

Chapitre 5

L'indexation multimédia

5.1. Introduction

L'indexation est traditionnellement associée à la recherche d'information dans la mesure où la pose d'index permet d'améliorer la recherche de documents répondant à une requête donnée. Les concepts et méthodes de l'indexation se sont principalement constitués dans un contexte où les documents textuels, sur des supports de type papier, sont prédominants. Sous l'effet de la numérisation des contenus et de l'informatisation de leur exploitation, le cadre conceptuel et méthodologique doit évoluer profondément pour affronter des problèmes inédits et tirer partie des nouvelles possibilités offertes par le cadre technologique émergent. Plusieurs faits sont à l'origine de ces mutations :

– la dématérialisation des contenus : la notion classique de document s'est élaboré dans un contexte où le support permettant d'enregistrer et de conserver un contenu est le même que celui qui permet de le restituer (le papier que je lis est le même que celui que je range). Dans le contexte numérique de dématérialisation des supports, le support de conservation (le disque dur) n'est pas celui qui permet de restituer le contenu (l'écran, ou le papier imprimé). Entre les deux s'est introduit la médiation d'un dispositif reconstruisant le document à lire à partir du document enregistré. Si bien qu'au fondement même du document numérique gît la problématique de publication d'un contenu à partir de son enregistrement. C'est la raison pour laquelle l'évolution du document papier au document numérique a pour conséquence l'évolution d'une indexation dévolue à la recherche d'information à une indexation conçue pour la publication électronique ;

– l'intégration de différents médias sur un même support : jusqu'au numérique, chaque média était confiné sur un support dédié sans avoir d'interaction avec les autres médias. L'audiovisuel, enregistré sur un support magnétique (la cassette vidéo) ou argentique (le film), n'était associé à aucune autre information : textuelle, graphique, etc. Réciproquement, un texte pouvait difficilement inclure des photos (par exemple, des pages de qualité différente au milieu d'un livre) et encore moins de l'audiovisuel. Le numérique permet à chaque média de sortir de son splendide isolement. Outre des problèmes techniques nouveaux, cette nouvelle situation renvoie à des difficultés fondamentales sur l'écriture et la lecture multimédias : comment intégrer différents médias dans une écriture pensée pour une lecture ? L'indexation prend dans ce contexte le rôle d'une instrumentation pour la lecture : l'indexation ne sert pas tant à retrouver de l'information qu'à l'organiser pour la lecture ;

– la numérisation des objets temporels : par objet temporel, il faut comprendre les objets sonores et audiovisuels, qui imposent le rythme et l'ordre de la lecture. Au lieu de se construire dans un espace comme un texte, les objets temporels se construisent dans et par une durée. La numérisation de ces objets permet de revoir l'indexation : en effet, dans un contexte classique analogique, l'indexation audiovisuelle est essentiellement une identification du document (que l'on appelle souvent « catalogage ») et une description globale du contenu (de quoi ça parle) sans qu'il y ait de description segment par segment. En effet, l'accès technique à une partie quelconque du document étant coûteuse et difficile dans ce contexte analogique, il n'est d'aucune utilité de référencer et localiser finement le contenu sur son support puisque de toutes manières il faut avoir le document en entier pour consulter l'une de ses parties. Le numérique permet d'avoir un accès aléatoire au contenu, en pratique à une partie du document sans avoir à consulter l'ensemble du document. Cette possibilité amène à revoir profondément indexation et documentation.

L'indexation adopte par conséquent de nouvelles méthodes pour de nouvelles finalités. Les méthodes, ce sont les moyens d'associer des index à des contenus multimédias et des objets temporels. Les finalités renvoient à l'organisation du contenu, à l'aide de ces index, pour sa publication et sa lecture. La recherche d'information ne devient qu'un cas particulier, l'une des tâches nécessaires pour publier et lire des contenus multimédias.

