

textualité

>>> categorie = papersreview

>>> keywords = | navigations | finder | balisages |



La question de la mémoire extériorisée que constituent les supports de savoirs, qui ne sont pas de simples réceptacles d'une pensée constituée hors d'eux et avant eux, mais qui sont la surface même de cette pensée "dialoguant avec elle-même", par l'objectivation qui se projette sur et est réfléchi par eux, qui fait d'eux ce qu'Artaud appelait un "subjectile", cette question est aussi celle de l'accès à cette mémoire accumulée, et la "navigation" dans les savoirs qu'elle contient, et pour laquelle des techniques d'assistance à l'orientation sont indispensables.

On a pu comparer l'événement de la Bibliothèque d'Alexandrie à la réalité géopolitique émergente d'Internet. Si une telle comparaison est sans doute osée, pour ne pas dire risquée, elle fait aussi tout le sens de la trajectoire qu'a tracée Christian Jacob dans le programme des métamorphoses du lecteur.

Et c'est en ce sens que je voudrais insister sur le fait que l'un des éléments techniques d'émergence de l'Internet actuel, à savoir la norme HTML, est issu d'un métalangage de formalisation et en ce sens d'annotation des textes : SGML, dont HTML n'est qu'une version enrichie de la capacité à gérer et à formaliser des liens hypertextuels.

Pour autant, je voudrais montrer pourquoi il me semble qu'il n'y a pas encore de base théorique herméneutique et philologique disponible aujourd'hui qui réponde à l'extrême singularité de la mémoire électronique en réseau, et qui en réponde tout en sachant véritablement rendre compte de ce que ce nouveau contexte a de profondément commun avec les plus anciens, et notamment celui d'Alexandrie.

C'est pour esquisser un mouvement vers une tentative qui irait en ce sens que je proposerai les concepts de "sociétés d'auteurs" et de "sémantique située".

L'annotation électronique. SGML (standardized generalized markup language) est un métalangage de formalisation des documents structurés qui décrit des formes inscrites dans un espace matériel : à ce titre, il est intéressant de lui appliquer le concept de "mise en forme matérielle des énoncés" qu'a élaboré Jacques Virbel pour mettre en évidence une véritable performativité de l'écrit dans sa matérialité, et qui montre que la sémiotique typodispositionnelle propre aux énoncés écrits engendre un sens qui n'est pas contenu dans l'énoncé oral. La performativité mise au jour par Austin montre que dans le cas de l'énoncé performatif, "dire c'est faire", avec le concept de Virbel apparaît une performativité inversée : faire c'est dire, un geste d'inscription matérielle produit un sens.

J.L. Lebrave avait mis en évidence le rôle qu'a joué l'apparition du bon à tirer avec la généralisation de l'imprimerie, dans les

rapports au texte de l'écrivain et du lecteur, accentuant notamment la séparation de l'écriture et de la lecture, le recul du manu-script engendrant une régression du rôle de la main . Parchemin manuscrit ou papier imprimé sont des supports statiques. Or, aujourd'hui, avec la numérisation, le texte connaît une nouvelle époque : celle des supports dynamiques où la lecture rejoint "naturellement" l'écriture.

Et c'est ce que signifie SGML du point de vue que j'adopte ici : si cette norme était d'abord simplement appréhendée comme un format d'échange, elle est devenue un instrument de navigation automatique, avec la possibilité de générer des tables de matières au moyen de balises, etc. Mais surtout, dans le cas qui m'intéresse, son usage a pu être étendu vers les langages d'annotation dans une expérience à laquelle Jacques Virbel, Christian Jacob et moi-même avons participé avec la Bibliothèque Nationale de France, il y a neuf ans : la conception de postes de lecture assistée par ordinateur.

L'annotation électronique, qui exploite les vertus dynamiques des supports numériques, permet d'inscrire les actions du lecteur dans ce qu'il lit. Ce qui signifie qu'ici, lire et écrire deviennent proprement inséparables. Lire en inscrivant sa lecture à même le texte lu est typiquement ce que fait un lecteur "savant". Je ne pense pas seulement ici à un écrivain ou à un universitaire, mais aussi à un ingénieur, un documentaliste, un médecin, etc.

Il n'est cependant pas "naturel" d'écrire sur un livre ou un dossier. Le livre est certes très souvent annoté, mais le support imprimé n'est pas fait pour cela, la rupture du bon à tirer consacre une séparation entre ce qui est annoté et les notes, même si elle est toujours transgressée.

