

# Lexicographie et formes opératoires de l'hypertextualité\*

David Piotrowski

## 1. Introduction

Si l'on s'en tient aux apparences, on conviendra que les dictionnaires se prêtent "naturellement" à une "hypertextualisation". La compatibilité des formes et des opérations de l'hypertextualité avec, respectivement, les caractéristiques organisationnelles et les conduites de lecture d'un fonds lexicographique semble en effet s'imposer à l'évidence. On en connaît bien les motifs.

D'un côté, l'hypertextualité est communément définie par trois propriétés, à savoir (i) *la fragmentation du contenu* : un hypertexte est un agrégat d'éléments d'information, de préférence de taille réduite, entretenant des connexions diverses, (ii) *l'informatisation* : le contenu d'un hypertexte est installé sur support électronique, et (iii) *la non-linéarité des lectures* : le contenu et la matérialité d'un hypertexte n'imposent aucune directive de lecture, et les fonctionnalités du dispositif informatique permettent d'effectuer des parcours multiples dans le fonds enregistré. Ainsi, "sera désigné comme hyperdocument tout contenu informatif informatisé dont la caractéristique principale est de ne pas être assujéti à une lecture préalablement définie (...). Un hyperdocument est donc tout contenu informatif constitué d'une nébuleuse de fragments dont le sens se construit au moyen d'outils informatiques à travers chacun des parcours que la lecture détermine" [Balpe, 1990, p. 6].

De l'autre côté, celui des dictionnaires, se trouvent des textes articulant deux niveaux de structure : la macrostructure, à savoir, principalement, la succession des entrées suivant l'ordre alphabétique, et la microstructure, à savoir, schématiquement, l'ordonnement des contenus d'article en une hiérarchie d'indicateurs métalinguistiques, d'unités de traitement (simples ou complexes) et d'informations lexicographiques (de diverses sortes). On sait que l'ordre de la macrostructure n'est qu'un artifice dicté par des exigences pratiques : il ne participe en rien de la rationalité du texte lexicographique. La sélection des articles qui rythme la lecture d'un

\*Cet article est une version remaniée de notre contribution au rapport d'étape 1996 du programme CNRS «Cognition, Communication intelligente et Ingénierie des langues», Axe III : «Hypertextes et Dictionnaires». La position problématique qui y est développée reprend et synthétise dans un cadre élargi deux précédentes études [Piotrowski, 1994, 1996].

dictionnaire n'est donc pas déterminée par un fil textuel immanent, mais laissée à l'initiative du lecteur. Par ailleurs, les contenus d'articles ouvrent chacun sur une multitude de pistes de lecture, plus ou moins "fléchées", parmi lesquelles le lecteur, guidé par ses objectifs, sélectionne la plus pertinente. Enfin, sous format électronique, les données et les informations lexicographiques se laissent atteindre suivant d'autres canaux que celui des seules entrées — ce qui a pour effet d'accroître les latitudes de cheminement dans le texte.

C'est sous cet angle qu'un dictionnaire informatisé constitue "naturellement" un hypertexte : les subdivisions du fonds lexicographique se distribuent en une "nébuleuse" (*ibid.*) de parcelles d'information interconnectées qui, par le biais des procédures d'accès, sont autant d'étapes possibles pour d'innombrables parcours de lecture.

Mais il est difficile de se satisfaire de telles évidences, pour au moins deux raisons.

D'abord, parce que la définition précitée des hypertextes n'est pas recevable : les caractéristiques qu'elle prend en considération, à savoir la *fragmentation* et la *non-linéarité*, ne sont pas le propre des hypertextes. On sait en effet que les lectures non linéaires dans des recueils d'unités textuelles (comme, par exemple, les articles d'encyclopédie ou de dictionnaire) ont pu être pratiquées dès que le format *codex*, se substituant au *volumen*, a rendu matériellement possible l'accès à des composantes "locales" indexées. La numérotation des pages, l'ordre alphabétique des index, les renvois ou les tables de matière ont administré très tôt des parcours de lecture "discontinus", des parcours affranchis du dirigisme d'un support linéaire ou d'un fil textuel obligé.

Ensuite, cette conception des formes de l'hypertextualité s'avère par trop limitative dans ce qu'elle permet d'appréhender des opérations sémantiques et cognitives dont la textualité est le siège. En effet, en tant qu'elles sont définies par les seules caractéristiques d'*organisation interne* (ensemble de composantes entretenant des connexions variées) et de *procédure d'investigation* (conduites d'exploration faiblement canalisées : chaque élément d'information est au carrefour de nombreux chemins de lecture), les formes de l'hypertextualité ne retiennent qu'un moment artificiellement figé des processus mis en œuvre dans l'activité de lecture. Car si la matière textuelle est susceptible d'être appréhendée sous la forme d'un complexe de composantes délimitées et interconnectées, un tel complexe ne constitue pas une donnée absolue, mais le produit toujours révisable d'un certain nombre de choix interprétatifs, le résultat d'une activité d'appropriation et d'explicitation du texte. La structure hypertextuelle, telle qu'elle est ici envisagée, est donc à considérer comme le produit d'un travail de lecture — produit dont elle constitue un aboutissement possible.

De plus, l'ordre de l'hypertextualité ne se réduit pas à une distribution statique d'éléments interconnectés : les fonctionnalités de l'appareil

informatique introduisent une dimension dynamique qui commue le système des connexions en réseau de circulation. L'hypertexte, comme structure résultant d'un processus d'explicitation d'une matière textuelle, est donc aussi un espace de données à investir suivant une multitude de voies. Et sous cet angle, l'hypertexte se situe à la source de nouvelles activités interprétatives.

Il serait en effet absurde de concevoir les navigations dans l'hypertexte comme des déplacements mécaniques d'étapes en étapes sans autre fin que de circuler, en somme comme l'activité d'un robot sillonnant sans but tous les chemins possibles. Car ce qui est essentiellement introduit avec l'hypertextualité, comme il est d'ailleurs implicitement entendu dans la citation précédente de J.-P. Balpe, ce n'est pas une panoplie de techniques facilitant la collecte de données éparses, mais bien de nouvelles modalités interprétatives, de nouveaux moyens de production de sens : autour des dispositifs hypertextuels s'élabore un nouvel espace de textualité et, corrélativement, de nouvelles pratiques de "travail" des contenus.

Définie comme mode d'organisation et d'investigation d'une collection de données, l'hypertextualité se situe donc à l'articulation de deux dynamiques d'appropriation des textes ; dynamiques qui, en retour, lui confèrent sa pleine teneur. Elle constitue d'une part l'*aboutissement* d'un travail d'explicitation d'un certain matériau textuel, et, de l'autre, la *matière* et le *principe* de nouvelles approches interprétatives. Ainsi, l'hypertextualité est indétachable des processus interprétatifs qu'elle suture, et sa définition en tant que forme d'organisation autonome (statique et dynamique) s'avère par conséquent trop restrictive.

La conception ici critiquée, qui ne s'attache qu'aux caractéristiques organisationnelles les plus générales d'un hypertexte (fragmentation, informatisation, non-linéarité) ne fournit donc pas les outils nécessaires pour aborder la question cruciale des pratiques sémantiques attachées aux formes opératoires de l'hypertextualité.

Pour se donner les moyens de traiter exhaustivement et adéquatement les différentes questions que soulève l'hypertextualisation d'un fonds lexicographique, il sera donc nécessaire de se doter au préalable d'une problématique de l'hypertextualité plus complète : une problématique qui intègre explicitement la dimension des opérations interprétatives, c'est-à-dire la dimension du lecteur, sans pour autant ignorer les caractéristiques des formes "internes" (structure de données et processus de parcours) de l'hypertextualité.

Pour ce faire, nous examinerons trois conceptions (ou "modèles") des hypertextes. Sans négliger les problèmes de cohérence que pose l'hypertextualisation d'un fonds lexicographique, c'est-à-dire, plus précisément, les problèmes de compatibilité entre les différentes composantes fonctionnelles constitutives d'un dictionnaire-hypertexte, nous concentrerons chaque fois notre attention sur la position et le rôle

qu'occupe dans chaque modèle le sujet interprétant, en somme : l'utilisateur de l'hypertexte — et c'est d'ailleurs dans l'ordre croissant de l'importance conférée à celui-ci que nous introduirons ces différents modèles, à savoir, tour à tour, le modèle (à 3 couches) de Dexter [Halasz, Schwartz, 1994], le modèle à "3 modules" et le modèle (à 2 couches) de M. Nanard [1993].

## 2. Le modèle de Dexter

### 2. 1. Présentation

Le modèle de Dexter (ainsi nommé parce qu'il est issu d'une série d'études et de confrontations amorcée à l'occasion d'un séminaire tenu au *Dexter Inn* de l'état du New Hampshire) est un modèle générique d'architecture hypertextuelle. Plus précisément, il s'agit d'un schéma formel spécifiant et articulant les principales composantes fonctionnelles de l'hypertextualité.

Le modèle de Dexter (dorénavant modèle dD) a été conçu pour servir de cadre de référence aux travaux sur les hypertextes : il définit une base de comparaison des différents systèmes existants, introduit une terminologie unifiée dans un domaine encore fluctuant et propose des formats d'échange standards.

L'organisation précise de l'architecture hypertextuelle définie dans le modèle dD ne nous intéresse pas directement. Seuls nous importent les éléments par rapport auxquels se situe l'activité de lecture. Du modèle dD, nous ne montrerons donc qu'une vue schématique et finalisée.

Pour l'essentiel, le modèle dD décrit un hypertexte comme une superposition de trois couches fonctionnelles. Au sommet : la couche *d'exécution* (*run-time layer*), en position médiane : la couche *d'enregistrement* (*storage layer*), et, à la base : la couche du *contenu* (*within-component layer*) (cf. **figure 1**).

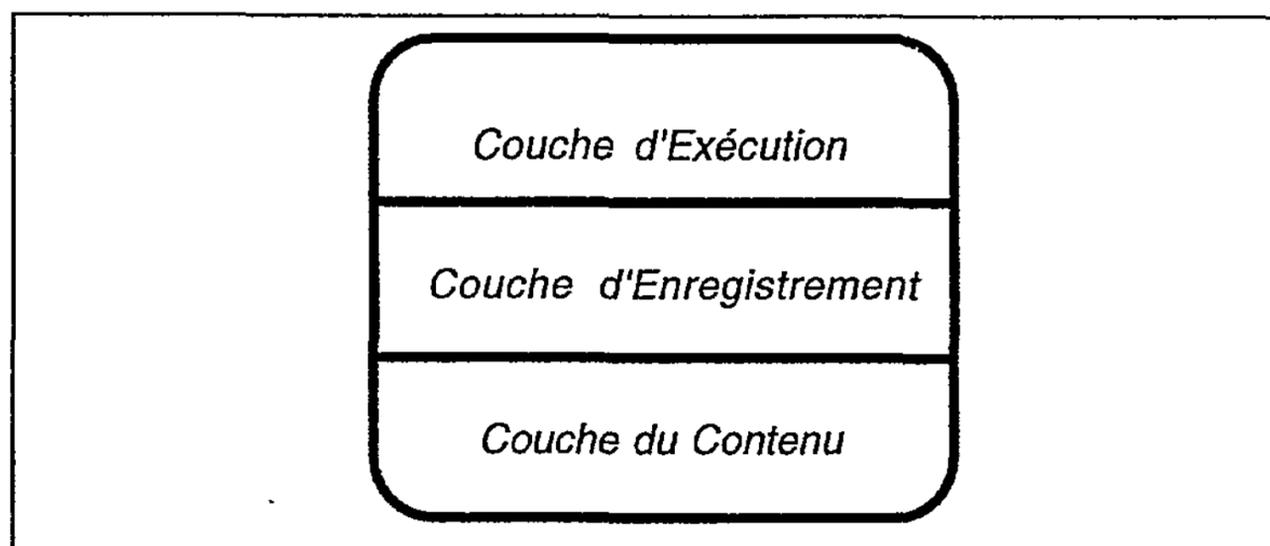


figure 1

Pour comprendre le principe de cette tripartition, revenons à la conception élémentaire d'un hypertexte comme réseau d'unités informatives accessibles suivant des parcours de lecture multiples. Un tel réseau peut être considéré suivant trois points de vue complémentaires. D'abord, en tant que "pure" structure formelle : en tant que regroupement d'unités abstraites (nœuds) interconnectées. Ensuite, suivant un second point de vue, seuls les contenus informatifs attachés aux nœuds sont pris en considération. Enfin, troisième point de vue, on ne retient que la structure dynamique du réseau, à savoir les modalités de circulation dans le système des nœuds et des liens. Ces trois facettes de l'hypertextualité correspondent respectivement aux couches d'*enregistrement*, du *contenu* et d'*exécution*.

Le modèle dD fournit une caractérisation et une articulation précises de ces trois points de vue. Tout particulièrement, les couches "en contact" entretiennent des rapports d'interface qui fondent l'unité du système. La couche d'exécution et la couche d'enregistrement sont raccordées par des *instructions de présentation (presentation specifications)*, et la couche du contenu est rattachée à la couche d'enregistrement suivant des mécanismes d'*ancrage (anchoring)*. Par ailleurs, c'est au niveau des deux couches supérieures et des rapports d'interface que s'articulent les formes de l'hypertextualité "en général", *i. e.* les structures opératoires communes aux différents systèmes hypertexte. Quant aux contenus spécifiques de chaque hypertexte, ils sont localisés dans la couche inférieure. Le type (texte, image, film, son, graphique...) et la structure de ces données pouvant être des plus divers, la couche du contenu échappe à toute détermination générique et sort du champ de l'hypertextualité. Il en sera donc très peu question par la suite.