On comprendra par conséquent que l'indexation multimédia est avant tout un problème conceptuel avant de renvoyer à des difficultés techniques. C'est pourquoi ce chapitre se concentre dans un premier temps à bâtir un cadre dans lequel poser les enjeux d'une indexation multimédia. Nous verrons ainsi successivement :

– qu'est-ce qu'un document ? Nous y définissons les différents types de document et en particulier nous introduisons la notion de métadonnée ;

– qu'est-ce que l'indexation ? Nous y définirons les différents types d'indexation, à savoir l'indexation conceptuelle, documentaire et par le contenu ;

– qu'entraînent les possibilités multimédias du numérique ? En particulier quelles sont les difficultés posées par l'indexation d'un objet temporel, d'une image, d'un flux audiovisuel ? Quels sont les différents niveaux d'indexation, selon sa propre nature sémiotique et celle du contenu indexé (par exemple indexer un son par un son ou un mot) ?

Nous rassemblerons les conclusions auxquelles aboutissent ces considérations conceptuelles dans une dernière partie consacrée à l'indexation audiovisuelle et multimédia telle qu'elle est abordée dans des normes émergentes comme MPEG-7.

5.2. Document

L'indexation permet d'instrumenter des documents pour les exploiter dans le cadre de services et usages allant de la recherche d'information à la publication éditoriale en passant par la navigation et la consultation de bases documentaires. Pour mieux appréhender la complexité inhérente de l'indexation documentaire, et des problèmes rencontrés quand on aborde des contenus multimédias, il convient de s'interroger sur la nature des documents et de leurs principales caractéristiques.

Étymologiquement, « document » vient du latin *doceo*, « j'enseigne ». Principalement destiné à fixer une information permettant de justifier un droit ou un contrat, le document est originellement utilisé dans un cadre juridique comme trace permettant d'administrer une preuve. A présent, un document a une vocation plus large, mais il conserve de ses origines la fonction d'inscrire dans la permanence une information pour des consultations futures. Mais il possède d'autres propriétés ou fonctions qu'il nous faut dégager. Risquons pour cela une définition dont nous analyserons les composantes. Un document, c'est un *contenu* institué par un *acte de publication*, inscrit sur un support matériel, et possédant une *délimitation temporelle* et *spatiale*. Précisons à présent.

Contenu

Un contenu est une forme inscrite sur un support se prêtant à une interprétation à travers laquelle elle fait sens pour quelqu'un ou une communauté.

C'est donc une forme matérielle interprétable. C'est d'ailleurs pour cette raison que le terme de contenu sert parfois à qualifier non pas la forme interprétable, mais son sens, c'est-à-dire le résultat de son interprétation. Si l'on parle de contenu, la notion de « contenant » n'est pas loin. Le contenant sera dans ce contexte d'une part tout ce qui permet de fixer ou réifier en un format manipulable le contenu, et d'autre part les outils pour manipuler ce format et le produire, le reproduire, le transformer

et le transmettre. Par exemple, un contenu pourra être un texte écrit, en considérant particulièrement la forme matérielle des lettres, de la mise en page et autres déterminations de ce qui fait sens. Le contenant sera le papier, l'encre, et les outils permettant d'imprimer, reproduire, etc.

Acte de publication

Une forme matérielle interprétable est un document quand elle est mise à disposition d'un public, même potentiel, par son auteur ou créateur. Cette mise à disposition institue la forme matérielle comme étant une forme *interprétable* car émanant d'une intentionnalité.

Délimitation spatiale

Pour qu'un contenu soit un document, il doit être possible de déterminer avec certitude où il commence, où il finit, ce qui lui appartient ou ce qui ne lui appartient pas.

Cette caractéristique est essentielle car elle constitue la possibilité de la lecture et de l'interprétation : on ne peut lire un contenu que lorsqu'on peut faire la part entre le document lu et la lecture qui en est faite, entre le discours et le métadiscours que l'on tient à son endroit. C'est à partir de la finitude documentaire que commence l'interprétation, interprétation qui, quant à elle, n'est jamais terminée. La finitude et fermeture documentaire renvoient à l'ouverture indéfinie de l'interprétation.

Délimitation temporelle

Un document doit être persistant dans le temps. Autrement dit, la forme matérielle interprétable le constituant ne doit pas varier en fonction du moment où s'effectuent la consultation et la lecture. Toute modification du contenu tient lieu d'une republication du document.