Tout se passe alors comme si les supports dynamiques et ce que je nomme l'hypertraitement de texte (HTT) conduisaient à un retour à la situation pré-gutembérienne des scriptoria. Dans une base de données engendrée par un HTT, les actions antérieures du lecteur (ces événements de lecture que sont les événements "souris" ou clavier) sont intégrées à la fois comme données et comme instruments de navigation. Ce qui permet de résoudre le problème du Petit Poucet que posent toutes les bases de données numériques (et dynamiques) aux utilisateurs, à savoir leur caractère virtuel et les problèmes de navigation que cela pose. Le Petit Poucet est perdu dans la forêt (il n'en a aucune appréhension cardinale) tant qu'il n'a pas compris qu'il lui faut y laisser ses propres traces pour pouvoir s'y orienter. Mais s'il abandonne sur son chemin des miettes de pain, les oiseaux les mangent et cela ne va pas beaucoup mieux. Seuls les cailloux blancs, visibles dans la nuit claire et inconsommables par le milieu animal de la forêt, lui permettent de résoudre le problème. L'hypertraitement de texte repose sur un principe comparable.

Naviguer est cependant déjà un problème dans l'espace du lecteur traditionnel : il marque ses passages dans un texte, les "passages" du ou des textes qu'il lit, par où il passe, au moyen de diverses techniques traditionnelles : de la croix au fichier, etc. Si l'on veut approcher le travail de l'écrivain sur ses supports traditionnels, il faut fréquenter aussi ses outils et lieux d'étude : stylo, papier, carnet, chemises, dossiers, rayonnages où ils s'accumulent avec les ouvrages marqués, cornés, annotés, tiroirs où sont rangés des fichiers, et tant d'autres formes d'aide-mémoire composant les supports des codes de "métascriptions" en quoi consistent les miettes de pain qu'il sème sur son chemin, et qui forment un processus global d'archi-annotation, auquel aujourd'hui s'ajoutent évidemment les disquettes, mais encore les photos, bandes magnétiques audio

et vidéo et "échantillonnages" numériques que les extensions multimédia des systèmes micro-informatiques font à leur tour entrer dans le champ d'une annotation électronique généralisée au sein de dispositifs dits hypermédias. Mais je dis miettes de pain dans cette mesure où le lecteur traditionnel perd ses propres marques à mesure qu'il les sème, du fait de la finitude de sa mémoire (de ses capacités de rétention), et du caractère statique des supports qu'il emploie pour y suppléer. Bien avant l'hypertexte et l'électronique, tout un espace rétentionnel et virtuel, statique et cependant opératoire jusqu'à un certain point, encadre physiquement l'écrivain travaillant à sa table. De subtiles techniques d'annotation, corrélation et classement, organisent dès son origine l'écriture des œuvres, qui peuvent être aujourd'hui transposées dans le domaine des supports numériques au plus grand bénéfice du lecteur-scripteur : elles s'en trouvent incommensurablement dynamisées.

Qu'il soit écrivain, professeur, étudiant, "intellectuel", mais aussi ingénieur, avocat, journaliste ou administrateur, un lecteur "professionnel" utilise diverses techniques de balisage et d'orientation que matérialisent des "habitudes graphiques" et/ou spatiales, qu'il en systématise ou non l'usage : signes d'annotation de significations diverses (croix et traits en marges, soulignements aux tracés divers dans le corps même du texte imprimé, etc.), techniques de résumé et de synthèse, fichiers, dossiers, etc. Il glose les textes, les indexe, les met en rapport par des systèmes de corrélation (les fichiers), y extrait des passages pour des citations et utilise des instruments de recherche (bibliographies, revues spécialisées, dictionnaires, encyclopédies).