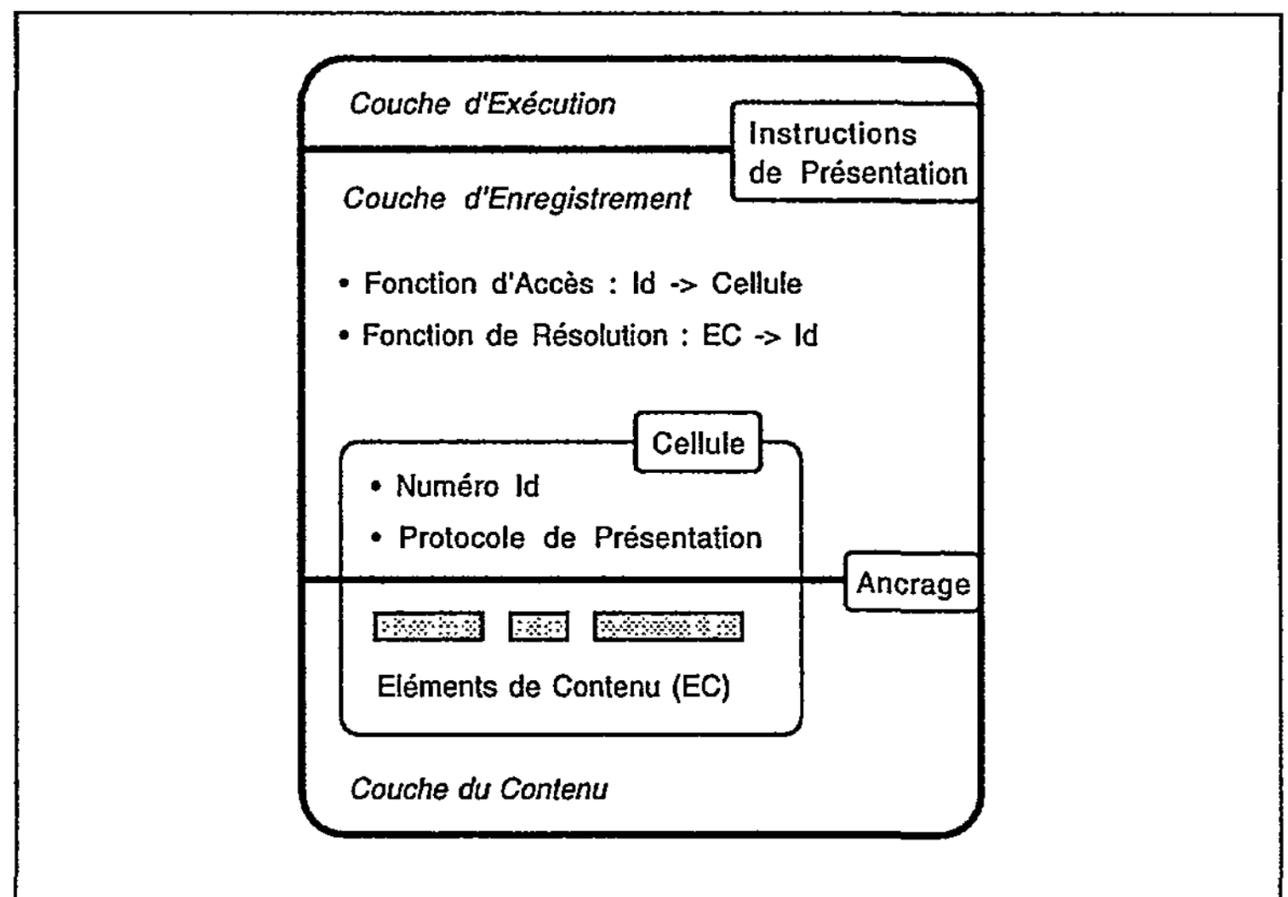
Comme le modèle dD articule trois grilles d'analyse d'un même objet empirique, on ne s'étonnera pas que ses unités constituantes présentent parfois une nature hybride : elles montrent telle ou telle facette suivant l'éclairage qu'on leur apporte, en d'autres termes, elles participent de différentes couches. Abordons les en partant de la couche médiane d'enregistrement.

La couche d'enregistrement renferme la structure "statique" de l'hypertextualité, à savoir, schématiquement, un certain réseau d'éléments interconnectés. Les unités de cette couche, appelées *cellules*, sont soit des liens, soit des cellules atomiques (primitives), soit des complexes construits récursivement à partir des cellules atomiques. Chaque cellule est qualifiée par un numéro d'identification absolu (un *identificateur* noté *Id*). Envisagées sous l'angle du contenu, donc en tant que grandeurs de la couche du contenu, les *cellules* atomiques constituent le réceptacle d'une certaine matière textuelle, sonore, etc. Par ailleurs, les *cellules* comportent des spécifications fonctionnelles activables à partir de la couche supérieure d'exécution où se trouvent définies les procédures d'accès, de présentation

et de manipulation du système hypertexte. Ce sont les *protocoles de présentation* qu'on examinera plus loin. Généralement, une *cellule* reçoit aussi des qualifications d'ancrage (cf. *infra*).

Pour ce qui concerne les modes opératoires, la couche d'enregistrement est administrée par deux procédures primitives : d'une part, la *fonction d'accès* qui reçoit comme argument un *identificateur* et "produit" l'accès à la cellule qualifiée par cet identificateur, et, d'autre part, la *fonction de résolution* qui prend comme argument un "élément de contenu" (par exemple un fragment du texte) et produit l'identificateur de la (ou les) *cellule* contenant cet argument.

A ce stade de notre présentation, l'architecture du modèle dD peut être illustrée par la figure suivante (**figure 2**) :



**figure 2**

Tournons-nous maintenant vers les mécanismes *d'ancrage* et leur application principale : l'organisation des liens. Pour ce faire, observons d'abord que dans un hypertexte les liens sont généralement établis entre des fragments de contenu. Conjuguant ainsi des caractéristiques du contenu et les formes relationnelles du réseau global, les liens constituent des entités composites dont il importe de distinguer les deux facettes, chacune au niveau de sa propre couche fonctionnelle. Le mécanisme d'ancrage est introduit à cette fin. Une ancre est un attribut de cellule défini par une paire de propriétés : un *identificateur d'ancre* et une *valeur d'ancre*. La *valeur d'ancre* est un "élément de contenu" pointé dans la cellule où se trouve cette ancre. L'*identificateur d'ancre* est le numéro de l'ancre dans le contexte de la cellule en question. La détermination complète du contenu pointé par une ancre requiert donc deux

informations : l'identificateur (absolu) de la cellule, et l'identificateur d'ancre dans cette cellule. Cette structure d'ancrage permet alors d'effectuer des modifications au niveau du contenu (changement de la valeur d'ancre) sans altérer l'ordre et la consistance du réseau hypertextuel défini dans la couche d'enregistrement (conservation de l'identificateur d'ancre).

Il est maintenant possible de détailler la structure des liens. Pour ce faire, il faut d'abord introduire une composante intermédiaire intitulée *spécifieur* : un *spécifieur* est un quadruplet d'attributs, à savoir (i) un *identificateur* de cellule (résultant de l'application de la *fonction de résolution* à une certaine grandeur de contenu), (ii) un *identificateur d'ancre*, (iii) un attribut de direction (apportant des informations sur l'orientation du lien ; par exemple SOURCE ou BUT), et (iv) un *protocole de présentation* (cf. précisions *infra*). Un lien est alors défini comme une *cellule* (portant un numéro d'identification absolu) regroupant au moins deux *spécifieurs*. On notera que cette définition permet de construire des liens d'arité quelconque. La structure d'un hypertexte peut maintenant être représentée comme suit (cf. **figure 3**) :

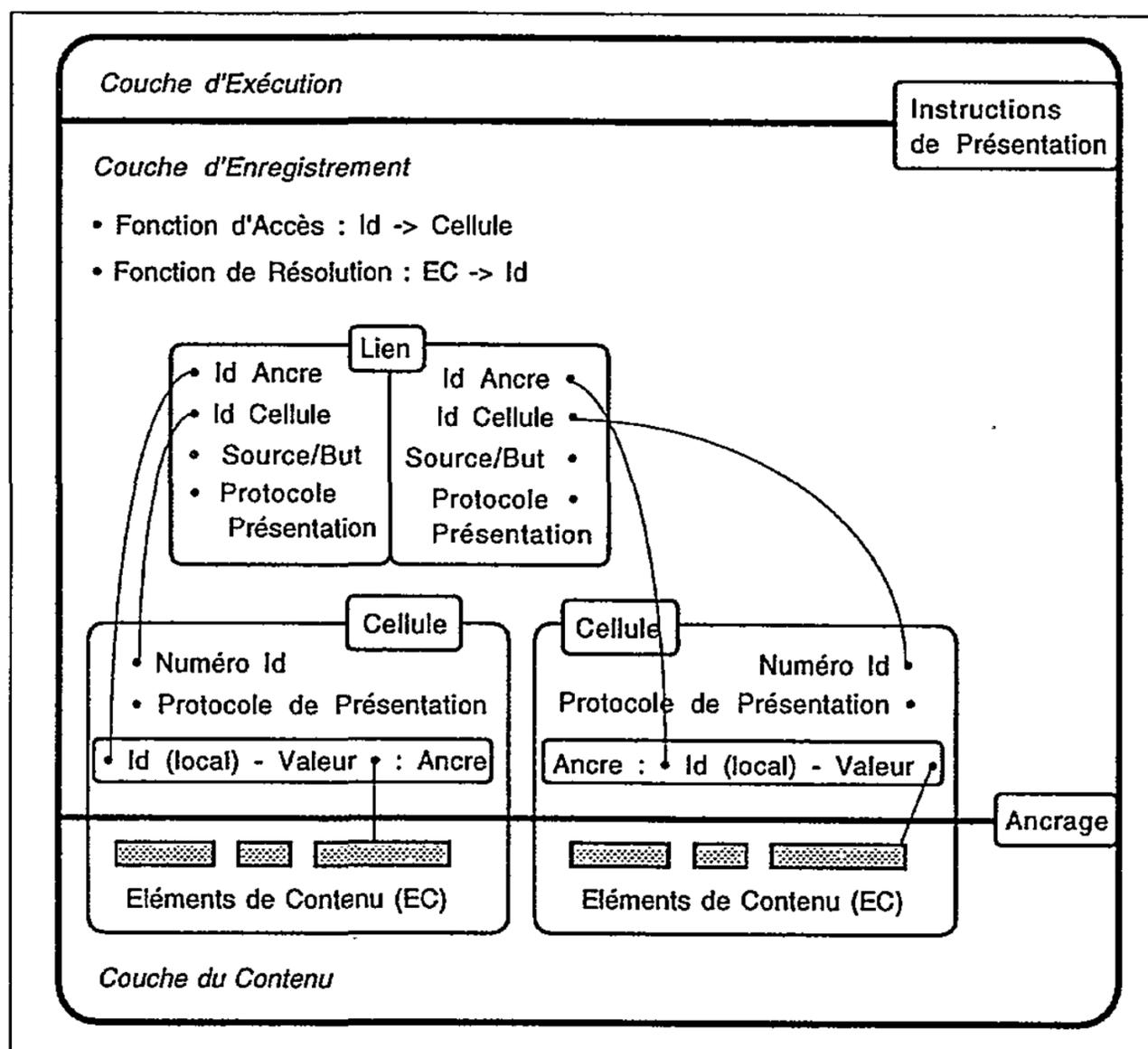


figure 3

C'est au niveau de la couche supérieure que le contact entre le système hypertexte et l'utilisateur est établi.

La couche d'exécution renferme les procédures de manipulation du système hypertexte, à savoir les procédures de sélection, de visualisation et de traitement des données et de la structure de l'hypertexte : "(...) [it] provides tools for accessing, viewing and manipulating the network structure" [Halasz, Schwartz, 1994, p. 32]. Au niveau de cette couche, "the Dexter model provides [the] bare-bones model of the mechanism for presenting a hypertext to the user for viewing and editing. This presentation mechanism captures the essentials of the dynamic, interactional aspects of hypertext systems (...)" [*ibid.*].

La couche d'exécution peut être décrite comme un "cadre de présentation" où des "images" partielles de l'hypertexte, tel qu'il est constitué dans les deux couches inférieures, sont *produites* et *manipulées*. Ces "présentations" des composantes (cellules ou liens) d'un hypertexte sont appelées *instanciations*. Pour pouvoir faire l'objet de traitements divers (au niveau de la couche d'exécution), les instanciations portent un numéro d'identification (*identificateur d'instanciation* noté *IdI*).

Lorsque le lecteur fait usage de l'hypertexte, il ouvre une *session* de travail qui, dans le modèle dD, désigne une structure fonctionnelle définie par (i) une correspondance entre les *IdI* des *instanciations* produites et leur corrélat (cellule ou lien) dans les couches inférieures, (ii) un historique des opérations accomplies, (iii) une fonction de *résolution* (l'analogue de la fonction de même nom dans la couche de l'enregistrement), (iv) une fonction d'*instanciation* (cf. *infra*) et (v) une fonction de *réalisation* (cf. *infra*). Examinons ces deux dernières fonctions.

Dans ses grandes lignes, la connexion entre les couches de l'enregistrement et de l'exécution s'effectue par le canal des *instanciations*. En effet, une instanciation est une entité hybride : en tant que *présentation manipulable* elle appartient à la couche d'exécution, en tant que *grandeur présentée* elle relève des couches inférieures. Pour garantir la cohérence du modèle, ces deux facettes doivent donc être clairement situées au niveau de chaque couche. Précisément, le produit et le "moteur" des instanciations vont relever de la seule couche d'exécution, et les arguments du "moteur" d'instanciation, des autres couches. Mais il faut aussi tenir compte du fait que les composantes d'un hypertexte sont passibles de différentes formes de présentation. Par exemple, dans une pièce de théâtre "hypertextualisée", les répliques de tel personnage pourraient être affichées en caractères rouges, et celles d'un autre personnage, en caractères jaunes. Comme les indications de présentation sont spécifiques à chaque composante, elles constituent donc, au niveau de la couche d'enregistrement, une propriété de *cellule* dénommée *protocole de présentation* (cf. *supra*). Le moteur d'instanciation qui produit les images manipulables des différentes cellules en fonction de leur protocole de présentation s'intitule fonction d'*instanciation*.

Enfin, le système de traitement hypertextuel doit donner la possibilité de "reverser" les transformations effectuées sur les entités virtuelles

manifestées au niveau de la couche supérieure (les *instanciations*). Cette opération est accomplie par la fonction de *réalisation* (l'inverse de la fonction d'*instanciation*) qui "takes an instantiation and returns a (new) component [*i. e.* cellule] that «reflects» the current state of the instantiation" [*ibid.*, p. 37].

Pour obtenir une vue plus complète du modèle dD, il suffira d'enrichir la figure précédente en ajoutant aux cellules un protocole de *présentation* et en inscrivant dans la couche supérieure les spécifications d'une *session*.