C'est en cela aussi qu'un contenu n'est pas nécessairement un document. Un contenu ne devient un document seulement quand une intention le fixe et le livre à la consultation d'un public. Un contenu renvoie à une écriture, non à une publication : le modifier n'est pas le republier. Les usages du monde de l'édition le confirme : un livre publié n'est plus modifiable, il doit être réédité pour cela.

Enfin, le document est inscrit sur un support. L'articulation du contenu au support sur lequel il est inscrit varie en fonction de la technologie mise en œuvre pour l'élaboration du support et le procédé d'inscription utilisé. Cette articulation détermine le type du document et ses propriétés. Plusieurs dimensions permettent de décrire cette articulation.

Le support d'enregistrement du document

Un document est l'inscription matérielle d'un contenu. Le support d'enregistrement est par conséquent le support d'inscription, le support que l'on charge de préserver et conserver le contenu inscrit. Par exemple, un livre a pour support d'enregistrement le papier, un document audiovisuel le support d'une bande magnétique vidéo, ou le support d'un film argentique, le document numérique enfin la mémoire adressable d'un support informatique.

La forme d'enregistrement

Le support d'inscription permet de consigner le contenu dans une certaine forme. La forme d'enregistrement est la forme sous laquelle le contenu est inscrit sur le support d'enregistrement. Pour un livre, c'est la typographie d'un répertoire alphabétique, pour un document audiovisuel, le signal magnétique sur le support vidéo, ou enfin le code numérique binaire des documents numériques.

Le support d'appropriation du document

Le document ne préserve et ne conserve un contenu que pour le rendre accessible et public, puisqu'il résulte d'un acte de publication. Un document doit par conséquent pouvoir être lu, consulté, visualisé sur un support le permettant. Ce support est celui où un utilisateur peut faire sien le contenu, se l'approprier. Un support d'appropriation est l'écran de télévision, le papier du livre, un haut-parleur, etc.

La forme physique d'appropriation du document

Le support physique présente le contenu. Pour cela, ce dernier doit être présenté dans une forme physique compatible avec le support physique, pour que le contenu puisse être appréhendé par l'utilisateur. Par exemple, à partir du signal magnétique (forme d'enregistrement) de la cassette vidéo (support d'enregistrement), la télévision (support d'appropriation) reconstruit un signal visuel (les points de couleurs de l'écran) regardé par le spectateur.

La forme sémiotique d'appropriation du document

La représentation affichée sur le support d'appropriation respecte une structure ou une forme telle qu'elle est directement intelligible par l'utilisateur. Cette forme est donc la forme d'appropriation permettant à un utilisateur de s'approprier le contenu. La forme d'appropriation correspond à une forme directement interprétable par un utilisateur, dans la mesure où elle appartient à un registre sémiotique dont l'utilisateur a déjà fait l'apprentissage culturel ou scolaire. Lorsqu'un document mobilise plusieurs formes sémiotiques d'appropriation, on dira que le document est multimédia. Par exemple, l'audiovisuel est multimédia car il mobilise l'image, la musique, le bruit et la parole. L'image peut elle-même être multimédia si elle comporte des textes et des structures iconiques.

La modalité d'appropriation

La forme sémiotique d'appropriation s'adresse à une ou des modalités perceptives. Quand il y a plusieurs modalités d'appropriation, le document est multimodal. Par exemple, l'audiovisuel est multimodal car il s'adresse à la vue et à l'ouïe.

Traditionnellement, nous sommes habitués au document textuel sur un support papier. Ce type de document est, bien qu'il ait incarné longtemps ce qu'il fallait comprendre par document, particulier, car dans ce cas :

- support d'enregistrement et support d'appropriation sont confondus : le support sur lequel on lit est celui que l'on range et stocke ;
- les formes d'enregistrement et d'appropriation sont confondues : ce que l'on lit est bien ce que l'on a inscrit sur le support.