Toutes ces techniques de lecture visent à créer des liens qualifiés entre des documents ou des passages dans des documents. Les techniques hypertextuelles, combinées avec les langages de description de documents structurés, permettent d'intégrer rigoureusement, en les automatisant, ces opérations. Il en résulte des langages d'annotation électronique, ou hypertraitement de texte. L'avantage en est que la mémoire de la machine n'oublie pas, tandis que celle du lecteur est essentiellement faillible. Une fois les cahiers, livres, fichiers, dossiers refermés dans les rayons de la bibliothèque ou sur le bureau, toutes les interventions sur le support même du corpus en quoi consistent les diverses gloses, de la croix en marge au commentaire quasiment rédigé sur le cahier, se disséminent dans l'espace de travail, s'atomisent et s'ignorent superbement. Le génie du lecteur sera d'en faire malgré tout la synthèse. Le génie de la machine permet sinon de vérifier ce génie, encore moins de le remplacer, en tout cas de lui assurer les plus rigoureuses conditions d'exercice : pour le lecteur traditionnel, la visibilité exacte du texte ne porte que sur quelques pages ou dizaines de pages en amont et en aval du passage actuellement lu. Au-delà de ce champ, la fidélité du lecteur à son corpus est irrémédiablement livrée aux infortunes de sa subjectivité. L'assistance de la machine inaugure au contraire une époque de haute fidélité de la lecture. La lecture de la machine est sans faille, sans délais. Sa visibilité du texte est totale et instantanée. Cela ne veut évidemment pas dire que la lecture devient objective - mais qu'elle gagne en rigueur et en lucidité sur sa propre opération.

L'espace rétentionnel statique est virtuel et très difficilement actualisable, si ce n'est par une orientation physique qui s'appuie sur une perception intuitive de l'espace. Dans l'espace dynamique, la virtualité est à la fois plus affirmée et manifeste

(parce qu'il n'y a plus d'appréhension immédiate et intuitive), et plus actualisable (parce que je peux confier à la machine des instructions d'actualisation : mémoire toujours ouverte, etc.). Mais cela suppose une très rigoureuse organisation de l'orientation et donc de l'annotation : on connaît le défaut des hypertextes, la création inflationniste de liens où celui-là même qui les a créés ne s'y retrouve plus ; il s'agit autrement dit d'inventer une géographie (c'est-à-dire un art de la cartographie) d'un nouveau type, ou, pour parler comme Jacques Virbel, de nouvelles normes de mises en forme matérielles.

On peut analyser, modéliser, automatiser et intégrer sous formes de fonctionnalités primitives d'annotation numérique les interventions manuscrites sur supports imprimés qui correspondent à l'ordinaire d'un travail de lecture amorçant un processus d'écriture (et il peut porter sur un texte dont le lecteur est lui-même l'auteur).

Dire que la lecture (d'étude ou savante) est une opération d'inscription, signifie qu'elle consiste d'abord en une appropriation du texte par intervention directe sur le support statique du papier. Dynamique, le support numérique permet une démultiplication et une exploitation automatique et systématique des opérations d'annotation modélisées selon les techniques de l'informatique.

Les actes d'annotation engendrent rapidement, au-delà des seuls repères ou codes graphiques, de l'écriture à proprement parler : notes en marges, mots-clés, commentaires. Les mots-clés créent par ailleurs des liens, des corrélations. Ils donnent également lieu à des listes ou thésaurii, qui sont autant de systèmes de "navigation" dans la mémoire bibliothétique du lecteur. Un système de lecture assistée par ordinateur peut alors reproduire ces techniques traditionnelles en les systématisant et en les intégrant. Il en résulte de nouvelles possibilités instrumentales d'orientation, par combinaisons et extrapolations, dont j'exposerai plus loin quelques aspects.

En première approche, on peut distinguer deux grandes classes d'interventions sur le texte : . Des opérations de hiérarchisations, qui pondèrent le poids de passages du texte, auxquelles correspondent sur le support de papier les soulignements dans le corps du texte et les traits verticaux en marge ;

. Des opérations de qualifications, qui consistent à attribuer des valeurs sémantiques à ces pondérations, par divers moyens : insertions de mots-clés; insertions de notes personnelles; rédactions de commentaires; liens avec d'autres documents (autres passages du même texte ou autres textes, par exemple : sources manuscrites, variantes éditoriales, textes mentionnés en référence, traductions, gloses canoniques, références bibliographiques, etc.).

Ces interventions ont été modélisées dans une maquette informatique, LECAO, réalisée à l'Université de Compiègne , où les opérations de hiérarchisations consistent à utiliser des caractères en couleurs, tandis que les opérations de qualification aboutissent soit à des opérations de surlignages en couleurs du texte commenté, soit à la création de liens entre documents visibles par l'ouverture d'une fenêtre des "textes en rapport" avec le document commenté, ou par l'apposition de mots-clés. Bien entendu, chaque intervention sur le texte est enregistrée par la machine, qui peut alors la traiter comme une information, ce qui rend possible par exemple des recherches portant uniquement sur un niveau donné de la hiérarchisation, et plus généralement, la combinaison de très nombreux critères s'appliquant aussi bien au texte lu qu'aux qualifications textuelles engendrées par le lecteur.

