

Source

Meunier. JG. La lecture et l'analyse de texte assistées par ordinateur (LATAO) comme système de traitement d'information. *BOULAG TAL et sciences Cognitives*, 1996-97 no 22: 211-223

La lecture et l'analyse de texte assistées par ordinateur (LATAO) comme système de traitement d'information.

Jean Guy Meunier

Université du Québec à Montréal
Laboratoire d'Analyse Cognitive de l'Information (LANCI) .
1997

La venue récente des technologies du cédérom et encore plus de l'inforoute a mis en évidence les difficultés du traitement l'information textuelle électronique. Il ne suffit plus simplement de stocker des textes ou de les retrouver encore faut-il souvent en découvrir le contenu. Par quelle stratégie peut-on atteindre ce contenu d'un texte? En première approche , il semble facile de répondre à une telle question. On pourrait en effet croire qu'un ordinateur armé d'outils linguistiques pourrait extraire ce contenu, qu'un "moteur de recherche " aurait pour fonction de découvrir le contenu des textes, que des "agents" artificiels intelligents pourraient lire des textes, etc. Malheureusement, la réalité de l'accès au contenu d'un texte n'est pas aussi facile que ne laisse croire les métaphores minières de "l'extraction" du contenu ou de les métaphores mécaniques des "moteurs" qui roulent sur l'inforoute et ou des métaphores maritimes qui naviguent dans le cyberspace. Les TGV de la lecture de texte ne sont pas aussi faciles à construire qu'on le pense. Malgré la richesse et la pertinence du développement technologique, peu de réflexions ont été faites sur le fondement théorique de cette technologie le la lecture et de l'analyse de texte assistées par ordinateur (LATAO). Virbel (1993), Stiegler (1991), Rastier (1987), Balpe et al (1996) sont parmi ceux qui, du moins en France, se sont interrogés sur cette opération cognitive de la lecture assistée par ordinateur. Ils défendent une position théorique qui manifeste l'influence de la longue tradition académique européenne sensible aux dimensions rhétoriques, discursives et philologiques d'un texte. Pour notre part, (Meunier, 1996, Gastaldy et Meunier, 1994) nous avons tenté de montrer la complexité de ce qui était impliqué dans des stratégies informatiques d'accès au contenu textuel. Dans la présente recherche, nous étudierons l'hypothèse selon laquelle l'accès informatique au contenu d'un texte pourrait être conçu comme un *système traitement d'information* (STI). L'objectif visé par une telle analyse est de comprendre davantage les paramètres qui constituent la technologie informatique la lecture et l'analyse de texte assistée par ordinateur.

1-L'accès au contenu de texte comme système de traitement d'information.

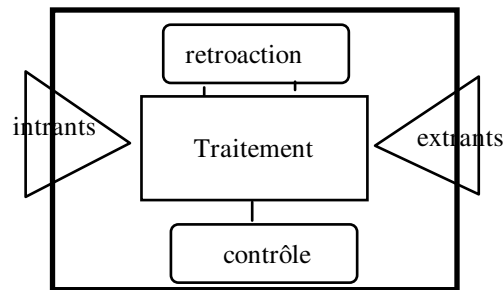
Dans les sciences cognitives avec orientation technologique, il est devenu classique de concevoir une opération cognitive comme un traitement d'information. (STI).

"La science cognitive en général cherche à comprendre les fonctions cognitives humaines en termes d'états mentaux et de processus. ie. en termes d'algorithmes qui réalisent la transformation des données en données de sortie. . . ." Rappaport 1983, cité dans Imbert 1987.

Dans une telle perspective, on peut alors voir les actions cognitives d'accès au contenu d'un texte comme un *traitement d'information* (dont les données d'entrée sont des chaînes de caractères et les données de sortie des " interprétations". Ce type d'approche permettrait de quitter une analyse qui ne porterait que sur la dimension physique de la machine (ordinateur ou cerveau) et pour aborder les *fonctions* effectuées dans le traitement. Dans cette perspective, on pourra constater que cinq grands modèles influencent la conception de ce type de traitement , à savoir : *cybernétique, documentaire, linguistique, IA (intelligence artificielle), émergentiste*. Nous présenterons ces modèles fonctionnels et nous les discuterons en regard de la lecture et de l'analyse de texte assistées par ordinateur.

1. 1- Le modèle cybernétique .

Le conception le plus classique d'un STI est le modèle cybernétique (Ashby, 1956; Lemoigne,1990,etc.). L'architecture d'un tel système se présente comme une boîte de traitement qui reçoit des *intrants* où opèrent *des fonctions de contrôle*, des *rétroactions* et finalement produit des *extrants*. Dans ce type de modèle, l'accès au contenu de texte sera vu comme une *machine abstraite* dont les intrants sont des chaînes de caractères (mots, phrases, paragraphes) du texte qui subissent des transformations par l'application de fonctions de contrôle internes ou de rétroaction.