## 2. 2. Position fonctionnelle du lecteur

La position fonctionnelle que le lecteur occupe dans le modèle dD est très rapidement circonscrite. En effet, dans ce modèle, les hypertextes sont envisagés sous l'angle fermé de leur structure interne, statique (réseau hypertextuel) ou dynamique (commandes d'accès et de traitement). Partant, le lecteur se trouve quasiment exclu : il reste extérieur à cette superposition serrée de couches au sein de laquelle, en tant que sujet interprétant, il n'arrive pas à s'insérer. De fait, seul le niveau supérieur rencontre l'utilisateur, mais à travers le filtre d'une "boîte à outils" de lecture/écriture. En d'autres termes, le modèle de Dexter laisse hors de son champ la signification et la finalité des activités qu'il autorise. Il réduit le lecteur à un automate capable de manipuler aveuglément les outils disponibles au niveau de la couche supérieure, sans autre contrainte que celle de leur bonne mise en œuvre.

En caractérisant les hypertextes comme une structure de données doublée d'une batterie de procédures qui permettent de l'investir, le modèle dD traite la cohérence de l'hypertextualité au seul palier des régimes de régulation internes et occulte la dimension des pratiques d'investigation et d'appropriation des contenus. Il faudra s'attacher à combler cette lacune.

## 3. Le modèle à trois modules

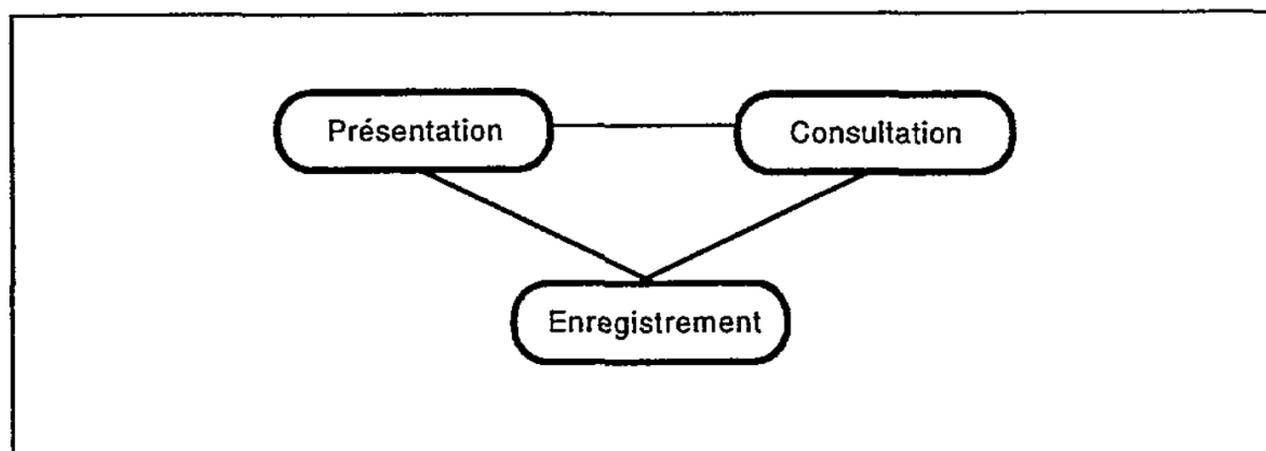
### 3. 1. Présentation

Le modèle à trois modules (dorénavant modèle à3M) est issu d'une étude empirique [Piotrowski, 1994] portant sur les défaillances et les limitations de quelques dictionnaires électroniques disponibles. Il s'agissait, entre autres, de définir une structure fonctionnelle permettant de situer et de qualifier un ensemble de difficultés auxquelles l'utilisateur se trouve confronté dans ses investigations d'un fonds lexicographique informatisé. Bien que l'attention soit essentiellement portée sur les inconsistances internes du dictionnaire-hypertexte, inconsistances qui constituent autant d'écueils dans les parcours d'exploration du fonds, le modèle à3M alloue une place explicite et précise au lecteur — mais, nous

verrons, de façon encore trop limitative pour fonder un cadre problématique satisfaisant.

Les modules qui composent le modèle à3M sont (i) le module de *présentation* (module P), (ii) le module de *consultation* (module C) et (iii) le module d'*enregistrement* (module E).

Le module E est une structure de stockage des données. Le module P désigne un dispositif d'interface par le biais duquel les informations stockées sont visionnées, *i. e.* sont manifestées au lecteur. Le module C désigne l'appareil fonctionnel qui permet d'accéder aux données enregistrées, autrement dit, c'est le dispositif qui administre la lecture. Le modèle à3M distribue ses différents modules aux sommets d'un triangle dont les côtés représentent leurs interactions mutuelles (cf. **figure 4**).



**figure 4**

Avant d'examiner le système des trois modules, notamment en vue de déterminer le rôle fonctionnel qu'il confère au lecteur, nous précisons de quelle conception de la textualité sur support électronique relève le modèle à3M, pour la confronter ensuite à la perspective qui est celle du modèle dD.

### 3. 2. Principes du modèle à3M

Le modèle à3M procède des deux observations suivantes : (i) un texte sur support papier superpose jusqu'à les confondre les dimensions de l'enregistrement, de la consultation et de la présentation, (ii) le transfert du "texte-papier" sur support électronique ; son inscription dans les schémas fonctionnels du dispositif informatique — en somme son hypertextualisation (abstraction faite des opérations d'écriture) — a pour effet de dissocier les trois précédentes dimensions. Voyons cela.

Concernant d'abord le premier point : dans un texte sous format "papier", le lecteur a sous les yeux ce qui lui est "présenté", c'est-à-dire ce qui est imprimé sur les pages, soit encore, ce qui y est inscrit, *i. e.* ce qui y est "enregistré". Or ce qui est inscrit, ce qui remplit les pages, l'est suivant les règles de la typographie, de la mise en page, etc. Autrement dit, ce qui est inscrit est *identique* à ce qui est présenté "couché sur le papier".

L'enregistrement et la présentation constituent donc un seul et même système fonctionnel. Ensuite, on remarque que la lecture (ou "consultation") d'un ouvrage "traditionnel" se règle sur sa présentation. En effet, l'activité de lecture prend nécessairement appui sur les éléments typographiques "affichés" — que ce soit des indices de structure ou les symboles alphabétiques qui composent un texte. Aussi, la conduite de consultation est bien isomorphe à la présentation qui n'est elle-même que le corrélat matériel de l'enregistrement. Sur support papier, les formes constitutives des systèmes de présentation, d'enregistrement et de consultation se recouvrent donc parfaitement.

Le transfert sur support électronique a pour effet de disjoindre ces trois dimensions de la textualité et de les établir en modules fonctionnels pourvus de formes d'organisation propres. Pour argumenter ce point, reprenons après N. Ide, J. Le Maître et J. Véronis [1991] l'exemple de l'implémentation d'un fonds lexicographique dans le cadre des bases relationnelles. Rappelons au préalable que, très schématiquement, une base de données relationnelles se définit comme un ensemble de relations caractérisées chacune par un "tableau", à savoir un croisement de colonnes et de lignes où les colonnes correspondent à des attributs (ou propriétés) et où les lignes portent les valeurs que prennent ces attributs.

Comme le notent les auteurs susmentionnés, la structure des bases relationnelles dissout l'unité et la cohérence des articles sous format imprimé. En effet, l'inscription des informations lexicographiques dans des "tableaux" contribue à dissocier des spécifications qui sont connexes (*i. e.* adjacence spatiale *et* sémantique) dans le texte imprimé. Prenons par exemple une version simplifiée de l'article *discipliner* dans le *Petit Robert*, soit :

**Discipliner** v. tr. 1. Accoutumer à la discipline ; donner le sens de l'ordre, du devoir, de l'obéissance à. *Discipliner une classe. Discipliner une armée.* 2. Plier à une discipline intellectuelle ou morale. «*La religion et les beaux-arts disciplinent les instincts rebelles*» (Maurois). 3. *Discipliner les cheveux*, les maintenir bien coiffés.

Cet article, qui comporte trois subdivisions, sera "enregistré" sous la forme des deux tableaux suivants (Attributs : Ved = vedette, NsD = numéro de subdivision, Def = texte de définition, EI = élément d'illustration (exemple, citation...), UTI = unité de traitement interne) :

Ved	NsD	Def
discipliner	1	Accoutumer à la discipline
discipliner	1	Donner le sens de l'ordre, du devoir, de l'obéissance à
discipliner	2	Plier à une discipline intellectuelle ou morale
discipliner	3	Les maintenir bien coiffés

Ved	UTI	NsD	EI
discipliner	discipliner	1	<i>Discipliner une classe</i>
discipliner	discipliner	1	<i>Discipliner une armée</i>
discipliner	discipliner	2	« <i>La religion et les beaux-arts disciplinent les instincts rebelles</i> »
discipliner	<i>discipliner les cheveux</i>	3	

Comme le soulignent les auteurs (qui traitent un exemple en tout point analogue), on voit bien que les acceptions dont rendent compte les deux définitions du premier niveau de subdivision “(...) are more closely related to each other than to senses 2, and 3 (...), but the tabular format obscures this fact (nous soulignons)” [Ide *et al.*, 1991, p. 9]. Alors que la présentation du texte imprimé relate bien la connexité sémantique de ces deux contenus définitionnels, dans une représentation en tableau, qui a pour effet de “(...) fragmenting the view of the data” [*ibid.*, p. 8], ceux-ci se retrouvent dégroupés et isolés. Il en est d’ailleurs de même s’agissant des rapports de portée (factorisation des informations) : “flattening this structure [of an article] into a tabular form obscures the derivational relations captured in the nested arrangement [of an article]” [*ibid.*, p. 9].

Mais, si l’implémentation en base relationnelle dissout l’unité informative d’un article, ce n’est qu’au niveau de sa “présentation”. En effet, la cohésion de l’article est certes moins flagrante lorsque celui-ci est présenté sous forme de tableaux, mais elle reste toutefois vérifiée au plan de la forme logique : “(...) the different relations represented in different tables are not explicitly connected, but are only logically connected (...)” [*ibid.*, p. 8]. Il est tout à fait possible, par recouvrement des valeurs des différents attributs, de reconstituer la structure hiérarchique d’un article et, partant, de rétablir les relations de connexité entre ses éléments pour, au besoin, les “afficher” de manière satisfaisante.

L’exemple de l’implémentation en base relationnelle illustre donc clairement la séparation des modules d’enregistrement et de présentation qu’une informatisation entraîne. L’enregistrement distribue les informations suivant certaines modalités organisationnelles (par exemple en tableaux) conçues non pas pour être directement visualisées mais pour permettre la mise en œuvre de procédures de consultation complexes. La forme apparente de l’enregistrement n’a aucune pertinence, et la présentation des informations échoit donc nécessairement à un module indépendant dont “l’argument” est un ensemble d’informations atteintes suivant certaines requêtes (dont la plus simple est “rechercher un certain article”) et le “produit”, une certaine configuration de textes et de marqueurs à l’écran.

Les modules d’enregistrement et de consultation constituent donc bien des fonctionnalités autonomes. Examinons maintenant ce qu’il en est du module de consultation.

L'unité des modules E et P une fois dissoute, le module de consultation se retrouve dans la position d'un tiers terme. En effet, l'activité de consultation implique et couple les composantes d'enregistrement et de présentation : c'est dans la banque de données et suivant ses régimes d'organisation propres que s'effectue la recherche des informations, et, corrélativement, c'est sur la base de ce qui est accessible, *i. e.* présenté, que se poursuit la lecture. Aussi, de deux choses l'une : soit les systèmes d'enregistrement et de présentation sont isomorphes, et le processus de lecture qui les implique conjointement leur est logiquement assimilé (il en est ainsi sur support papier), soit les systèmes d'enregistrement et de présentation constituent des modules autonomes et non superposables, et dans ce cas, le processus de lecture se détache de leur unité rompue pour s'investir de formes propres. Dans l'impossibilité d'être conjugué aux structures de deux modules *a priori* mutuellement irréductibles, le module de lecture se trouve alors condamné à constituer le troisième pôle d'un triangle d'interactions fonctionnelles.

### 3. 3. Modèle dD vs Modèle à3C

Le modèle dD et le modèle à3M ont en commun de produire une analyse de la textualité sur support électronique. Ils visent tous deux à expliciter les formes et les régimes constitutifs de ce nouveau genre textuel. Mais ils n'ont pas en vue le même objet, ou tout du moins, les mêmes aspects d'un phénomène global, à savoir l'activité de lecture sur dispositif informatique.

En effet, le modèle dD se focalise sur l'outil. L'objet dont il rend compte est un "hypertexte" considéré comme complexe fonctionnel susceptible de diverses manipulations, *abstraction faite du facteur de ces manipulations* (motivations et finalités d'un sujet interprétant). Plus précisément, on a vu que cet objet articule trois niveaux d'organisation : celui du contenu (couche du contenu), celui du réseau (couche d'enregistrement) et celui des connexions avec l'utilisateur (procédures d'affichage et commandes de la couche d'exécution). La "cible" du modèle dD est donc un système textuel ouvert sur l'utilisateur, mais en rupture des processus d'appropriation (et d'interprétation) des contenus que celui-ci met en œuvre. Autrement vu, si on accepte de décrire le fait empirique de la lecture comme la "somme" d'un système hypertexte (l'ouvrage), d'une activité de manipulation (de l'hypertexte) et d'une activité d'interprétation (des contenus atteints), le modèle dD ne retient que le premier terme de ce fait global ; les deux autres, dont la somme est égale aux aptitudes du lecteur, étant évacués. En définitive, dans le modèle dD, le lecteur voit sa position réduite à une place vide : il ne vaut que comme pôle virtuel d'une connexion mécanique avec le système hypertexte.

Le modèle à3M vise quant à lui l'activité fonctionnelle de lecture. Son approche de l'hypertextualité dépasse le seul système-outil pour englober (mais partiellement, nous le verrons) la pratique de lecture. Plus précisément, le modèle à3M conserve du processus global de lecture, envisagé précédemment comme somme de trois termes, les deux premiers et abstrait le troisième. En effet, alors que dans le système dD la superposition serrée des couches n'offrait aucune place au lecteur (parce que les trois couches sont fonctionnellement rattachées), le modèle à3C, dont les modules sont disjoints, laisse l'utilisateur s'introduire entre les composantes C et P (cf. figure 5).

