, 1997
- Debray, R., *Cours de médiologie générale*, Paris, Gallimard, 1991.
- Deleuze, G. & Guattari, F., *Mille Plateaux*. Editions de Minuit, Paris, 1980.
- Deleuze, G. *Logique du sens*, Minuit, Paris, 1969.

- Dessales, J-L., *Aux Origines du langage, une histoire naturelle de la parole*, Hermès Science, Paris, 2000
- Douglas, M., *How Institutions Think*. Routledge & Kegan Paul, London, 1987
- Eco, U., *Segno*, ISEDI, Milan, 1973
- Eisenstein, E., *The Printing Revolution in Early Modern Europe*, Cambridge University Press, 1983
- Engelbart, Douglas, *Augmenting Human Intellect*, Technical Report, Stanford Research Institute, 1962
- Fensel D., Hendler J. A., Lieberman H. and Wahlster W. (Eds.), *Spinning the Semantic Web, Bringing the World Wide Web to Its Full Potential*, MIT Press, 2003
- Foerster (von), H., *Observing Systems*, Intersystems, Seaside CA, 1981
- Foray, D., *L'économie de la connaissance*, La Découverte, Paris, 2000
- Fuller, S. *Knowledge Management Foundations*, Butterworth-Heinemann, Boston, 2002
- Goody, J., *The Domestication of the Savage Mind*, Cambridge University Press, 1987
- Goody, J., *The Logic of Writing and the Organization of Society*, Cambridge University Press, 1987
- Gross Stein, J., Stren, R., Fitzgibbon, J., Maclean, M., *Networks of Knowledge. Collaborative Innovation in International Learning*, University of Toronto Press, 2001.
- Hayek, F. *Economics and Knowledge*, in *Economica* IV (new ser., 1937), 33-54.
- Hayek, F. *Law, Legislation and Liberty*, 3 vol, Routledge & Kegan Paul, London, 1979
- Hayek, F. *The Use of Knowledge in Society*, *American Economic Review*, XXXV, No. 4; September, 1945, 519-30
- Heylighen F. "The World-Wide Web as a Super-Brain: From Metaphor to Model", in *Cybernetics and Systems '96*, R. Trappl (ed.), World Science, Singapore, 1998
- Himanen, P. *The Hacker Ethic and the Spirit of the Information Age*. Random House, NY, 2001
- Jackendoff, R., *Languages of the Mind*, MIT Press, 1992
- Johnson, Steven, *Emergence, The Connected Lives of Ants, Brains, Cities and software*, Scribner, NY, 2001.
- Johnson-Laird, P. N., *Mental Models*, Harvard University Press, England, 1983
- Kelly, K., *Out of Control. The New Biology of Machines, Social Systems and the Economic World*, Addison Wesley, NY, 1994
- Kerckhove (de), D., *Connected Intelligence*, Somerville House, Toronto, 1997 (french translation: Kerckhove (de), D., *L'Intelligence des réseaux*, Odile Jacob, 2000)
- Langacker, R. W. *Foundations of Cognitive Grammar, Vol 1*, Stanford University Press, 1987
- Langton, C. G., (ed.) *Artificial Life*, Santa Fe Institute Studies in the Sciences of Complexity Proceedings, vol. 6, Addison Wesley, Redwood, Calif., 1989
- Langton, Christopher G. (Ed.) *Artificial life: An overview* Cambridge, Mass. MIT Press, 1995
- Latour, B., *Science in Action*, Open University Press, Milton Keynes, 1987
- Leroi-Gourhan, A. *Le Geste et la parole*, Albin Michel, Paris, 1963
- Levine, R., Locke, C., Searls, D., Weinberger, D., *The Cluetrain Manifesto, the End of Business as Usual*, Perseus Books, Cambridge, Mass., 1999
- Levi-Strauss, *Anthropologie Structurale*, Plon, Paris 1958
- Lévy, P. "IEMML, Finalités et structure fondamentale", in *Intelligence Collective*, Rencontres 2006, coordonné par Jean-Michel Penalva, Presses de l'école des mines de Paris, Mai 2006, p.117-136
- Lévy, P. *L'Intelligence collective. Pour une anthropologie du cyberspace* La Découverte, Paris, 1994. English translation : Lévy, P. *Collective Intelligence: Mankind's Emerging World in Cyberspace*, Perseus Books, Cambridge, Mass. 1997
- Lévy, P., *Analyse de contenu des travaux du Biological Computer Laboratory (BCL)*, in *Cahiers du CREA*, 8, Paris, 1986, p. 155 à 191
- Lévy, P., *Cyberculture*, Odile Jacob, Paris, 1997. English translation : Lévy, P.