Bien que un peu âgé, ce modèle n'en demeure pas moins celui qui sous-tend la vulgarisation des technologies de l'information. En effet, cette modélisation cybernétique voit ce type de système comme un "moteur" d'accès au contenu, un peu comme le sont les "moteurs" de recherche " sur le WWW, ou dans les cédérom. Tout comme ces derniers, ce "moteur d'accès au contenu" celui-ci parcourrait " un texte, " mais, parce "sensible à la "signification présente dans le texte", il pourrait en "extraire" le contenu. C'est pourquoi on

lui donnerait le nom d'"agent intelligent de recherche". Seule une échelle de complexité les différencierait ces deux types de moteurs. Malheureusement, comme nous le verrons, les choses ne sont pas aussi simples. Le "carburant" qui transforme des "moteurs" en "agents" ne se trouvent pas dans les stations services des autoroutes de l'information!

Ces métaphores aident certes à vulgariser les technologies de l'information mais elles masquent aussi la véritable nature du traitement en jeu. En effet, le problème d'un système informatique d'accès au contenu d'un texte n'est pas de comprendre comment une machine peut manipuler des signaux électriques -si nombreux soient-ils- mais de comprendre comment elle manipule des signes ou des informations qui dénotent pour des humains un certain contenu signifiant. C'est une question analogue à celle de se demander comment une usine qui manipule physiquement des pièces rondes et minces, de 2 centimètres, en bronze et en nickel pourrait-elle aussi manipuler mathématiquement des "francs français" qui possèdent une valeur marchande. La réponse à une telle question ne se trouve pas dans une théorie physique des alliages de métaux ou le génie mécanique de la machine. Il faut plutôt la chercher dans une théorie sémiotique de la dénotation et théorie mathématique des opérations sur les nombres.

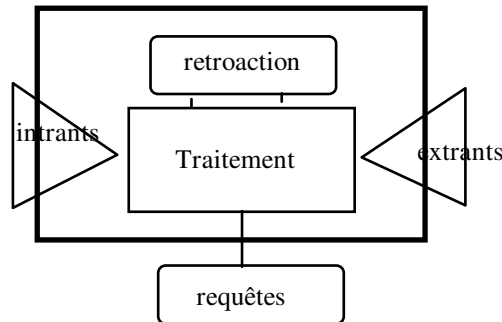
Le modèle cybernétique réussit-il à expliquer ce type de manipulation sémiotique? Malheureusement, non, comme l'a répété régulièrement la critique. En effet, une analyse à la loupe de ce système, montre que, contrairement à ce qu'il prétend, ce type de système ne manipule pas des *signes* mais uniquement des *signaux* physiques. En effet, les intrants d'un système informatique ne sont que des signaux électriques et non des » chaînes de caractères ", des "mots" ou des concepts). Pour affirmer le contraire, il faut que le système possède des règles linguistiques pour identifier ce qu'est un caractère (lettre) mot, idée .etc. Or ces règles ne sont pas explicitement présentes dans le modèle cybernétique.

Et si on acceptait l'hypothèse que les intrants d'un tel système sont des signes, (caractères , mots , concept ,etc) alors il faudrait préciser les fonctions de contrôle qui les traitent, c'est-à-dire qu'il faudrait préciser comment ces fonctions conservent aux signes leur signification et ce, malgré les multiples transformations qu'ils subissent ? Or le modèle est peu loquace sur ces fonctions. De fait, on ne trouvera pas dans ces types de systèmes le contenu de ces fonctions. De fait, elles lui seront données de l'extérieur et elles devront être définies par le constructeur du système. Or, la recherche nous montre bien que ceci ne sera pas facile à réaliser. En effet, il n'est pas assuré que les constructeurs de ces systèmes possèdent une connaissance précise de ces fonctions et qu'ils puissent facilement les identifier et leur donner une formulation adéquate. Les opérations fonctionnelles mises en jeu dans l'accès au contenu d'un texte ne sont encore un savoir des plus partagés par l'humanité.

Ainsi, comme on le voit, le modèle cybernétique n'est pas aussi limpide qu'il le semble. S'il n'est pas précisé davantage, il demeure ambigu sur la nature de ses intrants et silencieux sur ses fonctions de contrôle. Il présente donc une architecture acceptable mais il demeure très général. Il ne permet pas de comprendre comment des signes peuvent être manipulés pour en conserver toute la signification. Pour être plus adéquat le modèle doit approfondir la nature de ces fonctions de contrôle. Ce que tenteront de faire certains des modèles suivants.

1. 2- Le modèle documentaire .

Une des modèles de STI dont l'objet est directement défini en termes de traitement d'information est celui offert par les technologies de gestion documentaire. (Salton, 1987; DuPoirrier 1994, etc.) Ce modèle documentaire voit naturellement l'accès au contenu du texte comme un système de traitement d'information mais où les fonctions de contrôle et de rétroaction sont définies par les requêtes des utilisateurs ou de l'indexation. Ce modèle reprend l'architecture générale d'un STI qu'il adapte à l'accès au traitement de l'information textuelle.