Les inscriptions en marge ou dans le corps des textes lus, les cahiers de notes, les fichiers, les dossiers et leur organisation physique dans les rayonnages de la bibliothèque privée et sur le bureau, constituent autant de systèmes personnels d'orientation et de navigation dans l'espace à la fois matériel et spirituel (temporel et virtuel) du travail. Toute une mise en forme matérielle de la textualité et de sa lecture est ainsi opérée non seulement sur l'espace à deux dimensions du support de papier, mais aussi dans le volume tridimensionnel du bureau et de la bibliothèque.

Il résulte du transfert de cette mise en forme matérielle de la lecture-écriture sur les supports dynamiques de l'hypertraitement de texte une trans-formation majeure de l'accès, non seulement au texte, mais à la lecture passée de ce texte, telle qu'elle se matérialise elle-même textuellement, à la lettre, à travers toute la gamme des interventions décrites ici sommairement.

Le PLAO, dont les spécifications sont beaucoup plus riches que le point de vue général exprimé ici, a été réalisé sous forme prototypique par la société AIS (Advanced Information Systems) du groupe Berger-Levrault, sous la direction de François Chahuneau. Dans le système AIS, il n'y a pas de catégorisation a priori des systèmes d'annotation possibles : les outils de catégorisation sont très largement modulables (c'est la même logique que SGML), à l'exception évidemment des fenêtres de requêtes, des formes tolérées par le thesaurus, etc., ce qui donne une très grande généricité à l'instrument.

Mais, cette généricité présente aussi des inconvénients, et surtout pose un problème éditorial fort intéressant - propre aux supports dynamiques : le lecteur non-familier éprouve d'abord une difficulté certaine à appréhender la logique d'utilisation, ou la méthodologie du prototype, qui tient précisément à son ouverture : AIS a adopté le principe que l'utilisateur devait pouvoir définir lui-même les caractéristiques de ses types d'annotation, jalons, etc. Tout ce que François Chahuneau nomme unités logiques de découpage (ULD).

Une telle ouverture est un très grand avantage, qui n'est pratiquement jamais offert dans le domaine de la micro-informatique (si ce n'est de manière très limitée, comme avec Word, par exemple, qui permet de nommer et de définir des styles, tels que citation, note de bas de page, etc.). Mais cet avantage est aussi un grave inconvénient si l'utilisateur n'est pas clairement averti : et de la liberté d'intervention qu'une telle conception concrétise; et des grandes classes de possibilités qui s'offrent à lui.

Ainsi, la possibilité d'attacher un mot-clé est un attribut possible pour toute ULD, or l'utilisateur peut ne pas comprendre l'avantage considérable que cela présente. Il définira par exemple une ULD "mot-clé", sur laquelle il fera ensuite ses recherches, alors qu'une telle démarche n'est pas économique (il est beaucoup plus rationnel d'attacher des mots-clés à une unité de sens déjà constituée comme ULD que de définir une ULD spécifique pour saisir des mots-clés).

Autrement dit, des chaînes opératoires modèles doivent être proposées en standard, à l'utilisateur; elles ne doivent pas l'inciter à demeurer enfermé dans des rails, ce qui serait contraire à l'ouverture du système et à toutes les analyses qui ont été faites antérieurement par le groupe de travail de la BDF, prouvant que la maquette doit être modulable selon les besoins express des utilisateurs; mais ces modèles doivent lui donner des illustrations efficaces de la manière de construire des chaînes

opératoires personnelles en exploitant au mieux les concepts fondateurs du système (ULD, jalons, annotations, liens). Cela donne immédiatement la définition d'un travail qui reste à accomplir : il est indispensable de formaliser de grandes classes de procédures d'approche des textes, qui soient autant d'appropriations des éléments combinables offerts par l'interface, en fonction des grandes tâches qui caractérisent les modalités de scrutation des textes par lecture et écriture, propres aux différentes disciplines lectrices.