Le document papier est devenu un cas particulier quand il a fallu envisager un autre type de document, les documents temporels. En effet, il faut faire une importante distinction parmi les formes sémiotiques d'appropriation des documents :

- *les formes statiques et spatiales d'appropriation* : les structures interprétables sont présentées simultanément à l'utilisateur. L'ordre et le rythme dans lequel se déroule la lecture ou la consultation sont laissés à la discrétion du lecteur/utilisateur. Même si la succession linéaire des caractères d'un document papier suggère un ordre canonique de lecture, elle n'est pas une condition nécessaire et incontournable de la lecture ;
- *les formes temporelles et dynamiques d'appropriation* : les structures interprétables sont présentées successivement à l'utilisateur, dans un ordre et selon un rythme imposé à l'utilisateur par le document lui-même. En particulier, l'ordre et le rythme constituent ce par quoi le document fait sens pour un utilisateur. Accéder au sens du document temporel, c'est se conformer à l'ordre et au rythme du document.

Ce dernier cas correspond aux documents sonores et audiovisuels, documents qui enregistrent un cours temporel pour le préserver et le restituer. Ces documents rencontrent cependant la difficulté suivante :

- un support d'enregistrement est nécessairement matériel, donc spatial. Chargé de préserver dans le temps le contenu, il ne peut posséder l'écoulement du temps. La forme d'enregistrement est par conséquent spatiale ;
- la forme sémiotique est temporelle, c'est-à-dire que la temporalité fait partie intrinsèque du mode de signification du document qui doit par conséquent véhiculer par lui-même, en lui-même, la temporalité qui permet de restituer le contenu.

On en déduit que, contrairement au document papier, il ne peut y avoir identité entre le support d'enregistrement et le support de restitution, la forme d'enregistrement et la forme sémiotique d'appropriation pour le document temporel.

Par conséquent, il ne peut y avoir de document temporel qu'à partir du moment où l'on dispose d'un procédé permettant à partir de l'enregistrement de reproduire une forme temporelle. Ce procédé doit être un procédé mécanique, c'est-à-dire un procédé permettant de reconstruire un déroulement temporel à partir d'un ordonnancement statique et matériel d'éléments. Ainsi, le signal magnétique, statique, permet de piloter un magnétoscope et une télévision pour reconstruire une forme physique d'appropriation à laquelle fusionne une forme sémiotique temporelle.

On doit par conséquent distinguer la forme d'enregistrement du contenu, car elle n'est pas lisible comme telle : elle est destinée non à être lue, mais à être jouée par un mécanisme qui va reconstruire la forme temporelle du document. La forme d'enregistrement n'est donc accessible que par la médiation d'un dispositif de lecture, un *player*, qui permet de décoder cette forme pour reconstruire le document.

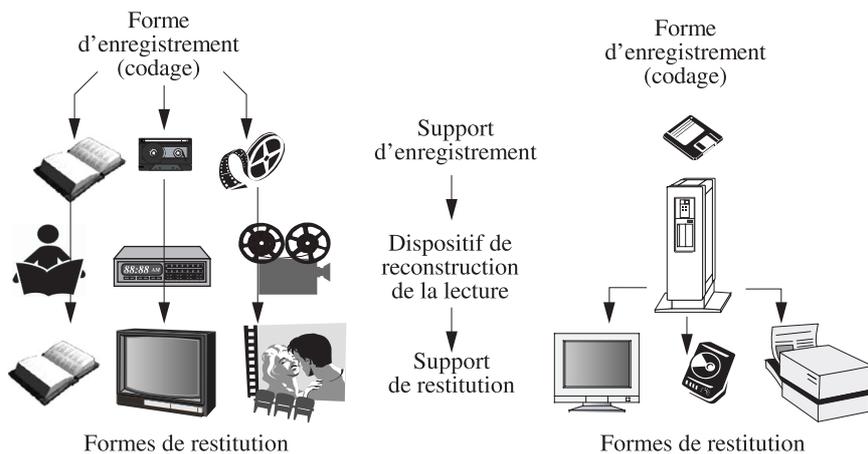


Figure 5.1. *Les dimensions documentaires*

La forme sémiotique d'appropriation est au contraire lisible, sa finalité est de renvoyer à un registre sémiotique maîtrisé et connu par l'utilisateur. On en déduit que pour les documents temporels, le document lu, c'est-à-dire la forme sémiotique d'appropriation adossée à la forme physique d'appropriation, n'est pas le document enregistré, la forme d'enregistrement sur le support d'enregistrement.