Ici, les intrants du système sont issus de la base documentaire mais peuvent se présenter soit en plein texte, soit sous forme de résumé ou soit sous forme de mots privilégiés (mots-clefs). Via des stratégies d'appariement (mesure de similarité, pondération, etc.) et d'outils (thésaurus, index, langages formels, etc.) le système répond aux requêtes des utilisateurs. Il produit en sortie des réponses avec un degré plus ou moins grand de précision ou de rappel. Il peut même être sensible, (par rétroaction) aux corrections ou reformulations des requêtes effectuées par l'utilisateur. Autrement dit, ce modèle est plus précis dans la définition des fonctions de contrôle et de rétroaction que ne l'était le modèle cybernétique.

Cette précision des fonctions de contrôle améliore assurément le modèle de STI cybernétique, mais elle fait apparaître aussi le problème que présente une telle précision. A savoir: son externalité. En effet, si on y porte bien attention au graphique précédent, on voit que la requête n'est plus une fonction interne du système, mais elle lui est imposée de l'extérieur. C'est l'utilisateur qui définit ce qu'il veut trouver dans base de données documentaire. Or dans le modèle cybernétique, même si elles venaient du constructeur, le système de traitement plaçait ses fonctions de contrôle à l'intérieur du système. Or, dans cette traduction documentaire d'un STI, les fonctions de contrôle sont placées à l'extérieur du système et dépendent des requêtes de l'utilisateur.

Cette externalisation de la fonction de contrôle pose alors un problème sérieux. En effet, le modèle documentaire suppose que celui qui effectue la requête possède une certaine image de la base d'information dans lequel il effectue une fouille. C'est un postulat implicite de ces systèmes. Par exemple, dans un base de données du domaine de l'aéronautique, un utilisateur normal sait qu'il ne peut demander des informations sur les *omelettes tonquinoises!* On sait d'ailleurs que le succès des requêtes effectuées dépendent grandement de la compétence de ceux qui font cette requête. Un technicien spécialiste ne procède aucunement comme un débutant ignare du domaine. Or, la situation est totalement différente lorsque l'utilisateur est placé devant un texte. La fonction première d'un texte n'est pas d'être

un reposoir d'informations mais un laboratoire de création d'informations. En effet, un texte est habituellement écrit pour communiquer quelque chose de neuf à quelqu'un. Il opère dans l'innovation et la création. Ce n'est qu'en second lieu qu'il sera jouera un rôle de reposoir. En conséquence, il faut modifier notre vision de l'attitude d'un utilisateur devant une base de données textuelles. Ici, l'utilisateur ne cherche pas simplement à retrouver ce qu'il sait exister dans la base de données. Au contraire, il est placé dans une attitude de découverte. Il possède une attitude épistémique de "*vouloir savoir*" et non pas simplement de "*vouloir retrouver ce qu'il sait*". L'attitude épistémique est donc totalement différente.

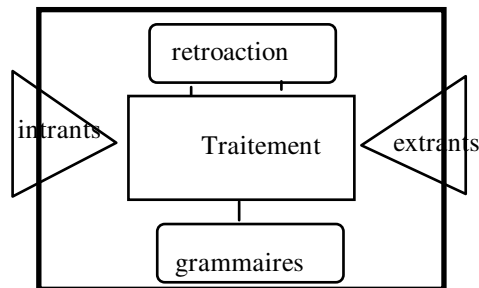
Ainsi, si le modèle documentaire nous amène à voir une différence importante entre la *recherche d'information* et la *découverte de l'information*. L'extraction d'information d'un texte dont on ne connaît pas le contenu n'est pas un processus cognitif similaire à celui de la recherche (rappel) d'information. Autrement dit, pour parler métaphoriquement, un *moteur de découverte d'information* n'est pas un *moteur de recherche d'information*. En conséquence, le modèle documentaire demeure limité pour comprendre comment accéder au contenu d'un texte. La stratégie de la requête par mot clef, ou langage contrôlé n'est pas un outil très utile dans ce contexte de la découverte.

1. 3 -Le modèle linguistique

Le troisième modèle de STI qui s'intéresse au traitement de l'information textuelle cherche à combler certaines lacunes des deux modèles précédents. D'une part, il rapatriera ces fonctions à l'intérieur du système et d'autre part, il les précisera en regard de la nature particulière des intrants.

En effet, ce modèle rappellera une chose évidente: Un texte est avant tout un phénomène linguistique; et un système de traitement de l'information textuelle doit posséder certaines des compétences du locuteur d'une langue. Dans une telle perspective, les intrants du STI linguistique ne sont plus vus comme des signaux physiques mais comme des entités sémiotiques c'est-à-dire des symboles de nature linguistique.

Ainsi, l'architecture de ce système aura comme intrants des symboles qui seront soumis à des fonctions de contrôle qui sont des grammaires (avec ou sans rétroaction) Et en sortie, le système produira des symboles classifiés c'est-à-dire catégorisés linguistiquement.