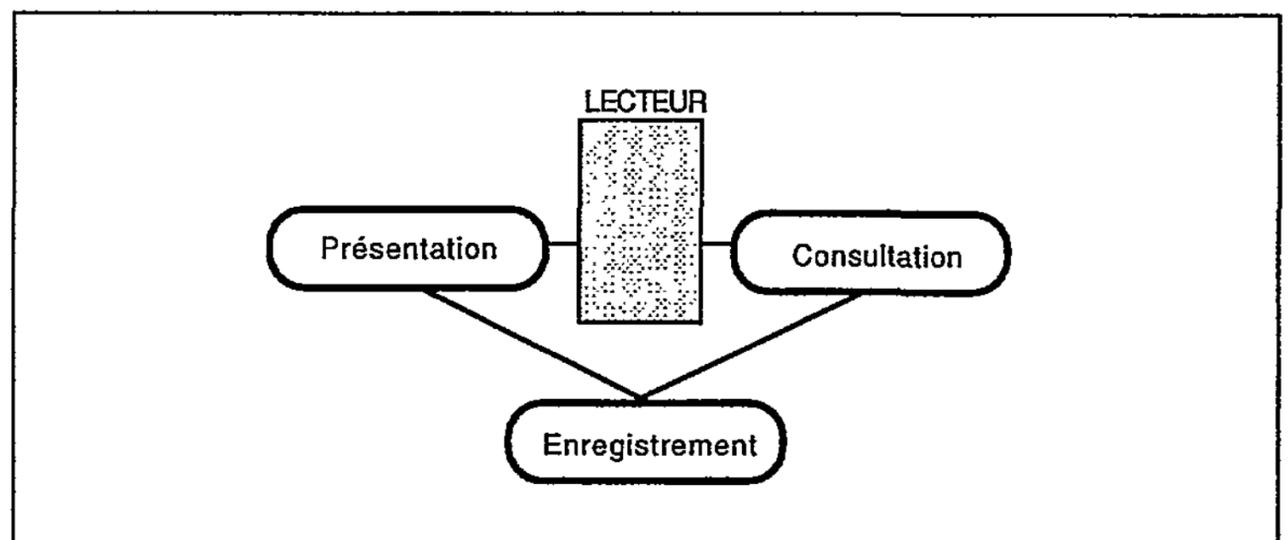


figure 5

La position fonctionnelle ainsi accordée au lecteur est aisément explicable : le lecteur est le sujet qui appréhende des informations "affichées" (module P) et qui, en fonction de ses finalités de lecture, actionne des commandes dans le but d'atteindre de nouvelles données (module C). Le lecteur apparaît donc comme un système "de transition", *i. e.* une instance décisionnaire, qui prend appui sur des informations affichées (marques activables ou données) pour choisir, en manipulant les "manettes" du module C, de poursuivre son travail de fouille dans telle ou telle direction.

Ainsi, dans le modèle à3M qui décrit le système hypertexte comme une triangulation de modules autonomes, une place fonctionnelle est clairement allouée au lecteur. Reste que celui-ci n'est qu'à "moitié" pris en considération, à savoir par le seul trait de son activité de manipulation (cf. *supra*). En définitive, alors que dans le modèle dD la place (vide) du lecteur *se raccorde* au système hypertexte par un *unique* canal "externe" : le lecteur "virtuel" se positionne "au dessus" de l'hypertexte qu'il aborde à travers l'écran fonctionnel de la couche supérieure, dans le modèle à3M l'utilisateur *s'intègre, s'incorpore* au système hypertextuel en engageant *deux* connexions participatives, respectivement avec les modules C et P. Mais si le modèle à3M se rapproche bien de l'utilisateur, il n'embrasse toutefois pas sa pleine intégralité : manque encore la facette interprétative.

Avant d'aborder le modèle "à deux couches" de M. Nanard, nous achèverons la description comparative des modèles dD et à3M avec une dernière observation : le modèle à3M comporte un seul niveau d'enregistrement, alors que le modèle dD en articule deux — ce qui illustre bien l'accent mis dans ce modèle sur la structure interne de l'hypertexte.

### 3. 4. Les dictionnaires électroniques à la lumière du modèle à3M

Le modèle à3M ouvre deux séries de questions. L'une concerne les modules en tant que systèmes autonomes, dédiés à des opérations et à des usages précis et, en tant que tels, objets de problématiques propres. L'autre traite des interactions entre les trois modules et de la cohérence du système qu'ils composent. Ces différentes questions ont été examinées à travers une étude de cas concrets dans [Piotrowski, 1994]. Nous en résumons ici l'essentiel.

#### 3. 4. 1. — Problématiques propres

##### — Le module P

Le module de présentation est la facette du système tournée vers l'utilisateur. Il détermine les modalités suivant lesquelles les divers éléments qui jalonnent l'activité de lecture sont portés à la connaissance du lecteur. Ces éléments, qui constituent les articulations des parcours de navigation, sont de deux sortes. Soit il s'agit d'éléments "statiques", à savoir des contenus enregistrés qui "remontent" du module E vers le module P à seule fin d'être "visionnés" (*i. e.* lus), soit il s'agit d'éléments "dynamiques", c'est-à-dire des grandeurs "activables" : des objets de différentes natures à partir de la présentation desquels la lecture va être relancée. Ce second type de données comprend par exemple les unités de contenu susceptibles d'être "développées" ou de conduire vers d'autres informations par simple action de la souris *sur leur présentation*, les cadres de formulation des requêtes, les menus déroulants, les boutons de défilement des informations (comme dans l'index des formes), etc. Dans sa dimension "passive", le module P s'adresse au module E, dans sa dimension active, il interfère avec le module C. Nous y reviendrons.

Considéré isolément (abstraction faite des interactions avec les modules C et E), le module P administre donc les configurations d'écran (fenêtres, couleurs, typographie, défilement, zones activables, commandes...) suivant lesquelles les données (statiques et dynamiques) sont manifestées à l'utilisateur. Les caractéristiques fonctionnelles du module P définissent ainsi une nouvelle économie de notation des textes dont les contraintes et la rationalité internes doivent être progressivement explorées. Tout particulièrement, les outils électroniques doivent trouver

leur point d'équilibre entre une multiplication et une complexification des fonctionnalités, qu'autorise à peu de frais le dispositif informatique, et les investissements cognitifs parfois démesurés qu'ils exigent de l'utilisateur. Dans ce sens, les travaux poursuivis en matière de lecture sur support informatique ont permis de mettre au jour un certain nombre de consignes simples.

La première consigne regarde la simplicité, l'homogénéité et la standardisation des commandes ; autant de traits qui conditionnent la "(...) facilité d'apprentissage et de mémorisation des fonctionnalités, [un] fonctionnement efficace, [une] possibilité réduite d'erreurs [et un] usage agréable" [Laufer, Scavetta, 1992, p. 102]. Concernant la présentation, l'espace restreint de l'écran et les embarras ergonomiques dont il est la source doivent être compensés par la "souplesse" des présentations. Tout particulièrement, le module P doit permettre "(...) [d']effectuer les opérations standard d'un système de fenêtrage : (...) repositionner les fenêtres, en modifier les dimensions, les fermer, les mettre en veilleuse, sous la forme réduite de petites icônes" [*ibid.*, p. 58]. Soulignons qu'il ne s'agit pas seulement d'agrément, mais aussi d'intelligibilité. Le support informatique détermine une nouvelle rationalité de lecture : les pratiques d'investigation du contenu textuel (attitude de lecture sur écran et collecte automatique d'informations) sont distinctes de celles suivies sur support papier et exigent donc des dispositifs d'affichage et de manipulation spécifiques, assurant notamment une gestion commode des multiples données rendues aisément accessibles et qui entrent alors dans une logique de superposition/confrontation.

#### — Le module E

Les questions afférentes au module E sont des plus complexes et leur traitement n'est pas sans incidence sur la compréhension de l'hypertextualité. Ces questions portent essentiellement sur le statut de la structure des textes. En effet, en tant que module autonome, le système de stockage est soumis à l'impératif de conformité : pour être en mesure de restituer fidèlement l'intégralité de la matière qu'elle consigne, la structure d'enregistrement doit nécessairement être compatible avec les formes constitutives de celle-ci. À partir de là se posent deux problèmes, l'un technique et somme toute trivial, l'autre théorique et ardu.

Le premier problème se formule ainsi : les formes de l'informatique sont-elles en mesure d'accueillir les données que l'on souhaite "hypertextualiser" ? Il est ici implicitement entendu que ces données sont dotées d'une structure propre et que c'est la compatibilité entre celle-ci et les schémas organisationnels de l'informatique qui est en question. Si problème il y a, il est donc technique et sa solution passe par l'élaboration de formats adaptés aux données traitées. Dans ce cas de figure, et en tant qu'il est considéré comme un système autonome, le module E d'un

hypertexte ne relève pas pleinement d'une problématique de l'hypertextualité.

Le second problème porte sur l'implicite du premier problème, à savoir le statut et le mode d'existence de la forme interne d'un texte. Il est clair que si l'on nie l'existence d'une forme textuelle immanente, *i. e.* l'existence d'une structure propre et arrêtée qui constitue une part fixe de l'objectivité textuelle, alors, en écho, le module E doit être reconsidéré dans sa fonction de réceptacle fidèle et exclusif. Nous aborderons ces différentes questions plus avant.

### — Le module C

A l'instar du module P, le module C (envisagé indépendamment des autres modules) doit satisfaire aux exigences de standardisation et de souplesse. Tout particulièrement, il sera souhaitable que ce module dont les structures gouvernent l'exploration des données se conforme à certains schémas opératoires dont l'efficacité, *i. e.* l'adéquation à la rationalité du support informatique, a été prouvée au terme d'une expérience bientôt cinquantenaire. Dans leurs grandes lignes, ces schémas ont les caractéristiques suivants :

(i) — l'utilisateur introduit une "matrice symbolique", aussi dénommée "patron de fouille", qui peut prendre les conformations les plus complexes. Dans le cas le plus simple, il s'agit d'une chaîne de caractères reproduisant la graphie de la forme autour de laquelle gravite la recherche. A un niveau de complexité immédiatement supérieur, certains symboles spéciaux ("joker") valant soit pour une lettre quelconque (en général le symbole "?") soit pour une chaîne de longueur quelconque (en général le symbole "\*") sont introduits — ainsi que les éléments de spécification des "champs" de requête. Enfin, il est aussi possible de composer les matrices précédentes suivant des schémas d'opération logiques et concaténationnelles. Ces schémas sont pour l'essentiel des schémas de cooccurrence (avec ou sans condition d'adjacence ou d'exclusion mutuelle) des matrices dans des environnements à délimiter.

(ii) — le système produit alors des nomenclatures qui permettent à l'utilisateur de visualiser les formes ou les contextes disponibles satisfaisant à la formule d'entrée. Cette seconde étape est donc interactive et a pour principale fonction de présenter à l'utilisateur les formes identifiées dans la base de sorte qu'il sélectionne (ou modifie et confirme une sélection automatique) les unités à partir desquelles il souhaite poursuivre son investigation. En général, lorsque l'entrée ne comporte pas de "joker", le système affiche la liste des formes proches de la composition orthographique de celle qui a été introduite — ce qui permet de "rattraper" les incertitudes que le système informatique ne tolère pas et d'informer sur la présence de formes similaires.

(iii) — sur la base des formes précédemment sélectionnées, l'utilisateur opte pour la recherche de telle ou telle sorte d'information, et il lui est possible d'accéder simultanément à plusieurs informations à travers des fenêtres ouvertes conjointement. Dans le meilleur des cas, ces fenêtres sont dynamiques et synchronisées : la sélection d'une unité particulière dans une fenêtre provoque l'affichage dans les autres fenêtres des informations relatives à l'unité choisie. Il s'agit là d'une des caractéristiques importantes des potentialités de consultation sur support électronique : la diminution de l'espace de lecture, l'inadéquation de l'écran aux textes linéaires et la prolifération des informations ponctuelles que le dispositif favorise sont contrebalancées par l'accès possible aux connexions que ces informations locales entretiennent.

### 3. 4. 2. — Problèmes de cohérence interne

Après avoir envisagé les modules E, P et C du point de vue de leurs exigences propres, donc en les considérant comme des modules autonomes et finalisés, nous les abordons maintenant sous l'angle du système qu'ils composent. Les problèmes que soulève l'intégration de ces trois modules en une unité systémique sont spécifiques à l'hypertextualité. En effet, sur support papier les dimensions P, E et C sont isomorphes, et, partant, la cohérence de leur combinaison est garantie *a priori*. Par contre, sous format électronique, ces trois dimensions constituent des modules fonctionnels autonomes. Aussi, leur synthèse requiert des ajustements réciproques qui sont comme le reflet des contraintes constitutives et de la rationalité "hypertextuelle". Nous donnons maintenant quelques indications sur les contraintes de compatibilité entre les trois modules.

#### — Modules P et C

Les modules P et C interfèrent principalement au niveau des éléments de présentation "dynamique", c'est-à-dire au niveau des éléments se présentant comme le point de départ possible de nouvelles consultations. Cette caractéristique fonctionnelle doit être mise en évidence par quelque artifice de présentation (couleur, surbrillance, aspect du curseur, etc.) qui aura un triple but : d'une part, indiquer quel objet (texte ou symbole) à l'écran possède cette propriété, d'autre part, déterminer une zone d'activation (ou, plus généralement, des commandes) pour "lancer" la procédure de consultation et, enfin, apporter des informations sur le type ou le contenu des données accessibles par activation de l'unité dynamique. Un exemple simple de dysfonctionnement de la connexion P—C est celui des index qui contiennent une succession de formes homographes sans autre spécification. Pour prolonger sa lecture, l'utilisateur n'a alors d'autre solution que de "cliquer" tour à tour sur ces différentes formes et de prendre connaissance des informations auxquelles elles conduisent.

### — Modules E et C

Le module de consultation fournit les cadres dans lesquels le lecteur va exprimer ses requêtes. Sous peine de graves dysfonctionnements, les requêtes formulées doivent être fonctionnellement compatibles avec les structures de stockage. S'il n'en est pas ainsi, la procédure de recherche, qui procède par superposition des éléments de la requête (structure et unités) et du fonds enregistré, ne peut aboutir, sinon en altérant (modification, effacement...) les éléments de la requête — et le module de consultation produit alors des réponses déviantes. C'est le cas, par exemple, lorsque le "formulaire" de consultation accepte des éléments ou des spécifications qui ne sont pas ou qui ne peuvent pas être recherchés dans le fonds.