Une fois que l'on a introduit la notion d'un mécanisme permettant de jouer le document à partir de son enregistrement, le passage au numérique coïncide simplement avec un changement de codage et de technologie. Ainsi, au premier abord, le numérique n'introduit fondamentalement rien de nouveau, rien que l'on ne

connaissait déjà avec les documents temporels. Pourtant, comme la section consacrée au numérique tâchera de le montrer, ce changement anodin en apparence contient la possibilité de mutations bien plus profondes.

Pour conclure cette section, une question peut être de savoir où gît la véritable nature du document, dans quelle dimension faut-il voir la quintessence du document. Il nous semble que ce sont les formes d'appropriation qui sont le véritable dépositaire du document, et que les dimensions attachées à l'enregistrement caractérisent non le document, mais les ressources permettant de reconstruire le document. Si l'on reprend la définition vue plus haut des documents, l'inscription matérielle d'un contenu n'est appréhensible qu'au niveau de l'appropriation, quand un contenu est présenté de manière intelligible, sémiotiquement compréhensible, pour un utilisateur. On arrive au paradoxe apparent qu'un document temporel n'existe que pendant son déroulement temporel : le document audiovisuel, le film par exemple, n'est un film que pendant sa projection au cinéma ou sa diffusion à la télévision. Il ne s'agit pas d'un paradoxe, mais simplement de la conclusion qu'il faut tirer de la nature temporelle du document. Ainsi, le film enregistré sur un support argentique, le document audiovisuel consigné sur une cassette vidéo, ne sont pas des documents, mais des outils, des ressources que l'on peut utiliser pour consulter les documents dont ils sont l'enregistrement. Ainsi, une cassette n'est pas l'enregistrement d'une réalité faite d'images et de sons, mais l'enregistrement d'un document audiovisuel, c'est-à-dire de l'acte de publier dans une certaine forme, pour exprimer un certain sens, d'un déroulement construit dans le temps.

Le document lu ou consulté est toujours le résultat d'une reconstruction effectuée par un dispositif technique, aujourd'hui majoritairement numérique. Ce qui est consulté ne correspond pas ce qui est enregistré, mais à sa transformation par le calcul¹. Un document devient par conséquent indissociable du processus informatique qui le reconstruit et des informations qui paramètrent et pilotent la reconstruction. Ces informations décrivent ce qui est nécessaire à propos du document pour que la reconstruction soit conforme à l'objectif visé. Cette information est donc une

1. Le fait qu'un document soit toujours le résultat d'un processus calculatoire ne va pas sans poser de nombreuses questions, en particulier sur l'authenticité d'un document. Dans quelle mesure un archiviste, un historien ont-ils affaire au document brut quand il est numérisé ? Qu'est-ce qui peut leur garantir l'authenticité du document consulté ? En fait, aucune réponse ne peut venir de la technique puisque, comme nous l'avons montré, le document numérisé n'est pas un objet stable et unique, mais un processus dynamique entre enregistrement et restitution. Ce n'est que dans la confiance que l'utilisateur met dans la reconstruction qu'il peut fonder son exploitation du document. L'inauthenticité introduite par la technique doit se compenser par les réseaux de la probité et de la confiance : parce que je sais qui m'a donné tel enregistrement, parce que je connais le programme que j'utilise, je crois ce que je vois sur l'écran. Plus que jamais le fonctionnement technique renvoie à une herméneutique de la confiance (sur ces questions voir notre article [BAC 00], p. 3-15).

métadonnée, information sur le document qui le rend exploitable pour une reconstruction donnée. Les métadonnées sont donc ces informations qui rendent l'information utile et exploitable. Cette notion généralise celle d'index et de recherche d'information.

Enfin, les métadonnées doivent être créées avec le contenu documentaire, en amont, et non aval comme la plupart des informations d'indexation. Autrement dit, les métadonnées s'intègrent au contenu, en constituent une partie indissociable si bien que l'on peut dire à leur endroit que les métadonnées sont des données.