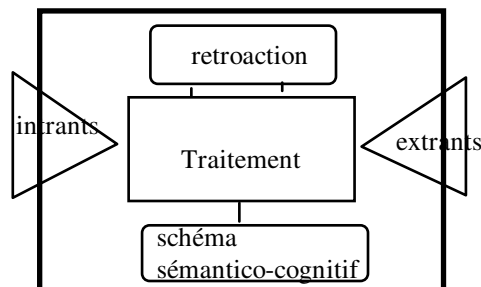
Dans un tel STI, les fonctions de contrôle seront des ensembles de règles dont le rôle est de détecter les propriétés morphologiques, lexicales, syntaxiques, etc. des données d'entrée. Comme ces règles deviennent internes au système, (elles sont dites simuler des compétences linguistiques) on croît alors que ce type de STI pourra, accéder au contenu du texte.

Malheureusement, ce postulat, issu d'une linguistique centrée sur la structure d'une langue (morphologie, lexique, syntaxe) est problématique. En effet, comme le montre la littérature technique sur la question, (Eco 1979, Habermas, 1989; Greimas 1983; Cullioli, 1990; Halliday, 1984; Rastier, 1987; Lenhert 1994) un texte présente une foule de dimensions (sémantique, pragmatique, narrative, rhétorique, argumentative, épistémique, etc.) qui ne relèvent pas comme tel de la grammaire d'une langue mais de son usage en discours. En conséquence, bien que les analyseurs linguistiques puissent traiter adéquatement la langue d'un texte, ils ne sont pas nécessairement en mesure de traiter un texte en tant que texte. De fait, les analyseurs linguistiques (parsers) ne livrent souvent qu'un marquage linguistique du texte. Tout le travail d'accès au du contenu reste à faire.

Ainsi, un ce modèle linguistique nous rappelle que les intrants d'un STI qui doit traiter du texte sont de nature linguistique. Il postule entre autre qu'on peut en préciser la nature linguistique par un marquage catégoriel. Cependant, ce modèle montre aussi sa limite. Si un STI possède des fonctions de contrôle qui ne sont que de nature strictement grammaticale au sens technique de ce terme, c'est-à-dire uniquement orientées sur des phénomènes de nature morpho -syntaxique, il ne peut atteindre véritablement le contenu des textes. Certains se demanderont même, si un tel marquage détaillé est toujours nécessaire. Pour de nombreux théoriciens (Salton, 1989; Spark Jones, 1973; Church, 1990) un marquage linguistique minimum, c'est-à-dire identification des chaînes de caractères qui sont des "mots" suffit pour la recherche d'information et même la découverte d'information, surtout quand le corpus est immense.

1. 4 -Le modèle IA .

Le quatrième modèle, celui que proposeront les chercheurs en intelligence artificielle classique tentera à son tour de toucher plus directement cette dimension " textualité " ou contenu d'un texte. Pour ce faire, le modèle réaffirmera que ces fonctions de contrôle sont internes au système mais il postulera qu'elles ne sont pas uniquement de nature grammaticale. Pour la tradition IA, ces fonctions de contrôle sont surtout de nature sémantico- cognitive. Ce seront des schéma sémantiques cognitifs fondés sur une "représentation des connaissances du domaine" Et les extrants seront des " interprétations " ou des réponses signifiantes.



Ainsi, dans ce modèle, on dira que l'accès au contenu de l'information met en jeu un processus de *compréhension*. Et celle-ci sera définie comme un processus d'appariement

entre le texte et un *schéma cognitif* possédé le système. Le schéma sémantico-cognitif représentera souvent les connaissances que le système possède. C'est par ce schéma que le système sera dit "intelligent"

Ce type d'approche sera appliqué dans l'analyse des textes. On le retrouvera en intelligence artificielle classique (Schank, 1977; Johnson Laird, 1988; Sabah, 1988) en psychologie cognitive (Kintch, 1977; Frederiksen, 1975; Denhière, 1992,) en théorie catégorielle (Desclés, 1989) en analyse du discours (Lenhert, 1994; Jansen, 1992; etc.) et même en génération de texte (Danlos, 1985), en linguistique computationnelle (Fuchs, 1993; Grosz, 1979)

La supériorité de ce STI sur les approches cybernétiques et même linguistiques est incontestable. Mais, l'expérience a aussi démontré ses limites (Lenhert, 1994). En effet, cette approche suppose un approvisionnement complexe du système en connaissances de toutes sortes: grammaticales, logiques, mondaines, pragmatiques, etc. Il devra être nourri d'*ontologies* multiples ou de métacognition (Desclés, 1989). Aussi en raison de l'ampleur de cette tâche, ce type d'approche ne sera souvent appliquée et applicable qu'à des textes restreints et dont la thématique générale ne présente que des connaissances dites "ordinaires", ou éminemment techniques.

Et même lorsqu'on atteint une "représentation des connaissances" fine et bien la critique (Dreyfus, 1989, Searle 1980) soutenu que ces connaissances étaient souvent une transposition des connaissances des constructeurs du système. La compréhension que pouvait élaborer le système en regard d'un texte serait vite confrontée à celle plus profonde et riche que présentent les vrais textes et les utilisateurs réels.