Ces catégorisations seraient autant de modèles à la fois scientifiques et éditoriaux. Langages à la fois d'annotation et de navigation : l'inscription de l'acte de lecture dans le support implique une normalisation ouverte et générique des modalités d'annotation, aussi bien pour ce qui concerne la rigueur de l'emploi de primitives, que pour les chaînes d'opérations (effaçant par là même le flou des annotations manuscrites, les aspects excessivement empiriques de leur usage), pour permettre l'échange entre stations et systèmes différents, et pour normaliser les modalités de navigation : on ne peut en effet ici distinguer le travail d'annotation effectué par le lecteur du système par lequel il donne accès au corpus. Autrement, c'est la cumulativité du savoir ainsi conservé et élaboré sur les supports dynamiques qui est en jeu.

Ces diverses formalisations peuvent être déclinées à divers niveaux :

- . Un fonds de chaînes opératoires commun à toutes disciplines lectrices.
- . Un fonds de chaînes opératoires commun à chaque discipline.
- . Dans chaque discipline, des modalités spécifiques de chaînes opératoires.
- . Des chaînes opératoires qui relèvent des idiosyncrasies de chaque lecteur en tant qu'écrivain - ce niveau de formalisation lui incombant alors pleinement.

Chaque lecteur a diverses manières d'aborder un texte. Cela tient à plusieurs éléments, dont en particulier :

- . Son corpus (la nature de ce corpus, de son accessibilité, etc.)
- . Sa formation, depuis l'école primaire jusqu'à l'université et au laboratoire,
- . La discipline dans laquelle il exerce son savoir,
- . Ses propres concepts, largement dépendants de ses méthodes de lecture, et inversement (bien que le plus souvent, le lecteur n'ait qu'une conscience extrêmement vague de cette dimension "matérielle" de son travail spirituel, conceptuel, ou "idéal").

La difficulté est alors de distinguer entre ces niveaux. Dans l'univers de la maquette AIS, l'outil contraint de lui-même à la distinction entre ces niveaux par certains aspects. Le fait de raisonner en termes de balisages et de liens, de représentations synthétiques d'un contenu, et de lecture-écriture, ce qui consiste à introduire de la sémantique personnelle du lecteur dans la sémantique du texte lu, détermine déjà certaines contraintes de distinctions.

Ces remarques nécessitent d'explicitier des considérations sur la maquette AIS à portée plus générale. Il faut distinguer entre ce qui concerne la structuration du texte, et la structuration de la grille de lecture qui devient elle-même un texte. De façon générale, plusieurs possibilités sont ouvertes pour aborder un texte, et elles correspondent 1) à des tâches différentes dans un cas, 2) à des alternatives pour une même tâche dans un autre cas. En voici deux exemples très simples :

1) Deux tâches différentes :

- . Faire une lecture approfondie d'une œuvre, pour en livrer une

interprétation globale. Cela peut se faire par appropriation systématique et réduction du contenu, cela consiste la plupart du temps et en premier lieu en une lecture séquentielle.

. Faire une lecture d'une œuvre au service d'une autre lecture d'une autre œuvre, ou au service d'une idée qui ne se constitue pas en premier lieu autour de l'œuvre lue, et par là, mobiliser des éléments de l'œuvre sans prétendre en faire une lecture globale. Il ne s'agit pas alors nécessairement d'opérer des réductions.

2) Une alternative entre plusieurs abords pour une même tâche.

. On peut imaginer qu'une œuvre qui a déjà été lue dans son ensemble, traditionnellement, linéairement, fasse ensuite l'objet d'un travail de LAO qui ne consistera pas à re-commenter séquentiellement toute l'œuvre, tout en demeurant dans une certaine séquentialité utilisant les moyens de l'annotation électronique (éventuellement en reportant des annotations déjà effectuées sur le support papier).

. Au contraire, on peut partir, dans le même cas initial (une œuvre qui a déjà été lue dans son ensemble), par exemple d'un dictionnaire constitué à partir du corpus numérisé, puis on ira y rechercher des éléments ponctuels en faisant plus confiance à la vision d'ensemble que l'on a acquise dans la lecture séquentielle traditionnelle qu'à une reformalisation informatique de la même opération.

. Etc. : il y a bien d'autres possibilités encore.

Il est évident, d'autre part, que ces diverses possibilités sont à reconsidérer selon que le corpus est en mode-image ou mode-texte. Je m'en tiens ici au mode-texte. Il s'agit de déterminer des standards de balisage et d'outils d'annotations, qui caractériseraient des méthodes pour approcher les corpus et les structurer. Jacques Virbel propose de désigner ces standards par l'expression description formalisée de structures (DFS). La question est de dégager des ensembles de DFS au niveau de l'utilisateur, sachant que cela concerne trois niveaux.