5.3. Indexation : caractérisations générales

5.3.1. Indexation traditionnelle

L'indexation a pour principal objectif de rendre accessible des informations, que l'on repère pour cela au moyen d'index. L'indexation est le processus selon lequel le contenu d'un document est analysé pour être ensuite reformulé dans une forme permettant d'accéder au contenu et de le manipuler. Le terme d'indexation qualifie à la fois le processus et son résultat. Une indexation est par conséquent la description d'un document effectuée dans la perspective d'une utilisation et exploitation donnée.

L'indexation repose traditionnellement sur deux étapes clairement distinguées :

- une étape d'analyse conceptuelle : le contenu est analysé et interprété par un documentaliste pour définir les principaux concepts permettant de le caractériser ;
- une étape de reformulation documentaire : l'analyse conceptuelle permet au documentaliste de reformuler le contenu dans une forme permettant sa manipulation.

Ce dernier point est essentiel : la nature des index composant une indexation est déterminée par le type de manipulation visée sur le document. Classiquement, la principale exploitation considérée est la recherche d'information : savoir où est l'information recherchée et extraire du fonds documentaire les documents correspondants. L'indexation a donc deux finalités : d'une part elle doit être directement exploitable pour déterminer où l'information recherchée, d'autre part elle doit permettre d'aller chercher cette information. Ainsi, dans une bibliothèque classique, chaque ouvrage est référencé par une catégorie déterminant son contenu, et une cote permettant de le situer sur une étagère donnée. Par ailleurs, des fiches rassemblent toutes les descriptions d'ouvrages : lorsqu'un lecteur réclame un ouvrage, le bibliothécaire consulte ses fiches pour déterminer quels ouvrages correspondent à la demande, et pour savoir sur quelles étagères les chercher. L'index a donc servi à déterminer l'information (les ouvrages) pertinentes et à la localiser (les étagères). Les fiches sont conçues pour être facilement consultable par le

bibliothécaire : l'indexation doit être elle-même manipulable pour permettre l'accès et la manipulation des contenus qu'elle référence.

Finesse et granularité d'indexation

La question habituelle à laquelle il faut répondre pour mener à bien une indexation est de savoir jusqu'à quel degré de finesse de description il faut conduire le processus. Cette finesse s'entend à deux niveaux : d'une part elle correspond à la richesse des concepts ou descripteurs mobilisés pour constituer une indexation ; d'autre part elle correspond à la granularité de la localisation des informations dans les documents décrits. Appelons la première acception la *finesse* de description, et la seconde la *granularité* de la description.

La finesse de description dépend de la fidélité avec laquelle il faut rendre compte du contenu des documents pour permettre de répondre adéquatement aux recherches d'information. Deux facteurs rentrent habituellement en ligne de compte : le répertoire conceptuel utilisé et la structure permettant d'articuler les index entre eux au sein d'une même indexation. C'est ainsi que la structure autorisée habituellement est l'articulation booléenne de concepts dans une requête. Le répertoire quant à lui renvoie au système de catégorisation utilisé.

La granularité de description dépend de la manipulabilité de l'information contenue dans les documents. Pour un ensemble de contenus, appelons *unité de manipulation* la plus petite entité directement accessible et manipulable. Par exemple, l'unité de manipulation d'un texte est le caractère, d'une image, le pixel, etc. L'intérêt d'une telle notion pour l'indexation est qu'il ne sert à rien de localiser une information plus précisément que l'unité de manipulation qui la contient. En effet, cette dernière étant la plus petite entité directement accessible et manipulable, il ne sert à rien de savoir où dans une unité se situe l'information recherchée puisqu'on ne pourra exploiter cette information de localisation dans le système d'accès au contenu. C'est pourquoi, dans la plupart des systèmes documentaires habituels, comme les bibliothèques classiques, l'unité de manipulation étant le livre, l'indexation effectuée, quelle que soit sa finesse conceptuelle et structurelle, ne possède jamais une granularité inférieure à celle du document, car la bibliothèque, comme système de recherche d'information, ne manipule physiquement que des livres, et non des pages par exemple. Ainsi il ne sert à rien, pour le bibliothécaire, de savoir que telle information se situe à telle page ou telle page d'un livre, car de toute manière, si c'est cette information qui est recherchée, c'est le livre en son entier qu'il faut sortir des rayonnages. Localiser l'information en termes de pages et non en termes de livres n'a un intérêt que si on peut manipuler les pages indépendamment des livres. Evidemment, le point de vue du lecteur est différent de celui du bibliothécaire : le système documentaire n'est plus, pour le lecteur la bibliothèque, mais le livre, au sein duquel il doit localiser l'information. L'unité de manipulation devient alors la page : la structure du *codex* permet en effet un accès direct (accès