" On my view the organization of world knowledge [of IA technologies] provides the largest stumbling block to AI precisely because the programmer is forced to treat the world as an object, and our know how as knowledge. " (Dreyfus 1979: 203)

En conséquence, même si ce modèle d'un STI est riche le travail préalable pour lui donner cette connaissance sera des plus lourds.

Plus profondément encore, ce modèle rencontre une difficulté similaire à celle du modèle documentaire. En effet, comme ce dernier, il est confronté à une contradiction de nature épistémologique. En effet, une des fonctions importantes d'un texte est de présenter et de transmettre à l'humain des nouvelles connaissances. Comment un utilisateur pourrait-il les découvrir dans un texte si pour ce faire, il doit les avoir mis au préalable (en partie ou en totalité) dans "représentation des connaissances" ? Autrement dit, comment peut-il construire une représentation de connaissances avant que le texte ne soit lu ?

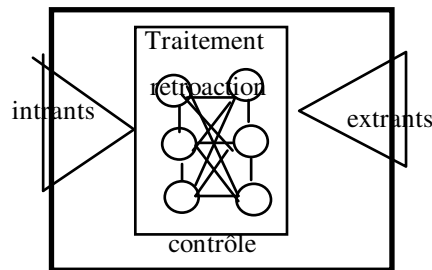
Ainsi, le modèle IA est riche mais, relativement au traitement de l'information textuelle, il met un utilisateur qui veut découvrir de l'information du système dans une position passablement contradictoire. En effet un tel utilisateur devrait avoir placée (via les constructeurs du système) dans le schéma sémantico-cognitif qui contrôle le système une bonne part de connaissances qu'il veut justement découvrir. Ce modèle sera assurément heuristique dans une situation de découverte d'information sur des petits corpus bien connus et classiques ou encore dans de la recherche d'information mais il est problématique dans une

situation de découverte d'information sur des grands corpus et surtout des corpus en constante évolution.

1. 5 -le modèle émergentiste .

Un dernier modèle de système de traitement de l'information est de type émergentiste. Ce modèle conçoit le traitement de l'information de manière radicalement différente des précédents modèles. Comme eux, il accepte aussi des mécanismes de contrôle. Ils sont aussi internes au système mais cette fois, le système les construit de lui-même. Il les fait émerger de son interaction avec l'environnement.

Face aux intrants qu'il reçoit, ce système s'adapte lentement et en vient à construire " de l'intérieur" les schéma contrôlant son traitement des signaux. Et lorsqu'il atteint un certain état de stabilité dans son apprentissage, il possède des états internes qui dirigent sa propre action. Ces systèmes s'inspirent de divers modèles formels. Les uns sont de type symbolique (Fisher, 1991) ou de type numérique: connexionisme avec supervision (les modèles connexionnistes à La Rumelhart McLelland, 1986,) ou sans supervision (à la Grossberg 1988, Kohonen 1987) ou encore génétique (Holland, 1975) markovien (Celappa 1991) etc.



On dira facilement de ces modèles qu'ils transforment véritablement des signaux intrants en des informations parce qu'ils sont en mesure de détecter et de classifier les intrants par des mécanismes autorégulation et qu'ils peuvent ainsi les rendre signifiants en regard de leur leur action (ses extrants) dans l'environnement. . . .

Si ce modèle a donné des résultats dans certains domaines particuliers : perception, reconnaissance d'images, etc. les applications dans des domaines textuels sont des plus limités. On en trouve quelques uns en psychologie cognitive (Tapiero, 1993; Miikkulainen 1990 , Croft, 1996) mais surtout en gestion documentaire (Salton et Buckley ,1994, William ,1990),en navigation hypertextuelle (Balpe et Lelu 1996, Burr 1987) . Des variantes non-dynamiques de ces approches ont par ailleurs été longuement explorées. On peut penser ici aux approches statistiques par cooccurrence, (Church 1994) par analyse multidimensionnelle (Lebart et Salem, 1994; Reinhert, 1995)

Dans cette perspective du traitement de l'information textuelle, ce type de STI sera vu comme un processus classificatoire dynamique qui cherche à faire émerger du texte des "patrons" généraux qui correspondent à des certaines régularités du textes telles, par exemple des régularités dans le lexique, dans les segments ,etc. .

On ne dira pas cependant que ces approches prétendent s'attaquer à la compréhension des textes. Il faut les voir plus comme des outils de découverte, de furetage, d'exploration mis à la disposition des utilisateurs en position de lecture et d'analyse. En cela,

ils sont plus humbles dans leur prétention. Mais de ce fait, aussi ils peuvent être plus utiles dans ce processus cognitif complexe, de découverte d'informations.

2 -La complexité de l'acte cognitif de lecture et d'analyse.