1. Il y a d'abord les DFS sur lesquelles le lecteur s'appuie, et telles qu'elles ont été produites par le fonds qu'il utilise. Exemple : les balisages BDF, soit en mode-image, soit en mode-texte.

Si elles n'ont pas été produites par ce fonds, il faut alors lui proposer des méthodes standards, ou normes, de production de telles balises par lui-même, sachant qu'il peut toujours modifier ces normes lui-même, étant donné l'ouverture de l'interface. Ce premier niveau est la description physique de la mise en forme matérielle (MFM) du corpus, en quelque sorte de sa géographie textuelle.

2. Le second niveau est le découpage syntaxique qui va permettre à l'utilisateur de se construire une grille de lecture, et de la formaliser pour lui-même, d'en faire donc un objet d'interrogation dynamique, étant donné que cette grille devient rapidement elle-même un objet hypertextuel complexe. Cette syntaxe est ce qui commande la conception des ULD par l'utilisateur, dans leurs rapports aux mots-clés, annotations, jalons et liens.

Dans le cas d'une lecture exhaustive d'une œuvre en vue d'une interprétation globale, cas auquel nous nous restreindrons ici, un moment essentiel et conditionnant tous les autres est le découpage du texte en ULD constituant des unités de sens (US). On peut imaginer que ce découpage soit automatisé, s'appuyant sur des éléments donnés par les DFS de la MFM, et mettant en œuvre la fonction DAZ (Découpage Automatique de Zones). Mais ce n'est pas toujours évident. Cela peut constituer une première phase, de dégrossissage, qui doit être ensuite affinée par une

lecture séquentielle. Une US peut rassembler plusieurs éléments de DFS/MFM, tout comme elle peut être plus granulaire que toute unité de ce niveau.

La syntaxe qui doit être ici élaborée est la MFM des pertinences - c'est-à-dire des différences, si l'on nomme pertinente une différence dans le sens de Troubetskoï, et j'indique cette référence moins comme argument théorique structuraliste que comme exemple - par l'utilisateur. Cette syntaxe est déjà la marque de l'utilisateur, elle concrétise par une mise en forme matérielle sa grille de lecture, et cependant, elle est encore enracinée dans la géographie du texte, ou dans son histologie, à la fois horizontale et verticale : il faut rendre compte à la fois de l'extension horizontale d'une unité de sens pertinente, par exemple tel alinéa de tel paragraphe dans tel chapitre de telle partie de tel ouvrage de tel auteur dans telle discipline, et de sa "hauteur", de son "relief", c'est-à-dire aussi de son poids qui relève d'une certaine verticalité : ce que j'ai nommé la hiérarchisation relève de cette dimension.

Cette syntaxe, il faut le souligner, permet déjà une réduction synthétique d'un texte dans la verticalité : si l'utilisateur a délimité des ULD de passages "très importants", leur seule sélection permet de donner un résumé, à la manière de la collection SUP Grands textes des PUF.

3. Le troisième niveau est celui qui permet :

- . Une élaboration sémantique chargeant de sens la "syntaxe" des DFS des deux premiers niveaux, et qui rende possible une réduction linéaire, par exemple par une sortie automatique de commentaire dans la séquentialité du texte commenté;
- . Une navigation dans ces éléments qui ne soit pas linéaire, puisqu'elle crée des corrélations entre eux permettant de sauter de saillances en saillances, et par exemple de rassembler toutes les ULD afférant à un concept du lecteur ou du texte commenté (et éventuellement de formaliser ces parcours par sauts en chemins de lecture) : dans le premier cas, il s'agit de rechercher des ULD spécifiées par un mot-clé (MC) de l'utilisateur (y compris si ce MC est présent dans le texte : le seul fait qu'il ait été sélectionné comme MC implique qu'il n'est plus un mot de l'auteur, mais aussi du lecteur, ou un lieu commun - au sens d'Aristote - des deux);
- . Une représentation synthétique, à divers niveaux, une cartographie autrement dit, mais aussi une radiographie, en quelque sorte, de la sémantique textuelle elle-même, que ce soit par nuages de points, graphes, ou plus simplement, listes de thesaurii, dictionnaires divers, etc. ;
- . Tout ceci vaut pour toutes les productions textuelles du lecteur, qui peut ainsi gérer ses propres notes comme il gère le texte annoté.