### — Modules E et P

Du module de présentation, on exigera avant toute chose qu'il produise fidèlement à l'écran la structure des données enregistrées ; surtout si cette structure est pensée comme une composante objective du texte. Par exemple, la structure hiérarchique d'un article sera présentée par une "mise en écran" appropriée (justifications, couleurs, passages à la ligne, etc.). Par ailleurs, les modules P et E interférant avec le module C (cf. *supra*), leur homogénéité conditionne la consistance des parcours de lecture. En effet, il importe que la présentation manifeste clairement les structures d'enregistrement sur lesquelles s'articulent les procédures de recherche. Si cette condition n'était pas satisfaite, le lecteur pourrait être amené à effectuer des "fouilles" dans des registres de stockage qui ne sont pas les bons.

## 4. Le modèle "à deux couches"

### 4. 1. Présentation

Pour clarifier notre démarche, nous avons précédemment suggéré une décomposition naïve de l'activité de lecture : à savoir comme la "somme" d'un système hypertexte (l'ouvrage), d'une activité de manipulation (de l'hypertexte) et d'une activité d'interprétation (des données atteintes). Nous avons alors montré (i) que le modèle de Dexter "vise" la seule première composante : la forme interne du système hypertexte, et (ii) que le modèle à trois modules ne fait abstraction que de la troisième composante : le modèle à 3M envisage le système hypertexte en tant qu'il est l'objet de manipulations. Nous présentons maintenant un modèle de l'hypertextualité qui appréhende les trois composantes de l'activité de lecture : le modèle de Marc Nanard, ou modèle à "deux couches" (dorénavant modèle à 2C) — dont on proposera une version étendue à 3 couches.

En intégrant la dimension des activités de manipulation, le modèle à3M a élargi l'angle d'approche de l'hypertextualité et, en conséquence, par rapport au modèle dD, son analyse de l'ordre interne des systèmes hypertexte a perdu en précision. Il en sera de même pour le modèle à2C qui, englobant la totalité du processus de lecture, ne fournira qu'un schéma très général des formes opératoires de l'hypertextualité.

Suivant les thèses de M. Nanard, un hypertexte "(...) est caractérisé par la projection d'une structure forte ou «réseau hypertexte» sur une structure faible ou «documents», et par l'emploi de la métaphore de la navigation pour *exploiter* (nous soulignons) cette nouvelle structure" [Nanard, 1993, p. 1]. Plus précisément un système hypertextuel se présente comme la superposition par voie d'*ancrage* d'un *ensemble de documents* (1ère couche) et d'une *structure de connaissance* (2ème couche) — cf. **figure 6**.

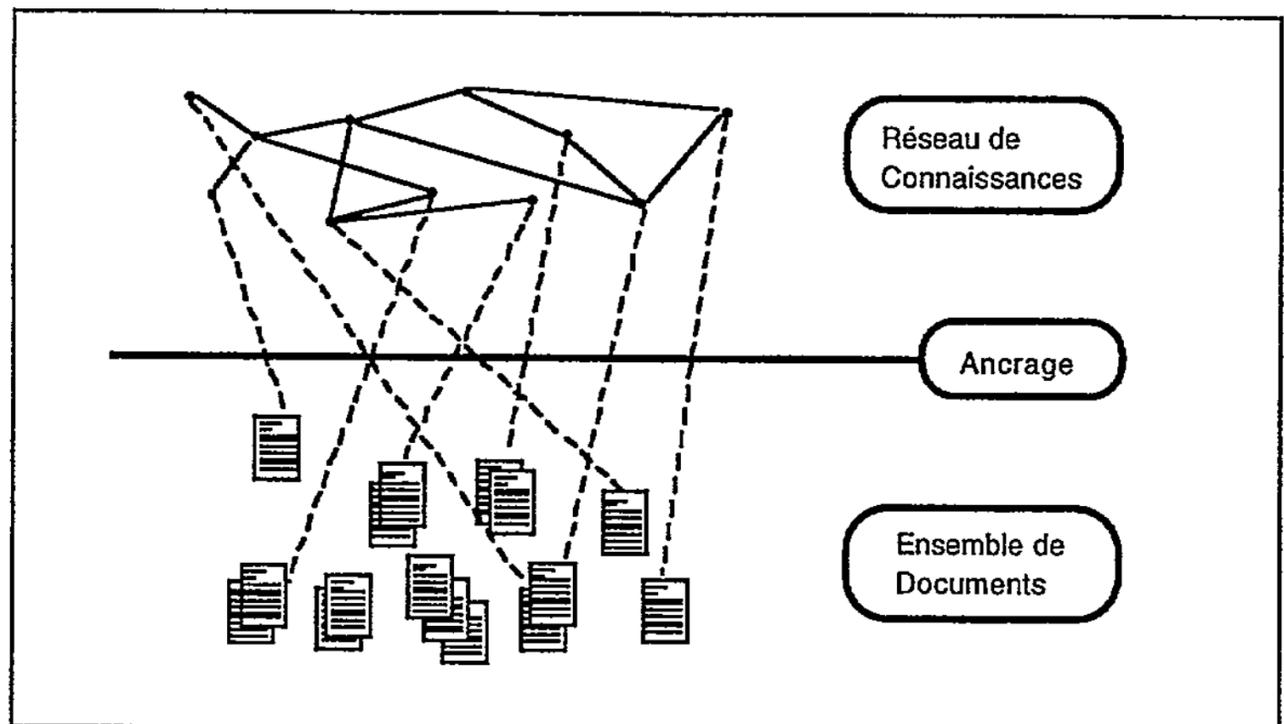


figure 6

. Par *document*, il faut entendre tout contenu textuel dont la signification est atteinte à travers un processus d'interprétation. En d'autres termes, les documents sont des entités (de taille quelconque) qui relèvent d'une saisie par un individu doté de facultés interprétatives : "les documents sont des entités fondamentalement destinées à l'homme" [Nanard, 1994, p. 10]. Les *documents* constituent donc ces données qui sont manifestées sous une forme ou une autre pour être appréhendées par un lecteur suivant le mode d'une interprétation.

La *structure de connaissance* est définie comme un système de *qualifications*, autonome dans son principe, informant la "couche de base" des documents et susceptible de traitements automatisés.

Un mécanisme d'*ancrage* permet alors d'associer aux documents les spécifications de la structure de connaissance. Ainsi, le lecteur pourra parcourir l'ensemble des documents en suivant les indications de lecture

que lui apporte le niveau de la couche des connaissances. En résumé, “la connaissance (au sens informatique) est un ensemble de relations spécifiquement destinées à la machine qui ont pour rôle de lui permettre de décider ce qui doit être proposé au lecteur comme suites possibles de sa lecture (...)” (*ibid.*).

Enfin, on sait que le terme de *navigation* désigne “(...) le déplacement à travers le réseau [que constitue un hypertexte], [il s’agit d’une] métaphore spatiale et intentionnelle qui souligne bien le rôle actif du lecteur” [Bruillard, La Passardière, 1994, p. 16].

La conception de l’architecture hypertextuelle proposée par M. Nanard articule donc trois concepts : les *documents*, la *connaissance* et la *navigation*. Discutons maintenant ces différents concepts.

On remarquera d’abord que ces trois concepts forment un ensemble cohésif. En effet, les documents sont successivement atteints au cours de “navigations” qui suivent les “directions” données (ou suggérées) par la grille des connaissances. Aussi, les interprétations, qui procèdent fondamentalement d’une mise en rapport des différents documents extraits au fil des parcours de lecture, se trouvent déterminées par le système des connaissances — dans la mesure où celui-ci constitue le principe des regroupements raisonnés de documents. *Connaissance, interprétation et navigation* sont ainsi étroitement liées — ce qui présuppose une forte adéquation entre le contenu des documents et le système des connaissances qui “soutient” le processus interprétatif. En effet, si les spécifications “ancrées” aux documents étaient incompatibles avec leurs contenus, il est peu probable que les navigations favorisent des collectes de documents susceptibles d’une interprétation globale consistante. Il s’ensuit que dans une architecture hypertextuelle, le système des connaissances doit d’une certaine façon refléter fidèlement, avec toute la précision et l’univocité requises pour faire l’objet d’un traitement automatique, le contenu des documents.

Mais cette vision demande à être nuancée. Les connaissances ne sauraient en effet être conçues comme des explicitations conformes du contenu des documents, explicitations sur lesquelles toutes les navigations consistantes devraient se régler, mais plutôt comme des explicitations *possibles* de ces documents. De cette façon, le système hypertextuel introduit une interface dans les rapports d’interprétation : entre le document et le lecteur, il glisse une couche intermédiaire de “connaissances” qui guide les approches du lecteur, mais ne constitue en aucun cas une grille de référence absolue. La couche “connaissance” est donc de ce point de vue indépendante de la couche “document” : elle est projetée sur les textes dont elle constitue en définitive une interprétation, un “éclairage” momentané choisi comme voie d’accès : “[la structure des connaissances] est projetée, en ce sens que l’ensemble des informations [les documents] qu’elle structure n’est pas affecté par cette organisation.

L'hypertexte en est une "vue", un "éclairage" particulier. Plusieurs organisations différentes peuvent être projetées simultanément sur le même ensemble d'informations" [Nanard, 1993, p. 1].

La structure fondamentale de l'hypertextualité est donc une *structure dynamique* : le réseau des connaissances qui reflète l'interprétation des documents à un moment donné du processus de lecture va orienter des parcours qui vont susciter de nouvelles interprétations, c'est-à-dire donner lieu à de nouvelles connaissances qui, à un moment ultérieur, pourront à leur tour jouer le rôle de guide de lecture, et ainsi de suite... Un hypertexte se définira donc comme une *structure d'interaction récursive* entre textes (documents) et interprétations temporairement "stabilisées" (ou connaissances "établies") destinées à produire de nouvelles élaborations sémantiques (interprétations).

#### 4. 2. Récapitulation et envoi

Les trois modèles que nous avons présentés se distribuent suivant un double gradient : le premier paramétré par la précision d'analyse des formes internes de l'hypertextualité, le second, par la largeur de la perspective englobant l'activité de lecture. Par ailleurs on a constaté que ces deux gradients sont en rapport inverse l'un de l'autre.

Que peut-on déjà conclure ? Avant tout que ces différents modèles ne sont pas mutuellement exclusifs. Ils offrent des images plus ou moins globales, plus ou moins focalisées et précises d'un même fait empirique : l'hypertextualité. Aussi, c'est bien dans leur combinaison que réside une problématique satisfaisante de l'hypertextualité : une problématique qui ne sacrifie aucune des dimensions constitutives de la lecture sur support informatique. Mais si la concordance des modèles dD et à3M est assez claire — les deux couches d'enregistrement du modèle dD sont ramenées à une unique composante (le module E), la couche d'exécution se dédouble en deux modules (C et P) et le démembrement des trois couches en trois modules alloue un rôle fonctionnel à l'utilisateur —, la concordance des modèles à3M et 2C est quant à elle moins immédiate. Sans doute, on en voit bien le principe, mais encore faut-il la fonder rigoureusement. Au risque d'anticiper, donnons-en tout de suite les principales lignes.

Dans le modèle à2C la couche des connaissances peut être divisée en deux sous-couches : d'une part la sous-couche des connaissances "établies", d'autre part la sous-couche des connaissances produites (interprétations). La sous-couche des connaissances "établies" projette sur la couche des documents une structure stabilisée qui va guider la lecture. Autrement dit, cette sous-couche des connaissances "établies" détermine et fixe une certaine organisation dans la couche des documents. En ce sens, elle définit un certain cadre d'enregistrement de ces données : elle se rapporte ainsi directement au module E du modèle à3M. Par ailleurs, la sous-couche des interprétations est le lieu où s'effectue l'appropriation des données. C'est donc le lieu où se trouve l'appareillage d'investigation.

Aussi, cette sous-couche supérieure correspond au module C et au module P qui, respectivement, administre l'accès aux données et les porte à la vue du lecteur. Ce que résume la figure suivante (figure 7) :

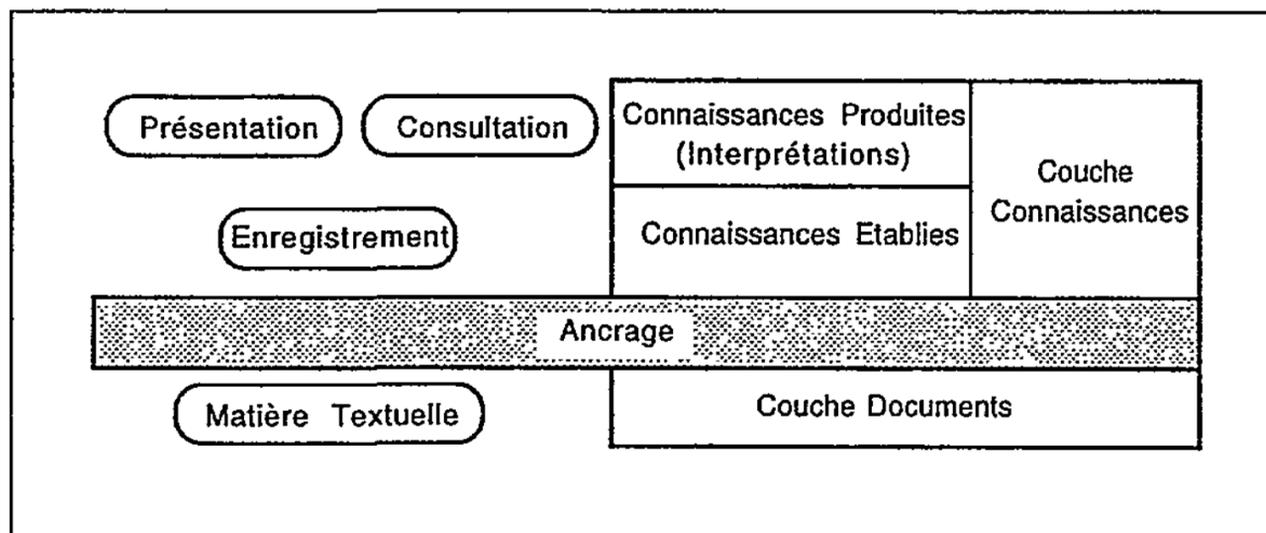


figure 7

Pour établir la correspondance entre les systèmes à3M et à2C, et ainsi rassembler les trois modèles dans une approche globale de l'hypertextualité, de nouveaux concepts doivent être introduits. Sur une base problématique élargie, il s'agira alors de caractériser avec précision les notions de "connaissance (établie)" et "d'interprétation", telles qu'elles sont introduites assez vaguement dans le modèle à2C. Il s'agira aussi d'examiner le statut des structures d'enregistrement et, dans cette voie, le statut des formes constitutives de la textualité. Pour ce faire, nous nous tournerons vers les dictionnaires et nous discuterons les deux approches antagonistes (*exploratoire* et *interprétative*) auxquelles ils se prêtent une fois instanciés sous format électronique. Nous montrerons que les concepts de *forme* et de *substance* apportent une caractérisation précise de ces deux approches et permettent de les conjuguer. Il sera alors possible de conclure (temporairement) le travail entrepris sur la problématique de l'hypertextualité.