aléatoire) à une page donnée arbitrairement. Le lecteur n'a alors plus affaire à un bibliothécaire assurant la médiation entre sa recherche et le fonds disponible, mais à des outils de navigation et de consultation comme la table des matières et les index. En associant ces outils au livre, ce dernier devient une base documentaire auto-décrite. Le lecteur peut alors assumer le double rôle de bibliothécaire et de lecteur.

5.3.2. Les conséquences de l'informatisation

5.3.2.1. Granularité et informatisation

L'indexation connaît une mutation profonde du fait de l'informatisation massive des contenus. Cette informatisation peut avoir une portée plus ou moins profonde, ou engendrer plus ou moins de problèmes fondamentaux selon la manière dont elle est menée. Il faut distinguer deux approches : la numérisation de la forme *physique* d'appropriation, et la numérisation de la forme *sémiotique* d'appropriation.

La numérisation de la forme physique correspond au fait que la forme sous laquelle le contenu apparaît est discrétisée. Les unités discrètes obtenues par la numérisation sont arbitraires² par rapport au sens véhiculé par le document : il n'y a pas de lien de direct entre d'une part l'interprétation du document et d'autre part la nature et la valeur des unités discrètes. Par lieu direct, il faut entendre deux choses : d'une part, l'interprétation du document ne permet pas de déduire la nature et la valeur de l'unité discrète, d'autre part l'altération de l'unité discrète n'altère pas l'interprétation du document. Ainsi, il n'existe pas de liens réguliers et stabilisés, même arbitraires, entre la forme discrétisée et son interprétation. Par exemple, le pixel est une unité discrète possédant un rapport arbitraire au sens de l'image dont il fait partie. En outre, modifier un pixel dans une image n'altère pas *significativement* l'image. C'est ainsi que les écrans à plasma tolèrent jusqu'à 5 points défectueux avant que la garantie du constructeur ne s'exerce : on accepte donc que 5 pixels défectueux par image affichée ne peuvent altérer significativement le contenu affiché.

La numérisation de la forme sémiotique correspond au fait que les unités discrètes obtenues entretiennent un lien avec l'interprétation du contenu. Ce lien peut être arbitraire, au sens où le signifiant linguistique est arbitraire, mais il est systématique : modifier une unité discrète modifie le sens. Ainsi, l'alphabet et les caractères sont des unités discrètes composant la forme écrite. Bien que possédant un lien arbitraire au sens, l'altération d'un caractère modifie le sens d'un document. Qui voudrait d'un écran qui altère cinq caractères par page affichée ?

2. La linguistique mobilise la notion d'« arbitraire du signe ». Cette notion signifie que la forme signifiante d'un signe n'est pas motivée par le signifié du signe. Ainsi n'y-a-t-il pas de lien entre le signifié du mot « chien » et la suite de lettres c.h.i.e.n. qui donne ce même mot.