C'est aussi bien la génération d'un nouveau type de tables de matières (TDM), l'une qui relève de l'auteur, l'autre du lecteur, les deux pouvant et même devant se confondre, mais pouvant et même devant être aussi distinguées.

Ce troisième niveau est celui des historiques de la lecture effective, constituée en objet hypertextuel qui est la grille de lecture dynamique.

Dès lors, on doit distinguer quatre grandes catégories d'opérations, qui sont des fonctionnalités de base de l'annotation électronique, et qui permettent par leurs combinaisons de réaliser effectivement ces trois niveaux :

- . Hiérarchisation : à la fois par la création d'US et par l'implantation d'ULD de mise en relief (horizontalité et verticalité).

. Qualification : MC, annotations, commentaires, etc.
. Navigation et recherche : création des liens et des protocoles de corrélation nécessaires : définition, par exemple, de catégories de liens spécifiant la nature de documents attachés : commentaires canoniques, traductions, manuscrits, références mobilisées par le texte commenté, etc.; fenêtres de recherche.
. Représentation (assistance à la navigation et à la recherche) : graphes divers, y compris les TDM, "vues" sur les corpus (y compris la structuration du finder - fonction qui manque ou qui reste pauvre dans la maquette AIS), mais telles qu'elles articulent syntaxes et sémantiques.
Ces formalisations sont autant d'hypothèses pour la description formalisée de documents structurés en tant que documents spécifiquement électroniques. Autrement dit, cela concerne une standardisation des formats d'édition électronique. Celle-ci est indispensable : il serait en effet inconcevable que les travaux d'annotation produits par de grands lecteurs des œuvres ne soient pas largement accessibles et publiables, l'annotation devenant ici partie intégrante de l'œuvre critique et même de l'œuvre tout court, en tant qu'elle rend accessible son propre travail de préparation. C'est une ère nouvelle qui s'ouvrirait ainsi pour l'élaboration et la transmission des savoirs.

On peut aller très loin dans l'hypothèse selon laquelle un système idiosyncrasique d'annotation reflète un parti théoriquement pris sur le corpus concerné : à chaque type de lecture correspondrait une technicité d'annotation et de navigation singulière. La question serait alors la possibilité d'une normalisation qui serait générique, c'est-à-dire qui permettrait à la fois le partage et l'échange des savoirs, ainsi que l'émergence des singularités.
En fait, c'est une question comparable à celle des thesaurii : certains font référence, d'autres non. Des écoles se constituent, avec des variantes, etc. Il faut à la fois de la norme et de la variabilité. Concevoir cette question des standards dans le cas des supports dynamiques ne saurait être la même chose que dans le cas des supports statiques : l'adaptabilité au lecteur, engendrée par la dynamique numérique, renouvelle entièrement la question éditoriale. En conséquence, il s'agit de définir des formats d'échange à la fois techniques et intellectuels, de déterminer ce qui relève strictement de la norme en termes de balisage et ce qui doit rester ouvert à la variabilité, tout en proposant des modèles dominants pour tous les types d'annotations et de structurations.

La sémantique située. La virtualité de la connaissance extériorisée (située), quel que soit son support, pose des problèmes de cardinalité, d'orientation, par et pour le lecteur, dans l'organisation des connaissances concrétisée par la structure physique des supports. Une connaissance extériorisée et conservée sur un support statique tel que le papier est spontanément appréhendée, lorsqu'elle est rangée sur des rayonnages, classée dans une bibliothèque, comme espace actuel d'organisation. À cet égard, elle est "moins virtuelle" qu'une mémoire numérisée, dont l'organisation devient imperceptible en tant que telle pour l'utilisateur, et pour laquelle il faut donc construire un système de représentation qui constitue un protocole de navigation.

Mais l'expérience montre que :

1. La mémoire écrite sur le papier demeure virtuelle au sens où elle reste cachée dans sa masse à l'utilisateur, bien que celui-ci dispose d'une sorte de "cardinalité" spontanée pour s'y orienter, en feuilletant les livres, en comprenant l'organisation des

rayonnages, en prenant une vue d'ensemble du corpus,
2. La mémoire stockée sur les supports dynamiques ne donne jamais une telle appréhension globale, et les systèmes de navigation traditionnels ne permettent pas vraiment la construction d'une cardinalité.