## 5. Du dictionnaire à l'hypertexte

### 5. 1. Introduction

Dans un "dictionnaire-papier", les conduites d'exploration sont guidées par la configuration matérielle du texte qui manifeste, plus ou moins précisément, les principaux aspects de la forme interne du fonds lexicographique. Il s'agit, pour l'essentiel, de la classification alphabétique des vedettes, du jeu des renvois et de la mise en page des articles suivant une hiérarchie de marqueurs typographiques. Sous format informatique en revanche, l'accès au contenu lexicographique se règle sur les schémas combinatoires, abstraits et complexes, des formules de requête.

On sait que ces nouvelles modalités d'investigation d'un fonds lexicographique donnent accès à des ensembles d'informations jusqu'à présent hors de portée. Comme l'écrit R. Martin, "(...) déliée des contraintes de la linéarité, la matière du dictionnaire informatisé s'ouvre à des interrogations croisées, à des affichages "intelligents", au regroupement de données éparses, inaccessibles dans le "dictionnaire-papier" " [Martin, 1994, p. 133]. L'informatisation d'un dictionnaire-papier aurait donc pour conséquence première d'en "déployer" de nouveaux espaces d'informations. Ainsi formulé, ce point de vue, communément admis et difficilement contestable, mérite toutefois d'être précisé, car il amalgame deux approches distinctes de l'informatisation des objets textuels, deux approches dont il importe de mettre au jour les principes et les interactions.

## 5. 2. L'approche "exploratoire"

### 5. 2. 1. — Introduction

Suivant une première approche, l'apport de l'informatisation est pensé dans la continuité des pratiques de lecture traditionnelles : les opérations de l'informatique libèrent le lecteur des entraves du support papier, mais les compilations, les comparaisons et les recoupements qu'elles permettent d'effectuer dans des délais raisonnables pourraient tout aussi bien être accomplis par un individu consacré à ces tâches — pour autant qu'il soit patient et doué d'une longévité exceptionnelle. L'établissement d'un fonds lexicographique sur support électronique en augmente donc la profondeur d'exploitation, mais dans le prolongement des voies d'accès "classiques".

Cette approche, que nous dénommerons "exploratoire", suivant laquelle l'informatisation aurait pour seul effet d'accroître les capacités matérielles d'investigation du lecteur, suppose implicitement que les sommes d'informations nouvellement accessibles sont *préalablement présentes* dans le fonds d'origine et, corrélativement, que les procédures de l'informatique restent extérieures aux contenus qu'elles manipulent : elles systématisent et dynamisent les conduites de lecture sur support papier, mais ne redessinent ni ne reconfigurent les contenus du fonds. Ceux-ci sont envisagés comme des données invariantes, comme des informations indépendantes des modes de leur matérialisation et des méthodes mises en œuvre pour les atteindre.

Fondamentalement, l'approche "exploratoire" relève d'une conception *immanentiste* : on fait l'hypothèse que le texte dont on vise la transposition sous format électronique est, en tant que tel, l'ordonnateur des contenus qu'il véhicule. Autrement dit, le texte est une entité pourvue d'une structure propre suivant les régimes de laquelle s'élaborent les significations dont il est doublement l'expression et le principe de

formation. Aussi, quelles que soient les techniques d'investigation employées, le texte offre au lecteur des contenus identiques. Et si les techniques d'accès diffèrent, ce n'est jamais qu'en termes d'efficacité.

La connexion entre la conception immanentiste et l'approche exploratoire semble donc directe. Elle mérite toutefois davantage d'attention. Tout particulièrement, il importe d'identifier avec précision sur quels concepts fondamentaux de la problématique immanentiste se fonde et s'articule la perspective "exploratoire".

### 5. 2. 2. — Forme et Substances

Admettons d'abord, d'une part, que le terme *objet* est synonyme de *objet de connaissance* et, d'autre part, conformément à la position structuraliste, que la connaissance n'a accès qu'à des faits relationnels. A partir de là, le texte pourra être *objectivement* décrit comme une "entité autonome de dépendances internes, ou, en un mot, une structure" [Hjelmslev, 1971, p. 29].

Dans une version plus nuancée de cette problématique, ce n'est que *jusqu'à un certain niveau d'analyse* que l'on pourra identifier dans le texte l'ordre d'une distribution relationnelle univoque et explicite, à savoir une *forme*. Passé ce niveau ultime au-delà duquel la procédure de reconnaissance des formes relationnelles s'avère inadéquate, on rencontre alors le complémentaire de la *forme*, à savoir, toujours suivant une optique immanentiste, une *substance*, dont la reconnaissance (*i. e.* le relevé de la forme interne) requiert l'application d'autres grilles d'analyse. Par exemple, on identifiera aisément dans tel roman ou telle pièce de théâtre une organisation en chapitres et paragraphes ou en actes, scènes et dialogues. On aura ainsi dégagé une forme dont le complémentaire par rapport au texte examiné, à savoir le contenu des paragraphes ou des dialogues, constitue une substance susceptible d'être elle-même soumise à d'autres modes d'analyse (narrative ou phrastique par exemple).

Plus précisément, comme un fait relationnel (à un niveau formel algébrique) comporte génériquement deux facettes, il faudra distinguer les *deux sortes de substance* qui en sont les complémentaires. En effet, considérons par exemple une relation binaire notée  $R(\alpha, \beta)$ , où  $\alpha$  et  $\beta$  représentent les arguments de  $R$ , *i. e.* les pôles de la relation  $R$ . En tant que principe d'organisation abstrait, la relation  $R$  est entièrement déterminée par certains régimes de fonctionnement qui peuvent être des plus divers, par exemple : la transitivité, la réflexivité, etc., ou bien à un niveau topologique et dynamique, l'oppositivité, l'extensivité, etc. Mais, ainsi considérée, sur la base de ses propriétés formelles, la relation  $R$  est dépourvue de spécifications de contenu : elle n'est jamais qu'un principe formateur. Aussi, il y a lieu d'envisager plus globalement les grandeurs qu'elle informe — grandeurs à la constitution desquelles elle participe. En d'autres termes,  $R$  doit être envisagée comme une spécification partielle

d'objets plus complexes, et ce "complément" de complexité par rapport à **R** porte, on a vu, le nom de *substance*. Or un tel complément peut s'établir relativement à **R** de deux façons distinctes. D'une part, ce peut être un "complément de relation". C'est le cas lorsque les caractéristiques relationnelles de **R** déterminent une facette partielle d'un fait relationnel **R** plus global. La "différence" (**R**-**R**) constitue alors la *substance relationnelle* de **R**. D'autre part, ce peut être un "complément d'unité". Il en est ainsi lorsque ce sont les arguments  $\alpha, \beta, \delta...$  de **R**, *i. e.* les unités que cette relation ordonne, qui sont dotés d'un ensemble de spécifications non captées par **R**. Si l'on note **A**, un objet indéterminé et  $\alpha$  l'objet **A** en tant qu'il est déterminé par les relations **R** qu'il contracte, alors la "différence" (**A**- $\alpha$ ) constitue une *substance d'unité*.

Les analyses de contenu telles qu'elles ont été conduites dans le cadre de l'intelligence artificielle (réseaux sémantiques, graphes conceptuels...) ont tout naturellement retrouvé ces deux types de substance. Ainsi, par exemple, dans un réseau sémantique, défini comme une structure formelle (graphe orienté) de liens (relations) et de nœuds (arguments), typer un lien (par exemple : *possède, sorte de, est un...*) ou assigner une étiquette à un nœud (par exemple : *animal, poule, poussin...*) revient bien à reconnaître des surcroûts de complexité qui dépassent les spécifications données par la seule structure relationnelle du réseau. Dans le premier cas (typage des liens), on adjoint au lien une *substance de relation*, dans le second (étiquetage), au nœud, une *substance d'unité*.

S'agissant des dictionnaires, tout du moins dans les cas de figure les plus simples, on retrouvera aussi cette tripartition en une forme et deux substances. Prenons par exemple l'article *bolchevik* dans le *Petit Robert* (on omet les informations étymologiques, phonétiques, ainsi que les datations. Précisons par ailleurs que dans le *Petit Robert*, "[le symbole  $\Rightarrow$ ] suivi d'un mot en gras, présente un mot qui a un grand rapport de sens : 1° avec le mot traité ; 2° avec l'exemple qui précède" :

**BOLCHEVIK** n. [...] 1. Autrefois en Russie, Partisan du bolchevisme\*.  $\Rightarrow$  **maximaliste**. 2. Russe communiste.  $\diamond$  PAR EXT. PÉJ. Communiste. ABRÉV. FAM. BOLCHO. *Les bolchos et les fachos*. «On dit qu'il est maintenant contre les bolchos» (Drieu La Roch.)  $\diamond$  HOM. Bolchevique.

En tant que *forme immanente*, cet article est caractérisé par une distribution hiérarchique à deux niveaux (présentée dans la **figure 8** ci-contre par des traits doubles), des relations à d'autres entrées d'article (flèches), des relations internes orientées (traits fins) et des rapports de portée (implicites). La *substance relationnelle* de cette forme est donnée par le type des relations (entre crochets), et la *substance des unités* (en italique) est en général un texte (définition, exemple...). Deux remarques : (i) les pôles des relations sont ici présentés par des "puces"; (ii) lorsque la relation est unaire, sa substance est un contenu de prédicat appliqué à un

pôle (la relation unaire est représentée par un carré entourant le pôle concerné). Ainsi les indications [nom], [hom.] et [par ext.], par exemple, caractérisent respectivement (i) la substance d'une relation unaire, (ii) la substance d'une relation binaire entre une unité (pôle) du système de l'article et une autre entrée d'article, et (iii) entre deux unités du système.

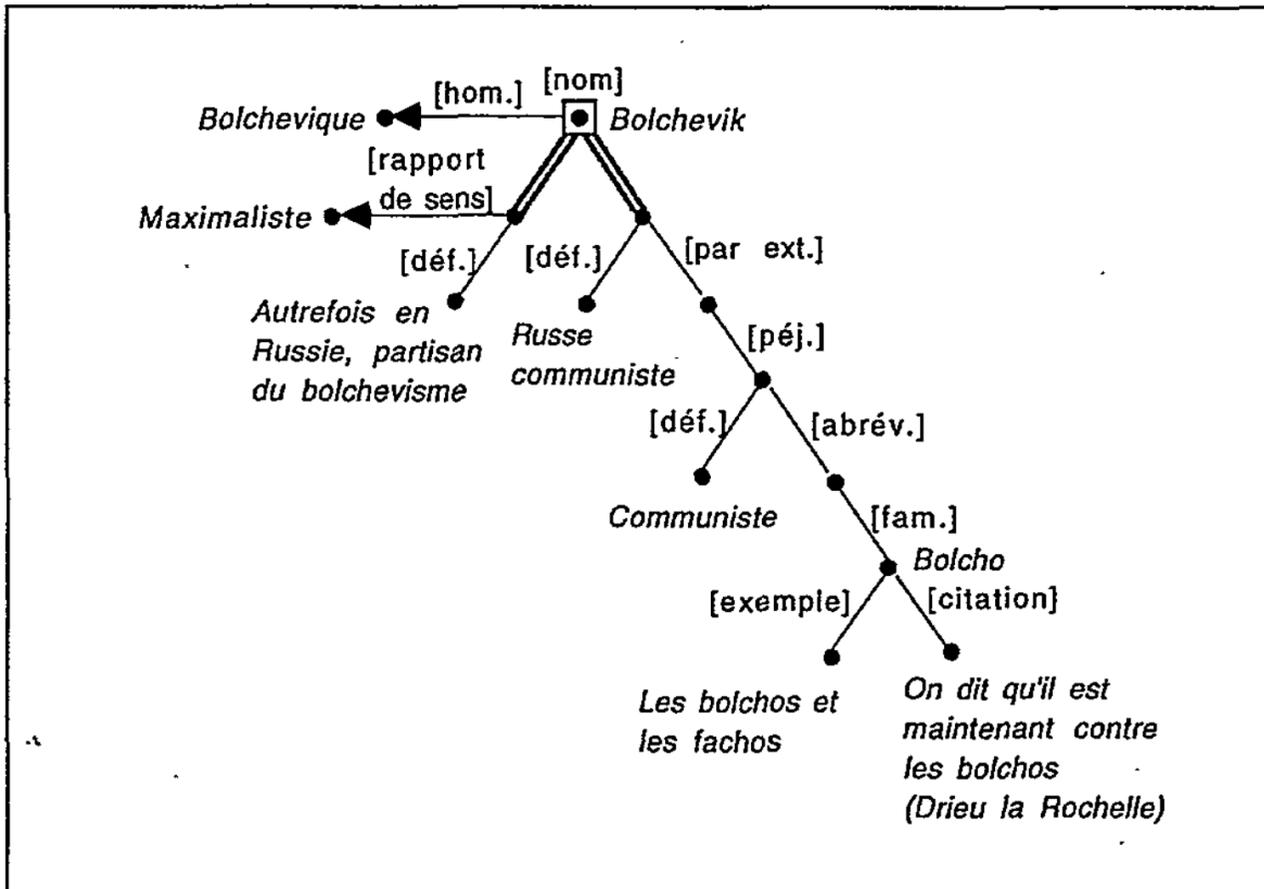


figure 8

En résumé, dans ses fondements, la grille d'analyse de la conception immanentiste articule trois ordres complémentaires : (i) l'ordre de la *forme*, (ii) l'ordre de la *substance de relation* et (iii) l'ordre de la *substance d'unité*.

Sur la donnée de ces trois ordres, on peut maintenant examiner plus précisément quels rapports entretiennent la conception immanentiste et la première approche "exploratoire" du traitement informatique des objets textuels.