- Cyberculture, University of Minneapolis Press, 2001
- Lévy, P., *Cyberdémocratie: Essai de philosophie politique*, Odile Jacob, Paris, 2002
- Lévy, P., *De la programmation considérée comme un des beaux-arts*, La Découverte, Paris, 1992
- Lévy, P., *La Machine Univers. Création, cognition et culture informatique*, La Découverte, Paris, 1987
- Lévy, P., *Les Technologies de l'intelligence. L'avenir de la pensée à l'ère informatique*, La Découverte, Paris, 1990
- Lévy, P., *L'invention de l'ordinateur*, in *Éléments d'histoire des sciences* (dir. Michel Serres) Bordas, Paris, 1989, p. 515 à 535.
- Lévy, P., *L'Oeuvre de Warren McCulloch*, in *Cahiers du CREA*, 7, Paris, 1986, p. 211 à 255.
- Lévy, P., *Qu'est-ce que le virtuel?* La Découverte, Paris, 1995. English translation : Lévy, P. *Becoming Virtual. Reality in the Digital Age*, Plenum Trade, New York, 1998
- Lévy, P., *Wittgenstein et la Cybernétique*, in *Cahiers du CREA*, 7, Paris, 1986, p. 257 à 285.
- Lévy, P., *World Philosophie: le marché, le cyberspace, la conscience*, Odile Jacob, Paris, 2000
- Licklider, Joseph, C R, & Robert Taylor *The Computer as a Communication Device Science and Technology*, April 1968
- Licklider, Joseph, C R, *Man-computer Symbiosis*, IRE Transactions on Human Factors in Electronics, volume HFE-1, pages 4–11, March 1960
- Lin, Nan, *Social Capital: A Theory of Social Structure and Action*, Cambridge University Press, 2001.
- Luhmann, N. *Social Systems*, Stanford UP, CA, 1995.
- Machlup, F., *Knowledge, its Creation, Distribution and Economic Significance*, 3 vol, Princeton UP, 1984
- Maturana, H.R., Varela, F.J., *The Tree of Knowledge: The Biological Roots of Human Understanding*, New Science Library, Boston, 1988
- Mauss, M., *Sociologie et Anthropologie* (avec une introduction de Claude Lévi-Strauss), Presses Universitaires de France, Paris, 1950
- McCulloch, W., *Embodiments of Mind*, MIT Press, Cambridge, 1965
- McLelland, J. L., Rumelhart, D. E. and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Micro-Structure of Cognition*, 2 vol. MIT Press, 1986
- McLuhan, M., *The Gutenberg Galaxy*, University of Toronto Press, 1962
- McLuhan, M., *Understanding Media: The Extensions of Man*, New American Library, NY, 1964.
- Memmi, D. *Le modèle vectoriel pour le traitement de documents*. in *Journal of Semantic Research*, à paraître.
- Minsky, M., *The Society of Mind*, Simon and Schuster, NY, 1997
- Mitchell, William J., *E-topia*, MIT Press, Cambridge, MA., 1999
- Morey, D., Maybury, M. & Thuraingham, B. (Ed.), *Knowledge Management, Classic and Contemporary Works*, MIT Press. Cambridge Mass., 2000
- Morin, Edgar, *La Méthode 4 - Les Idées, leur habitat, leur vie, leur moeurs, leur organisation*, Seuil, Paris, 1991
- Moulin, H. *Axioms of Cooperative Decision Making*, Cambridge University Press, 1991
- Nelson, Theodor Holm, *Literary Machines*, Mindful Press, 1982
- Oram, A., *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, O'Reilly & Associates, Sebastopol, CA, 2001
- Packer, R., Jordan, K. (eds), *Multimedia, from Wagner to Virtual Reality*, Norton, NY, 2001
- Pannaccio, C. *Les Mots, les concepts et les choses, la sémantique de Guillaume d'Occam et le nominalisme d'aujourd'hui*, Bellarmin-Vrin, Paris-St Laurent (Qc), 1992
- Peirce, C. S., *Ecrits sur le signe* (rassemblés par G. Deledalle), Le Seuil, Paris, 1978
- Popper, Karl, *La Connaissance objective*, traduction intégrale et préface de Jean-Jacques Rosat, Flammarion, collection Champs, Paris, 1998 (Objective