Il nous faut constater que la technologie informatique n'offre que peu de moyens pour véritablement atteindre le contenu des textes. Une des raisons profondes de ces difficultés, semble-t-il, est qu'on oublie souvent l'immense complexité de l'action cognitive impliquée dans l'accès au contenu d'un texte. Un texte est en effet un objet sémiotique qui se dévoile dans un acte interprétatif de plus complexe. De fait, l'accès au contenu de texte se fait essentiellement par deux cheminements distincts mais souvent parallèles : la lecture et l'analyse. Or, ces deux opérations ne dépendent pas de la structure matérielle du texte non plus que de l'instrument qui est utilisé pour le manipuler à savoir en l'occurrence, l'ordinateur. Elles appartiennent à la dynamique de la *compréhension* humaine.

La *lecture*, même la plus simple, est en effet un *parcours* et une *progression* où se joue le décodage progressif et heuristique d'un système sémio-linguistique complexe. (Virbel, 1993, Gervais, 1990, Stiegler 1991). Elle procède par étapes et retours, laissant lentement se révéler le contenu à saisir. Elle vise la compréhension, c'est-à-dire, l'intégration du résultat du décodage dans l'univers cognitif du lecteur [Gadamer 1976; Rastier, 1994; Beaudet et Denhière, 1992; Tapiero, 1993]. De fait, la lecture n'est pas que reconnaissance. Elle est construction d'un "objet mental" (Thérien, 1990), un modèle de situation (Kintch, 1992; Iser 1978) ou d'une problématique. Autrement dit, lire ce n'est pas simplement décoder un texte, c'est aussi construire [(Grize, 1981; Vignaux, 1988) un objet sinon même un schème mental.

L'*analyse* pour sa part, est une opération méta-cognitive. Son objet n'est pas la saisie du contenu lui-même mais l'extraction de certaines contraintes et propriétés qui régissent un texte. L'analyse déploie des stratégies de décomposition pour ne pas dire de déconstruction. Elle implique ainsi la saisie des divers constituants qui contrôlent les multiples niveaux, plans ou régimes sémiotiques du texte. Analyser un texte peut consister ainsi à en identifier le lexique, l'argumentation, la narration, la construction, les thématiques, etc. Ceci sera dépendant de modèles théoriques ou des paradigmes spécifiques. Virbel (1993) montre que ces deux opérations de lecture et d'analyse présentent un *rythme*, un *cheminement*, une focalisation, une dynamique *propre*, etc.

Autrement dit, l'accès au contenu d'un texte n'est pas algorithmique (Eco, 1985). Qu'il soit dans la lecture ou dans l'analyse, il est fait d'heuristiques et d'interprétations, de reconnaissances et d'intégration. Il existe une multiplicité de type de lecture et surtout une myriade de lecteurs. Chaque lecteur ne cherche pas la même "chose" dans un texte. Même le roman ne sera pas parcouru de la même manière dans la lecture de détente, la lecture technique, la lecture historique ou la lecture documentaire. Il est de la nature d'un texte d'être polyvalent et indéterminé dans sa signification. Une telle perspective s'inscrit dans l'herméneutique c'est-à-dire une théorie de l'interprétation des textes. A l'instar de Rastier [1994], nous dirons que cette herméneutique est de nature matérielle ou plus techniquement philologique. Elle se veut plus une démarche et une stratégie de lecture qu'une théorie métaphysique de la connaissance.

Une telle compréhension de la lecture et de l'analyse ne sont pas sans contrainte pour le design des systèmes informatiques de traitement d'informations textuelles. En effet, il devient évident que, au sens stricte, que *lire et analyser* ne sont pas des opérations qu'un ordinateur peut réaliser. Lire et analyser un texte sont des activités humaines reliées à l'adaptation et intégration de l'humain au monde. Cependant, il ne s'ensuit pas que les ordinateurs ne peuvent en rien être utiles dans ces processus. Il faut cependant les concevoir comme des outils capables de d 'assister les humains dans ces processus. Pour ce faire, il faut abandonner la voie des systèmes informatiques qui se donnent comme des robots-lecteurs au profit de celle de systèmes “adjuvants” de l'activité cognitive de la lecture humaine [Meunier, 1996], et qui laissent la maîtrise ultime du traitement de l'information entre les mains de l'expert lui-même. L'ordinateur devient alors un assistant dans la lecture, la compréhension et surtout l'analyse de texte.

Dans cette perspective un STI textuel devra sortir des ornières classiques du type autonomiste. De fait, il devra se dédoubler en deux systèmes STI: un premier, le lecteur-analyste humain qui entre en interaction avec un STI informatique adjuvant. Et le premier système est le lieu où les fonctions de contrôle et de rétroaction sont insérées et définies. Dans cette perspective, l'accès au contenu du texte est un phénomène synergétique entre un lecteur, un outil et un texte. Bref, on retrouve là le vieux schéma de l'acte philologique de lecteur ou la lecture est vue comme un aller-retour méthodique d'interprétation dynamique du texte. D'où la nécessité de construire des modèles de STI textuels qui respecte la spécificité des cette dynamique. Ce modèle ouvrirait la porte a une philologie électronique (Meunier 1996) et une herméneutique matérielle (Rastier, 1995) .