### 5. 2. 3. — Base immanentiste de l'approche exploratoire

En premier lieu, on comprend comment la position immanentiste fonde l'approche "exploratoire". En effet, considéré sous l'angle de sa forme interne, donc en tant que mode d'organisation abstrait, un texte est susceptible d'être instancié dans différentes matières empiriques (papier, écran, bande magnétique...). L'informatisation peut alors être considérée comme la projection, la matérialisation de la forme textuelle dans ce substrat particulier qu'est un ordinateur (pour une analyse de la "matière" informatique suivant une optique structuraliste, cf. [Piotrowski, 1993]). Mais qu'il soit réalisé sur support électronique ou sur support papier, le texte en tant qu'objet relationnel reste inchangé : ses caractéristiques

objectives sont conservées. En conséquence, comme le texte informatisé ne diffère en rien du texte imprimé, les procédures automatiques d'exploration ne sauraient atteindre des éléments d'information différents de ceux préalablement présents dans la version papier. Les opérations de l'informatique permettent de procéder à des fouilles complexes, exhaustives et rapides dans le fonds enregistré mais, en termes de contenu, la lecture électronique reste dans le prolongement de la lecture traditionnelle.

Mais un fonds lexicographique ne se réduit pas à sa seule forme interne. Il faut aussi tenir compte des deux compléments de substance. Il conviendra donc, en second lieu, d'examiner de quelle façon le dispositif informatique "accueille" les grandeurs de substance. Trois points devront être abordés, concernant respectivement (i) les modalités d'instanciation des substances, (ii) leur position dans le système de traitement et (iii) leur rôle fonctionnel.

Pour traiter ces différents points, nous choisirons comme référence le mode d'une implémentation en base de données relationnelles (cf. *supra*). Par exemple, l'article *bolchevik* (cf. *supra*) peut être représenté par les six tables relationnelles suivantes (on fait le choix des attributs suivants : Vedette (**Ved**), Catégorie Grammaticale (**Cat**), Niveau Hiérarchique (**Nv**), Définition (**Déf**), Exemple (**Ex**), Citation (**Cit**), Homonyme (**Hom**), Par Extension (**Ext**), Abréviation (**Abrév**), Familier (**Fam**), Péjoratif (**Péj**) et Rapport de Sens ( $\Rightarrow$ )) :

Ved	Cat	Nv	Hom	Déf
bolchevik	nom	0	bolchevique	autrefois en Russie, partisan du bolchevisme Russe communiste
bolchevik	nom	1		
bolchevik	nom	2		

Tableau 1

Ved	Nv	$\Rightarrow$
bolchevik	1	maximaliste

Tableau 2

Ved	Nv	Ext	Péj
bolchevik	2.1	communiste	communiste

Tableau 3

Ved	Nv	Abrév	Fam
bolchevik	2.2	bolcho	bolcho

Tableau 4

Ved	Nv	Ex
bolchevik	2.2	les bolchos et les fachos

Tableau 5

Ved	Nv	Cit
bolchevik	2.2	On dit qu'il est maintenant contre les bolchos

Tableau 6

Les comparaisons ligne à ligne de la valeur des différents attributs (ici, tout particulièrement, celle des attributs **Ved** et **Nv**) permet alors de rassembler les informations dégroupées dans les six tableaux pour reconstituer le texte structuré d'origine. En d'autres termes, l'ensemble des tableaux associés à un article constitue une réalisation particulière de sa forme interne.

L'établissement d'un article de dictionnaire sous forme d'un ensemble de tables permet de traiter les trois points précédents. D'une part, concernant l'instanciation des substances, on observe que les substances d'unité (par exemple, la dénomination des unités de traitement et les contenus d'information) se réalisent par des valeurs d'attribut et sont codifiées sous forme de chaînes de caractères. Les substances de relation sont, quant à elles, instanciées comme des étiquettes d'attribut (et codifiées aussi sous forme de chaînes de caractères). D'autre part, concernant la position des deux sortes de substance, on constate qu'elles sont clairement *disjointes et entrecroisées* : les grandeurs de l'une occupent des positions génériques (tête des colonnes), celles de l'autre, des places de valeur (lignes). Enfin, pour ce qui est des rôles fonctionnels, ils dérivent directement de la position que les substances occupent en structure. En effet, distribuées suivant le schéma de couples attribut/valeur qui reproduit l'organisation interne de l'article, les grandeurs des deux substances se prêtent à jouer le rôle de paramètres d'entrée dans des procédures de recherche croisée d'informations. Du fait de la tripartition en une forme et deux substances, les recherches peuvent alors porter sur trois sortes d'information. Par exemple : (i) recherche d'information sur la seule forme interne : quels sont les articles dont chaque nœud de la structure hiérarchique est le pôle d'une relation interne ? (ii) recherche

d'information sur la substance relationnelle : quels sont les articles dont une acception *péjorative* possède une abréviation dans un registre *familier* ? (iii) recherche d'information sur la substance d'unité : quels sont les articles dont une définition contient le terme *partisan* ? Par ailleurs, il sera bien évidemment possible de conjuguer ces trois types de recherches.

#### 5. 2. 4. — En conclusion

L'approche "exploratoire" procède bien d'une conception immanentiste. En premier lieu, l'hypothèse d'une *forme lexicographique interne* permet de concevoir le dictionnaire informatisé comme l'instanciation *conforme* d'une entité abstraite. Partant, les opérations de l'informatique appréhendent des données identiques à celles qui sont matérialisées dans la version papier. En second lieu, complémentaires d'une forme interne univoque, les substances du dictionnaire (relationnelles et d'unité) sont elles-mêmes déterminées avec précision dans leur nature et leur distribution. La connexion de ces trois types de données (de forme et de substance) fonde alors les procédures exploratoires : celles-ci se règlent sur les catégories de la substance relationnelle, les contenus de la substance d'unité et les régimes d'organisation de la forme.

Sous cet angle, un dictionnaire informatisé constitue en définitive une base de connaissances sur le lexique, c'est-à-dire un recueil d'informations (substances) structuré suivant des régimes univoques explicites (forme). On est ici en présence de ce que l'on pourrait dénommer *l'idéal analytique* de la lexicographie.

### 5. 3. L'approche "interprétative"

#### 5. 3. 1. — Introduction

La deuxième approche met l'accent sur les nouveaux rapports aux textes qu'introduit le dispositif informatique. Contrairement à l'approche précédente où l'informatisation est pensée en continuité des schémas de lecture traditionnels, la transposition sous format électronique se trouve maintenant à l'origine d'un renouvellement des rapports entre le texte et le lecteur. Dans cette optique, les fonctionnalités qui administrent la lecture sur support informatique sont à la source d'interprétations originales. Il est clair que, tout du moins dans sa version radicalisée, cette seconde approche s'oppose à l'hypothèse immanentiste : le texte n'est pas supposé investi d'une structure autonome qui le caractérise *en soi*. Son contenu ne se dévoile que dans le processus d'appropriation qu'est la lecture : "le texte [est maintenant pensé comme] un objet qui ne se constitue que dans les conditions d'accès qu'y ménage son support matériel, conditionnant la nature de la relation du lecteur à ce qu'il lit par une mise en forme

matérielle (...)” [Stiegler, 1994, p. 45]. Les nouvelles modalités d'accès au texte que fournit l'informatique et qui déterminent de nouvelles formes de compréhension, vont donc aboutir à des interprétations radicalement nouvelles. Entre le fonds dans sa forme imprimée et sa version informatisée, il n'y a plus cette fois continuité, prolongement, mais bien rupture, reconfiguration.

Dans l'absolu, les deux précédentes approches sont manifestement antagonistes. Mais elles ne sont pas pour autant exclusives et peuvent être appliquées conjointement au même objet. En effet, les textes offrent généralement le double visage de la systématité et de la maléabilité, de l'univocité et de l'ambiguïté... En termes plus précis, les textes se prêtent à de multiples saisies, chacune d'elles en révélant une forme particulière et, complémentaiement, une certaine substance. L'interprétation revient donc à reconnaître dans le texte l'existence d'une systématité qui en détermine les substances. Et il s'ensuit que chaque lecture, chaque moment interprétatif, pour autant qu'il suive des modalités explicites, détermine un univers de données textuelles susceptible d'être investi à travers des procédures exploratoires.

### 5.3.2. — L'exemple des dictionnaires

Un ouvrage d'importance comme le *Trésor de la Langue Française (TLF)* est de toute évidence passible de différentes lectures. Sa matière textuelle ne se présente pas sous une forme arrêtée : elle peut faire l'objet de différentes structurations qui relatent autant d'actes de formation et d'appropriation de ses contenus, en d'autres termes, autant d'actes de lecture et de saisie d'information.

Pour mettre en évidence ce caractère de “maléabilité” d'un fonds lexicographique, reprenons plus précisément le schéma d'organisation en une forme et deux substances précédemment introduit.

De façon générale (cf. exemple de l'article *bolchevik*), la substance des pôles (substance d'unité) correspond aux deux principaux types de données des dictionnaires sémasiologiques, à savoir (i) les unités de traitement et (ii) les contenus d'information concernant ces unités. La substance de relation est quant à elle caractérisée par les différents types d'indicateurs métalinguistiques (indicateurs grammaticaux, sémantiques, stylistiques, d'usage, de niveau de langue, de domaine...) ou par les indicateurs des types d'information lexicographique (définition, exemple, syntagmes types, synonymie-antonymie...).

L'approche “exploratoire” des dictionnaires suppose donc que le fonds est administré par une structure univoque qui, en corollaire, détermine des compléments en substance tout aussi univoques. Mais, un examen empirique contredit ce tableau. Ainsi, dans une étude minutieuse de la structure des articles du *TLF*, F. Henry relève de nombreuses “interférences” entre les différents types de composantes (unités de

traitement, indicateurs métalinguistiques et autres éléments d'analyse). Prenons par exemple le fait des "interférences entre [les] unités de traitement internes et d'autres éléments d'analyse" [Henry, 1993, p. 185], et limitons-nous au cas où ces derniers sont des indicateurs de construction. F. Henry note que "les mêmes faits ou des faits très similaires peuvent faire l'objet tantôt d'une présentation de type construction, tantôt d'une présentation de type unité de traitement interne" [*ibid.*, p. 193]. En d'autres termes, une même matière textuelle peut être appréhendée suivant deux modalités de mise en forme distinctes. F. Henry propose à ce titre de comparer les deux exemples suivants [*ibid.*, p. 193] :

**OBLIGATION** (niveau C))

- *Veilli. Avoir obligation (de qqc. à qqn)*

Etre redevable (de bons offices, de quelque bienfait à quelqu'un).

**ORDRE** (niveau III A 1 a)

• *Avoir, recevoir ordre de + inf. Avoir reçu la mission, la consigne de.*

On voit bien que la sous-structure de l'article *Ordre* présentée ci-dessus pourrait être réécrite en faisant usage d'un indicateur de construction similaire à celui de l'article *Obligation*, soit :

**ORDRE** (niveau III A 1 a)

• **Avoir, recevoir ordre de + inf.**

Avoir reçu la mission, la consigne de.

Dans ce cas, l'unité de traitement interne "*Avoir, recevoir ordre de + inf*" est remplacée par une indication de construction **Avoir, recevoir ordre de + inf**. Cette reconfiguration de la forme d'article et la redistribution des substances appariées peut être mise en évidence par les schémas suivants (**figure 9**) (on conserve la codification utilisée dans l'exemple de l'article *bolchevique*, cf. *supra*) :

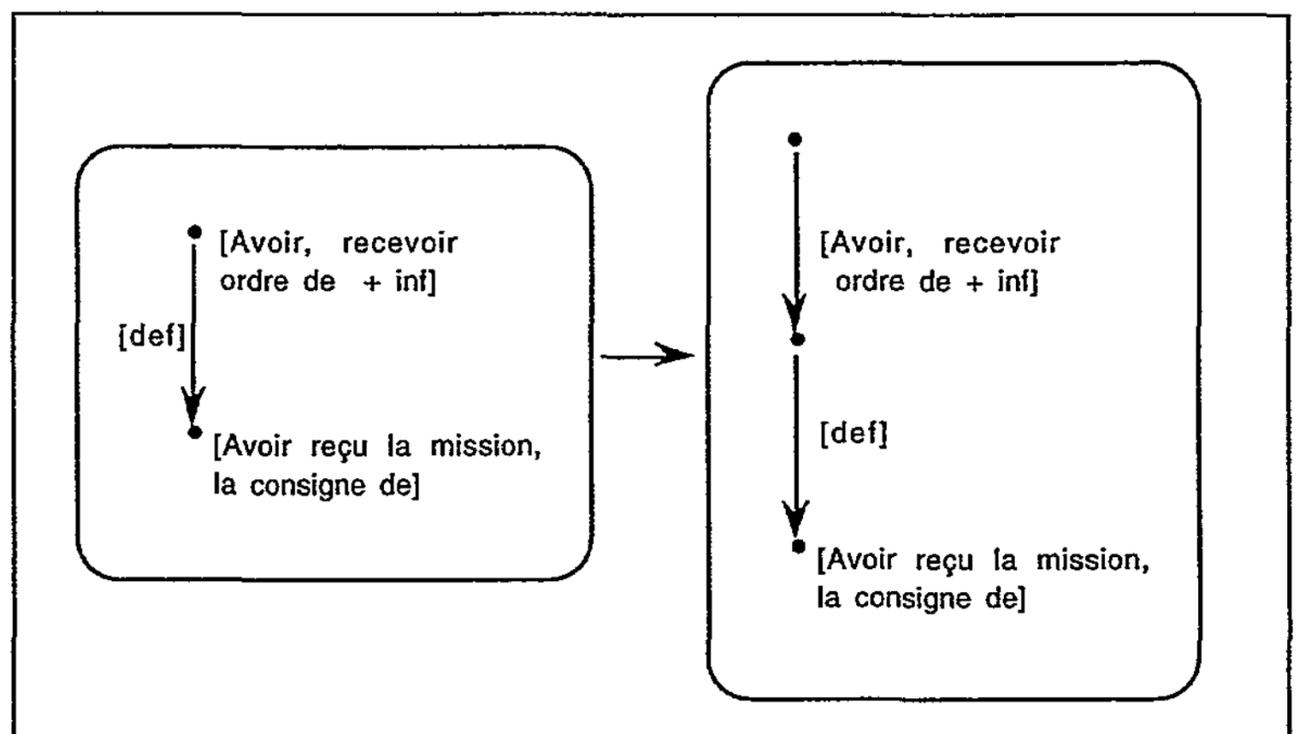


figure 9

Ainsi, force est de constater dans le dictionnaire de nombreux glissements et flottements qui ne peuvent se résoudre qu'à travers des actes d'interprétation, autrement dit, à travers des processus variés de mise en forme du texte et de détermination des substances.