L'avantage de la numérisation consiste en ce que le support, actif, permet de déléguer des fonctions de recherche assumées par l'utilisateur vers le système numérique lui-même. Mais les requêtes qui concrétisent cette délégation sont contraintes par le système de représentation de l'espace virtuel, soumises à la logique de navigation qui n'est pas le fait de l'utilisateur lui-même.

Nous faisons l'hypothèse que le système de navigation n'est optimal que lorsqu'il est construit par l'utilisateur lui-même, ou par la communauté des utilisateurs, qui devient une société d'auteurs lorsque la construction du système de navigation est assurée par l'usage collectif d'un dispositif d'annotation électronique de documents.

En effet, selon ce que nous avons nommé la logique du Petit Poucet, un espace virtuel est comme une forêt dans la nuit, où l'orientation peut être construite par celui qui y erre à la condition qu'il puisse y laisser des repères de ses passages antérieurs, et constituer ainsi par essais, erreurs et mémoire de ces essais et erreurs une carte de l'espace parcouru. Il peut évidemment aussi retrouver et emprunter dans ses propres parcours les parcours d'autres errants, passés avant lui, et bénéficier de leur expérience sur un mode secondaire.

Une base de données dans laquelle un utilisateur laisse ses traces est beaucoup plus rapidement appropriée par celui-ci. L'annotation des documents constituant le corpus est ce qui organise leur appropriation. Ici, le lecteur est aussi scripteur. Et la sémantique est un processus et se construit au cours du temps.

De fait, une base de données accessible par un système d'annotation électronique est ce qui permet de
. Construire par émergence (au cours de l'interprétation de la base en quoi consistent les parcours qui y sont effectués) un système de navigation,
. Construire ce système de navigation comme une sémantique hypertextuelle du domaine.

Cette sémantique peut être une construction collective, et le système d'annotation est alors coopératif. Dans la mesure où elle est constitutivement un système d'orientation par construction d'une cardinalité, nous l'appellerons une sémantique située dont il résulte ce que les catalographes et bibliothéconomes appellent des listes d'autorités - et ce que la tradition philologique plurimillénaire a conçu comme thésaurii.

La sémantique située est un système de navigation dans des connaissances situées, produit par les utilisateurs de celle-ci. Pour reprendre le langage de Jacques Virbel, c'est une sémantique "typo-dispositionnelle".

On peut parler de sémantiques "typo-dispositionnelles" dans la mesure où un énoncé écrit, par son organisation physique sur un support, engendre du sens qui ne peut être réduit au sens du même énoncé sous sa forme orale. La conception d'un système de navigation par annotation électronique est ce qui devra formaliser des sémantiques typo-dispositionnelles primitives, et en exploiter les possibilités nouvelles, offertes par le caractère dynamique du support.

Aujourd'hui, la gestion industrielle des savoirs, qui sont devenus la source principale de profits et de biens, a conduit les

entreprises a explorer les possibilités de la "capitalisation des connaissances" par l'intermédiaire des technologies numériques. C'est sur ces bases que se développent les instruments électroniques que nous, chercheurs, devons ensuite utiliser par exemple, pour accéder à l'information publiée sur Internet. Or, ces techniques sont actuellement conçues à partir d'une "théorie" implicite largement erronée : elles supposent, par exemple, que la connaissance peut être "capitalisée" après-coup, après avoir été élaborée. Or, l'hypothèse que j'ai proposé d'adopter pose que la capitalisation des connaissances sur supports numériques dynamiques ne peut être pleinement valide qu'à la condition que l'élaboration d'une connaissance ou d'un corps de connaissances soit en tant que telle, originairement et spontanément, l'opération de capitalisation de cette connaissance : les activités de conception doivent être d'emblée organisées comme activités de capitalisation et de partage de connaissances, et donc de coopération et de distribution des connaissances et de la décision.

L'enjeu de cette question dans le contexte actuel est de faire en sorte que les instruments d'aide à la navigation que prétendent être les moteurs de recherche laissent place à des instruments de construction collective d'autorités, par des sociétés d'auteurs qui sachent exploiter le potentiel immense qu'ouvre la dynamique des supports numériques. La question sous-jacente est évidemment l'immense question épistémologique que soulève l'hypothèse d'une "sémantique située".

(Ce texte reprend à la fois le contenu oral d'une intervention présentée à Alexandrie (Égypte), et des problématiques déjà développées dans d'autres articles rédigés au cours de ces dix dernières années, et dont il constitue une sorte de synthèse).