Jean Guy Meunier
Université du Québec à Montréal
1997

Références bibliographiques:

- Allen, R. B. & Riecken, M. E. (1988). *Interacting and communicating connectionist agents*. Elmsford, NY: Pergamon.
- Ashby, R. (1956). *An introduction to cybernetics*. New York: John Wiley.
- Balpe, J. P. , Lelu, A. , Papy, F. , & I, S. (1996). *Techniques avancées de l'hypertexte*. Paris : Hermes.
- Bertrand-Gastaldy, Paquin, L.C. S. ,Meunier, J.G., 1993 L'analyse de texte par opposition à la gestion des documents, ICO, vol 4, p 12-18
- Bouchaffra, D et Meunier, J. G. A Thematic Knowledge Extraction Modelling through a Rewell N & Min Tjoa Markovian Random Fields Approach, DEXA database and Expert Systems applications. (ed) p 329-339, 1996.
- Burr, D. J. (1987). "Experiments with a connectionist text reader". *IEEE First International Conference on Neural Networks*, San Diego, 717-24
- Burr, D. J. (1987). "Experiments with a connectionist text reader". *IEEE First International Conference on Neural Networks*, San Diego, 717-24
- Cellappa, R., Jain, A. Markov random fields : theory and application. Academic Press, INC. 1991.
- Church, K. , Gale, W. , Hanks, P. , & Hindle, D. (1989). "Word Associations and Typical Predicate-Argument Relations". *International Workshop on Parsing technologies*, Carnegie Mellon University, Aug. 28-31,
- Church, K. W. , & Hanks, P. (1990). "Word association norms, mutual information, and lexicography". *Computational Linguistics* 16, 22-29.
- Church, K. W., & Hanks, P. (1990). "Word association norms, mutual information, and lexicography". *Computational Linguistics* 16 , 22-29. Classification, *ACM Transactions on Information Systems* 296-333. [1,159,585 bytes]. , Classification, *ACM Transactions on Information Systems* 296-333. [1,159,585 bytes]
- Croft, W. B. L. , & Leah. (1996). " Combining Classifiers in Text Categorization in *Proceedings of the 19th International Conference on Research and Development in Information Retrieval (SIGIR '96)* Zurich, Switzerland, pp. 289-297. " . ,
- Culioli, A. (1990). *Pour une linguistique de l'énonciation, Opérations et représentations*. Paris: Orphrys.
- Danlos, L., Guez, S., & Sabbagh, S. (1985). "Conception d'un système intelligent". *Intellectica* 1,1 , 91-100.
- Denhière, G., & Baudet, S. (1992). *Lecture, compréhension de texte et science cognitive*. Paris: PUF.
- Desclées, J.-P. 1989 Intermediate Representation in the Cognitive Sciences. *Semiotica* 77(1-3), 121- 135.
- Dreyfus, H. A. and Dreyfus, L. 1988 Towards a Reconciliation of Phenomenology and AI. In *Foundational Issues in Artificial Intelligence*, D. Partridge and Y. A. Wilks (eds.), Cambridge: Cambridge University Press.
- DuPoirrier, G. (1994). *La technologie de la GED. L'édition électronique*. Paris: Hermes.
- Eco, U. 1979 *The Role of the Reader*. Bloomington: Indiana University Press.

- Fisher, D. H., & Pazzani, J. M. (1991). "Computational Models of Concept Learning". In D. Fisher H., Pazzani J. Michael, & P. Langley (Ed.), *Concept Learning: Knowledge and experience in unsupervised Learning*. San Mateo, California: Morgan Kaufman Pu.
- Frederiksen, Carl, 1975 *Semantic processing units in understanding text*. in Feedle (ed), p. 57- Gervais, B. (1990), *Récits et action. Pour une théorie de la lecture*, Montréal, Le Préambule.
- Fuch, C. (. (1993). *Linguistique et traitement automatique des langues*. Paris: Hachette.
- Gadamer, H. G. (1976) *Vérité et méthode*. Paris: Seuil; 1976.
- Gervais, B. (1990), *Recits et action. Pour une théorie de la lecture*, Montréal, Le Préambule.
- Greimas, A. J.(1983) Du sens, *Essais Sémiotiques*, Paris: Editions du Seuil.
- Grize, J. B. (1990)."Mais au fond qu'est ce que raisonner?". *Archives de Psychologie* 58, 115-122.
- Grossberg, S. (. (1988). *Neural Network and Natural Intelligence*. Cambridge: MIT Press.
- Grosz, B. (1979)."Focusing on dialog". *Computational Linguistics* 79, , 96-103.
- Habermas, J. (1989). *The Theory of Communicative Action, vol. 2*, Oxford: Beacon Press.
- Halliday, M. A .K. (1984) Introduction of Fonctionnal Grammar, Edward Arnold,
- Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, Mi.: University of Michigan Press.
- Imbert, M. (1987)."Neurosciences et sciences cognitives". *Le Débat* 47 (Paris,Gallimard),
- Iser, W. (1976). *The Art of Reading A Theory of Esthetic response*, Baltimore 1976: John Hopkins University.
- Jansen, S., Olesen, J., Prebensen , H., & Tharne, T. (1992). *Computational approaches to text Understanding*. Copenhagen: Museum Tuscalanum Press,
- Johnson-Laird, P. N. (1988), *The Computer and the Mind*, Harvard University Press.
- Kamp, H., & Reyle, U. (1993). *From Discourse to logic*. Dordrecht.
- Kintsch, W., *Memory and Cognition*. New York: Wiley, 1977
- Kohonen, T. (1987). *Self Organization and Assoicative Memory*. Berlin: Springer Verlag.
- Lebart, S Salem. S. A. (1994). *Statistique textuelle*. Paris: Dunod.
- Lehnert, W. G(Riloff, E. and 1994. , Information Extraction as a Basis for High-Precision Text Classification, *ACM Transactions on Information Systems* 296-333. [1, 159, 585 bytes]
- Lehnert, W. G., & Riloff, E. (1994).Information Extraction as a Basis for High-Precision Text
- Lelu, A. (1995). "Hypertextes: la voie de l'analyse des donn_es". In L. Bolasco. . S L, A. Salem (Ed.), *Anilisi statistica dei dati testuali vol2*. (pp. 85-96). Rome: CISU.
- Lemoigne, J. L. (1990). *La théorie du système général théorie de la modélisation*. Paris: PUF.
- Meunier, J. G. (1996). "La théorie cognitive: son impact sur le traitement de l'information textuelle". In V. Rialle et Fisetete, D (Ed.), *Penser l'Esprit, Des sciences de la cognition ^ une philosophie cognitive*. (pp. 289-305). Grenoble: Presses de L'Université de Grenoble.