- Knowledge, Clarendon Press, Oxford, 1972)
- Porat, M. & Rubin, M. The Information Economy, Government Printing Office, Washington DC, 1977
- Rastier, F. « La triade sémiotique, le trivium et la sémantique linguistique » Nouveaux actes sémiotiques, n# 9, 54 p., 1990.
- Raymond, E.R., The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary, O'Reilly & Associates, Sebastopol, CA, 1999
- Rheingold, Howard, Smart Mobs, The next social Revolution, Perseus Books, Cambridge Mass., 2002
- Rosnay (de), J., L'homme symbiotique, Seuil, Paris, 1995
- Serres, M. Le Système de Leibniz et ses modèles mathématiques, PUF, Paris, 1968
- Serres, M., La Communication, Minuit, Paris, 1968
- Serres, M., La Traduction, Minuit, Paris, 1974
- Shannon, CE, and Weaver, W., Mathematical Theory of Communication, U. of Illinois Press, Urbana, 1964
- Simon, H., Models of Bounded Rationality: Behavioural Economics and Business Organisation, 2 vol. MIT Press, Cambridge, MA, 1982
- Simon, H., Sciences of the Artificial, MIT Press, 1969
- Simondon, G. Du Mode d'existence des objets techniques, Aubier, Paris, 1958
- Sowa, John F., Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley, Reading, MA., 1984
- Sowa, John F., Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole Publishing Co., Pacific Grove, CA, 2000
- Sperber, Dan, La Contagion des idées, Odile Jacob, Paris, 1996
- Stehr, N., Knowledge and Economic Conduct. The Social Foundations of the Modern Economy, University of Toronto Press, 2002.
- Stengers, I. L'invention des sciences modernes, La Découverte, Paris, 1993
- Stewart, Thomas A. Intellectual Capital: The New Wealth of Organizations, Doubleday/Nicholas Brealy, NY, 1997
- Svenonius, E., The intellectual Foundation of Information Organization, MIT Press, 2000.
- Teilhard de Chardin, P. Le Phénomène humain, Seuil, Paris, 1955
- Tofts, D., Jonson, A., Cavallaro, A., (Eds), Prefiguring Cyberculture, an Intellectual History, MIT Press, 2003.
- Varela, F. Autonomie et connaissance, Seuil, Paris, 1989
- Varela, F., Thompson, E., Rosh, E., The Embodied Mind: Cognitive Science and Human Experience, MIT Press, Cambridge, Mass. 1991
- Vigotsky, L., Thought and Language, MIT Press, 1986 (1ère édition russe : 1934)
- Watts, Duncan J., Six degrees, the science of a connected Age, Norton, NY, 2002
- Weinberger, David, Small Pieces Loosely Joined, a Unified Theory of the Web. Perseus Books, Cambridge Mass., 2002
- Wellman, B. and Berkowitz, S.D. (Eds.), Social Structures: A Network Approach, Cambridge University Press, Cambridge, 1988
- Wellman, Barry (Ed.) Networks in the Global Village, Westview Press, Boulder, CO, 1999.
- Wenger, E. Communities of Practice: Learning, Meaning, and Identity. Cambridge: Cambridge University Press, 1998
- Whitehead, A. N., Aventures d'idées (trad. J-M. Breuvert et A. Parmentier), Cerf, Paris, 1993 (ed. originale 1933).
- Wiener, N., The Human Use of Human Beings: Cybernetics and Society, Doubleday, NY, 1950
- Wittgenstein, L., Philosophical Investigations, Blackwell, Oxford, 1958
- Yates, F. The Art of Memory, University of Chicago Press, 1974.

SÉLECTION WEBOGRAPHIQUE

Bien commun, sources ouvertes, archives ouvertes

<http://creativecommons.org/>
<http://www.fsf.org/>

<http://www.opensource.org>
<http://www.openarchives.org/>
<http://inteligenciacolectiva.bvsalud.org/>

Web sémantique et gestion des connaissances

<http://www.lornet.org/>
<http://www.semanticweb.org/>
<http://www.w3.org/>
<http://www.w3.org/TR/REC-xml/>
<http://suo.ieee.org/> (« ontologies » pour ingénieurs)
<http://www.boxesandarrows.com/> (sur l'architecture de l'information)
<http://www.bootstrap.org/> (autour de Douglas Engelbart)
<http://www.cio.com/research/knowledge/> (gestion des connaissances pour directeurs de l'information)
http://solutions.journaldunet.com/dossiers/pratique/knowledge_management.shtml (explications sur la gestion des connaissances)

Sciences de la documentation et bibliothéconomie

Le système de classification de la librairie du Congrès :
<http://www.loc.gov/catdir/cpsol/lcco/lcco.html>
le système de classification de Dewey :
<http://www.tnrndlib.bc.ca/dewey.html>
Dublin core metadata initiative :
<http://dublincore.org/>
Standards pour les musées en ligne
<http://www.chin.gc.ca/English/Standards/index.html>
Ressources éducatives, Canada :
<http://www.cancore.ca/en/>
Learning objects metadata :
<http://ltsc.ieee.org/wg12/>
La classification par facettes et Ranganathan :
<http://www.iawiki.net/FacetedClassification>
Sciences de l'information :
<http://www.urfist.info/>
La bibliothèque virtuelle en santé :
<http://www.bireme.br/bvs/I/home.htm>
Rameau :
<http://rameau.bnf.fr/indexPlan.htm>

Archiver le réseau :
<http://netpreserve.org/about/index.php>

Logiciels sociaux et coopératifs

<http://many.corante.com/>
<http://web2.wsj2.com/>
http://www.wikiindex.com/Wiki_Index
<http://www.lesreseauxsociaux.com/>
<http://www.cooperatique.com/>
<http://del.icio.us/>
<http://www.flickr.com/>
<http://www.connotea.org/>
<http://www.consensusview.com/>
<http://fr.newsutures.com>
<http://www.technorati.com/>
<http://www.feedster.com/>

<http://isdm.univ-tln.fr>