## 6. Conclusion

On a vu que les concepts de *forme* et de *substance* permettent de dépasser l'antagonisme entre les approches "exploratoire" et "interprétative" des dictionnaires informatisés. C'est aussi à partir de ces deux concepts qu'il sera possible d'établir une problématique unifiée de l'hypertextualité (recouvrement des modèles dD, à3M et à2C) et, corrélativement, d'avancer vers la spécification hypertextuelle d'un fonds lexicographique. En effet, les qualifications en termes de *forme* et de *substance* permettent d'opérer une double jonction : (i) entre le modèle à3M et le modèle à2C, et (ii) entre le schéma général d'une dynamique d'appropriation des textes sur support électronique, telle qu'elle est exposée dans le modèle à2C, et la matière d'un dictionnaire. Nous examinerons cela tour à tour.

### 6 1. Unification des problématiques de l'hypertextualité

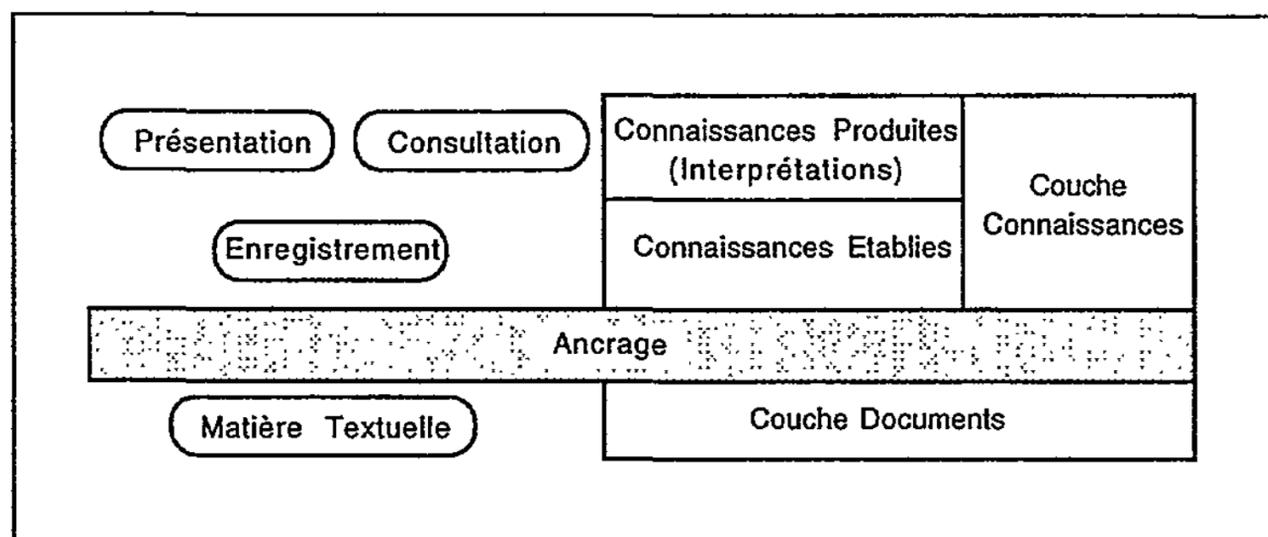
Les notions de forme et de substance déterminent adéquatement la boucle récursive des interprétations du modèle à2C. La *connaissance*, à un stade donné des processus d'appropriation de la couche des "documents", est caractérisée par une distribution univoque et réglée des contenus, en d'autres termes, par une certaine *forme* appariée à certaines *substances*. Et c'est sur la base de cette forme, qui par nature se prête aux opérations de l'informatique, ainsi que sur la distribution croisée des substances *relationnelles* et d'*unité* (cf. *supra*), que le système hypertextuel tisse un réseau de liens et procède à des calculs d'inférence en vue de suggérer au lecteur des cheminements conséquents dans le fonds informatisé. Par ailleurs, au stade de l'*interprétation*, pour apporter un nouvel "éclairage" au fonds, l'utilisateur va instaurer de nouveaux liens entre les éléments extraits au cours de la navigation ; en d'autres termes, il va instancier une nouvelle forme et, corrélativement, dégager de nouvelles substances.

Dans le modèle à2C, la couche des connaissances se subdivise donc en deux sous-couches : la sous-couche des formes et des substances "établies" (les connaissances posées en référence) et la sous-couche des formes et des substances "produites" (les interprétations). La sous-couche des connaissances "établies" constitue ainsi une distribution "arrêtée" (momentanément) de formes et de substances sur laquelle prend appui

l'activité de lecture. C'est sous cet angle qu'elle est l'analogie du module d'enregistrement du modèle à3M. En effet, le module E n'est pas un réceptacle passif accueillant des données brutes, mais un dispositif consignait une matière textuelle à travers une certaine détermination de ses formes et de ses substances. A l'instar de la sous-couche des connaissances "établies", le module E délivre donc une certaine compréhension de la matière textuelle qu'il renferme. Corrélativement, il renferme un système "figé" de spécifications relationnelles qui contribue à guider les parcours d'investigation des données.

Par ailleurs, c'est au niveau de la sous-couche des interprétations (du modèle à2C) que se développe l'activité d'appropriation du texte. C'est donc à ce niveau que le lecteur met en œuvre des procédures de recherche, de visualisation et de manipulation des données. La sous-couche des interprétations inclut donc les modules fonctionnels de "présentation" et de "consultation" — tels qu'ils sont définis dans le modèle à3M.

La figure suivante (antérieurement introduite) illustre la concordance entre les modèles à2C et à3M :



*En conclusion* : les problématiques des modèles dD, à3M et à2C se recouvrent en une vision unifiée de l'hypertextualité (la correspondance entre les modèles dD et à3M a été discutée en 2. B. 2. et 2. C. 2.). Les trois modèles conjuguent trois éclairages compatibles et complémentaires d'un même fait empirique : la lecture sur support électronique.

## 6. 2. Spécification hypertextuelle d'un dictionnaire électronique

Disposant maintenant d'une base problématique unifiée, il est légitime d'aborder la question des fonctionnalités hypertextuelles d'un dictionnaire.

Installer un fonds lexicographique sous format hypertextuel revient donc, pour l'essentiel, à l'inscrire dans un système d'interactions distribué sur trois niveaux, à savoir, à la base, la matière textuelle, au sommet, le lecteur en tant que sujet interprétant (mettant en œuvre les fonctionnalités des modules P et C), et au milieu, un système de connaissances "établies"

(ou module E) compatible avec les opérations de l'informatique. On sait par ailleurs que les fonctionnalités hypertextuelles d'un dictionnaire électronique doivent assister le lecteur dans son travail d'interprétation. Ce qui requiert deux types d'opérations : en *lecture* et en *écriture*. Nous précisons ces deux points.

D'une part, le système hypertextuel doit proposer des parcours de lecture qui favorisent des collectes consistantes de données. C'est, on a vu, le rôle de la couche des connaissances "arrêtées" (ou de la structure du module E) qui "éclaircit" la couche des documents. On notera que l'on retrouve ici les finalités de l'approche "exploratoire".

D'autre part, le système doit mettre à disposition de l'utilisateur des outils d'aide à l'interprétation, c'est-à-dire un ensemble de commandes destinées à produire de nouveaux rapports relationnels (*formes*) et, complémentirement, à instituer de nouvelles *substances*. Il s'agit, à travers une telle grille de fonctionnalités, de *canaliser les processus d'interprétation suivant des schémas de mise en forme cohérents et explicites*, et ainsi d'assister la production de nouveaux réseaux de qualification des données. Les outils d'écriture ne se réduisent donc pas à des commandes d'indexation ou d'annotation ; ce sont fondamentalement des instruments de remodelage d'une matière textuelle offerte au lecteur à travers la fenêtre "active" du système hypertexte.

Complémentaire aux fonctionnalités de lecture (navigation raisonnée), les fonctionnalités d'écriture visent à *objectiver*, c'est-à-dire à établir suivant des formats de structure et d'opération rigoureux et univoques, les processus et les résultats d'interprétations qu'accomplit le lecteur. Les outils d'écriture se situent donc au cœur du système hypertextuel conçu "comme dispositif potentiel favorisant l'ouverture de nouveaux espaces sémantiques" [Vignaux, 1995, p. 8].

C'est sur cette base que s'établit la corrélation entre les systèmes hypertexte et la matière lexicographique. Tout simplement parce que les concepts de *forme* et de *substance* subsument un matériau lexicographique précis (cf. *supra*) et se prêtent donc à une mise en œuvre fonctionnelle effective.

Plus concrètement, du côté des fonctions de lecture, on trouvera bien sûr les différents types de commandes qu'une informatisation "classique" (sous forme de base relationnelle par exemple) est en mesure de fournir (à savoir, pour l'essentiel, des recherches multicritères) doublées de procédures plus sophistiquées, issues des avancées de la linguistique formelle et de ses applications à l'informatique (par exemple, structuration sémantique et syntaxique des contenus).

Du côté des fonctions d'écriture, il faudra concevoir des modalités d'assignation de relation, de regroupement de données, de déplacement de contenus... qui permettent de transcrire des résultats de lecture sous une forme appropriée au sens que souhaite leur conférer l'usager. On voit de

quoi il peut s'agir à travers le cas très simple de la reconfiguration de la forme et des substances d'un fragment de l'article *ordre* (cf. *supra*). En effet, il est tout à fait concevable qu'un lecteur, pour homogénéiser ses données en vue de synthèses ou de confrontations cohérentes, éprouve le besoin d'ajuster ses différentes informations, c'est-à-dire éprouve le besoin de redessiner les formes et de recalibrer les substances des différentes portions d'articles qu'il a retenues. A cette fin, il sera utile, par exemple, qu'il dispose d'outils d'écriture lui permettant de transformer la structure de l'article *ordre* en établissant l'unité de traitement interne *Avoir; recevoir ordre de + inf* au rang d'indicateur de construction.

Suivant une même optique, de tels outils devront permettre à l'utilisateur de "travailler" ses données pour l'aider à mettre progressivement en évidence des jeux de rapports sémantiques (recouvrement, opposition, complémentarité, extensions, etc. dans différents domaines) qui peuvent s'établir, par exemple, entre quasi-synonymes. Un simple coup d'œil sur les principales articulations des articles *laver* et *nettoyer* du *TLF* (ci-contre, en **Annexe**) suffira sans nul doute à entrevoir la nature des opérations d'ajustement dont il est ici question : on observe que les vocables *laver* et *nettoyer*, qui partagent une substance d'expérience commune, relèvent de régimes linguistiques distincts. Leur inscription, à des fins interprétatives, dans un système qui explicite le rendement sémantique de leur juxtaposition, suppose des fonctionnalités permettant de rassembler et de reverser les différentes informations lexicographiques dans un cadre homogène dont la définition et l'élaboration, assistées par le système hypertexte, incombent entièrement au lecteur.

(CNRS-INaLF)

## Annexe

## NETTOYER

I. - [Le compl. d'obj. dir. désigne un subst. concr.]

A. - Rendre net, propre en débarrassant de tout ce qui salit, souille, ternit.

B. - *P. ext.*

1. Rendre net en débarrassant de tout ce qui gêne, encombre.

a) *En partic.*b) *Spécialement*

- AGRIC., SYLVIC.

- TECHN. MILIT.

2. Rendre les contours de quelque chose nets et distincts.

3. *Fam.*II. - *Au fig.*

## LAVER

I. A. - [Le compl. désigne ce qui est sale]

1. [Le suj. désigne l'agent]

a) [Avec un compl. prép. indiquant le moyen]

b) [Avec un compl. prép. locatif *dans* ou *à*]

c) [Sans compl. prép.]

d) [Avec un compl. prép. *de* désignant ce qui salit]e) *Emploi abs.*

2. [Le suj. désigne le moyen]

a) *P. anal.*b) *P. métaph.*

B. - [Le compl. désigne ce qui salit]

1. [Le suj. désigne l'agent]

2. [Le suj. désigne ce qui lave]

II - BEAUX-ARTS

III - *Fam. et arg.*

**Bibliographie**

BALPE (J.-P.)

1990, *Hyperdocuments, hypertextes, hypermédias*, Eyrolles.

BRUILLARD (E.), LA PASSARDIERE (B. de)

1994, "Hypermédias et éducation : des repères", *STE*, vol. 1, n°1.

HALASZ (F.), SCHWARTZ (M.)

1994, "The Dexter Hypertext", *Communications of the ACM*, vol. 37, n° 2.

HENRY (F.)

1993, "Analyse descriptive du *Trésor de la Langue Française*", in *Pour l'informatisation du Trésor de la Langue Française*, INaLF-Nancy.

HJELMSLEV (L.)

1971, *Essais linguistiques*, Minuit.

IDE (N.), LE MAITRE (J.), VERONIS (J.)

1991, *Outline of a Model for Lexical Databases*, Rapport n° 496 du GRTC, CNRS, Marseille.

LAUFER (R.), SCAVETTA (D.)

1992, *Texte, hypertexte, hypermédia*, PUF (Que sais-je ? ; n° 2629).

MARTIN (R.)

1994, "Présentation", *Le Français moderne*, t. LXII, n° 2.

NANARD (M.)

1993, "L'Interface des mondes réels et virtuels", *Cours n° 10, INFORMATIQUE' 93* (22-26 mars, Montpellier), France.1994, "L'Apport des travaux de recherche dans les hypertextes aux techniques éducatives", in *Actes du séminaire Hypermédia, Éducation et Formation*, E. Bruillard, B. de La Passardière, G. L. Baron, éd., IUFM Créteil.

PIOTROWSKI (D.)

1993, "Structuralism, Computation and Cognition : The Contribution of Glossematics", in *The Computer as Medium*, P. B. Andersen, B. Holmqvist, J. F. Jensen, éd., Cambridge University Press.1994, "Dictionnaires électroniques : questions d'architecture", *Le Français moderne*, t. LXII, n° 2.1996, "Opérations hypertextuelles et formes lexicographiques", in *Lexicographie et informatique : autour de l'informatisation du TLF*, D. Piotrowski, ed., Paris, Didier-Érudition.

STIEGLER (B.)

1994, "Machine à écrire et matières à penser", *Genesis*, n° 5.

VIGNAUX (G.)

1992, *Les Sciences cognitives : une introduction*, La Découverte.1995, "Spécification hypertextuelle d'un dictionnaire électronique", in *Rapport de Recherche du programme CNRS «Cognition, Communication intelligente et Ingénierie des langues», Axe III : «Hypertextes et Dictionnaires»*.