- Meunier, J. G. (1992). SATO: un philologue électronique. *Documentation et Bibliothèque*, Montréal, (avril-juin), 65-69.
- Miikkulainen, R. (1993). *Subsymbolic Natural Language Processing. An integrated Model of Scripts, Lexicon, and Memory*. Cambridge, Mass: MIT Press.
- Rastier, F. (1987). *Sémantique interprétative* Paris: PUF.
- Rastier, François. *Sémantique pour l'analyse : de la linguistique à l'informatique*. Paris, Masson, 1994, 240
- Reinheirt, M. (1995) Quelques aspects du choix des unités d'analyse et de leur contrTMle dans la méthode "Alceste". In L. L. S.Bolasco A Salem (Ed.), *Analalisi Statistica dei Dati Testuali*. vol 1, (pp. 19-27). Rome: CISU.
- Rumelhart , D. E., & Mc Clelland , J. (1987). *Parallel Distributed Processing :Explorations in the Micro Sctucture of Cognition*, 2 vol. Cambridge: MIT Press.
- Salton, G., Allan, J., & Buckley, C. (1994)."Automatic Stucturing and Retrieval of large Texte File". *Communications of the ACM* 37 (2), 97-107.
- Salton G., & Mc Gill, M. (1983). *Introduction to models of Information Retrieval*., New York: Mc Graw Hill.
- Salton, G. (1989). *Automatic Text Processing*. Addison Wesley:
- Schank.R.,Abelson R.,*Scripts, Plans, Goals, and Knowledge Understanding* , LawrenceErlbaum, Hillsdale.N.J. 1977.
- Searle, J., (1980) " Minds, Brains and Programs"The Behavioral and Brain Sciences, 3: 417-547.
- Seffah Meunier 1995 ALADIN :Un Atelier Génie Logiciel Orienté Objets Pour La Lecture & l'Analyse Intelligente de Textes. JADT, Rome 1995
- Spark Jones, & Martin, K. (1973). *Linguistics and Information Science*. New York: Academic Press.
- Stiegler, B. (1991). "La lecture assistér par ordinateur". In P. L. P. Chambat (Ed.), *Les nouveaux outils du savoir* ,. Paris.: Descartes.
- Tapiero, I. (1993). *Traitement cognitif du texte narratif et expositif et connexionnisme: expérimentations et simulations*. Thèse Université de Paris VIII,
- Thérien, G. (1990), "Pour une sémiotique de la lecture", *Protée*, vol. 18, n^o 2, pp. 67-81.33-45.
- Veronnis, J. , Ide, N. M. , & Harie, S. (1990). "Utilisation de grands réseaux de neurones comme modèles de représentatons sémantiques". *Neuronimes*,
- Vignaux, G. (1988), *Le Discours acteur du monde: énonciation, argumentation et cognition*, Paris, Orphrys,
- Virbel, J. (1987). "L'apport de conniasances linguistiques à l'interprétation des structures textuelles". *Structure des documents, Bigre++Globule* 53, 77-97.
- Virbel, J. (1993). "Reading and Managing Texts on the Bibliothèque de France Station". In P. Delany, & P. Landow (Ed.), *The Digital Word: Text Based Computing in the Humanities*. Cambridge, Mass: MIT Press.
- Virbel, J. E. , F. Pascual, E. (1992). *La lecture assisfée par ordinateur, Rapport de recherche*. Toulouse: Laboratoire IRIT.