La numérisation de la forme sémiotique est une discrétisation s'appliquant sur ce qui fait signe dans un contenu : altérer les unités discrètes modifie la signifiante du document. Une telle numérisation repose sur le fait qu'une sémiotisation a déjà été faite et a été fixée, inscrite sur le support. En effet, il faut qu'il y ait eu une définition des signes composant le document pour que la numérisation puisse produire des unités discrètes coïncidant avec ces signes. L'histoire de l'écriture nous apprend que telle est son rôle : elle a contribué à établir et fixer un répertoire de signes dont la combinatoire permet de produire la presque totalité des formes signifiantes linguistiques. De cette manière, on comprend que l'alphabétisation (le processus établissant une écriture alphabétique) est une sorte de numérisation, que la numérisation actuelle reconduit en apportant en plus la manipulation calculatoire et alphabétique. Réciproquement, on peut supposer que la numérisation des contenus puisse avoir un impact sur la détermination de formes signifiantes élémentaires, de signes à la base de la constitution des contenus. Autrement dit, la numérisation des contenus, en inscrivant dans le système technique les unités discrètes obtenues, contribue à voir en elles les signes alphabétiques à la base de l'écriture des contenus.

En anticipant sur les aspects multimédias que nous développerons ci-après, nous pouvons prendre comme exemple les normes de compression et de numérisation MPEG. MPEG-1 et 2 sont des normes opérant au niveau du pixel et exploitant la redondance de l'information associée à chaque pixel. Cette information est une information physique de luminance et chrominance sans lien direct avec l'interprétation. L'interprétation ne dégage jamais, dans sa sémiotisation, le pixel comme signe support du sens. En revanche, la norme MPEG-4 analyse le flux audiovisuel comme un ensemble d'objets reliés et articulés par des relations spatio-temporelles. Or, la caractérisation des objets se fait sur la base d'un point de vue selon lequel telle ou telle partie du contenu fait sens. Par exemple, pour des raisons liées à l'économie de la production, on considérera que le fond d'écran de télévision est un objet, distinct des autres, parce qu'il est construit de manière spécifique et, étant peu variable dans une émission, il peut être transmis seulement à intervalles espacés. Ces objets sont donc des signes pour un point de vue donné. Par conséquent, la segmentation en objets MPEG-4 du flux audiovisuel prescrit *a priori* une sémiotisation en signes porteurs d'un sens possible. Rien n'empêche de vouloir lors de la consultation de vouloir interpréter différemment le contenu et d'y voir une autre sémiotisation ; mais le spectateur doit négocier avec cette prescription avant toute chose, car la réception qu'il en a est conditionnée par elle (par exemple, les objets MPEG-4 sont interactifs si bien que le spectateur est amené lui-aussi à les considérer comme des objets à part entière.

L'audiovisuel a d'abord connu une numérisation de la forme physique et non de sa forme sémiotique. La raison est qu'il n'existe pas de systèmes de signes audiovisuels, stabilisés et normés culturellement, à partir desquels les contenus sont constitués. Il n'est donc pas possible d'établir des unités discrètes coïncidant avec

des unités signifiantes dans la mesure où ce qui fait signe peut constamment changer dans l'image selon le point de vue adopté, sans que des régularités se dégagent. Nous aurons l'occasion d'y revenir.

Qu'il s'agisse d'une numérisation de la forme physique ou de la forme sémiotique, elle apporte dans les deux cas la possibilité de définir autant niveau de granularité pour les unités de manipulation que nécessaire. Par conséquent, toute unité repérable comme étant signifiante, faisant sens pour un point de vue donné, peut être manipulée en tant que telle et extraite pour être exploitée indépendamment du contenu dont elle est issue. Cette possibilité accroît considérablement la complexité de l'indexation : alors que dans le cas classique, l'unité de manipulation est implicitement déterminée par le système (le livre ou le document), la numérisation implique qu'il est nécessaire de préciser explicitement la partie du document correspondant à l'information recherchée.

Les index ne sont plus seulement structurés en fonction des relations logiques et conceptuelles entre les descripteurs qu'ils mobilisent, mais en fonction des relations unissant les parties décrites d'un contenu. Ainsi, l'index ne dira pas seulement qu'un document par de l'apprentissage des langues (l'index étant structuré par le concept d'apprentissage, relié par une relation *a_pour_objet* au concept de langue), mais par exemple que la partie évoquant l'apprentissage du latin est incluse dans celle traitant des langues mortes.

Cette possibilité a eu pour conséquence, bien connue, de pouvoir repérer un contenu en le marquant par des balises insérées dans le document : un segment indexé est un segment compris en une balise ouvrante et une balise fermante, le libellé de la balise permettant de qualifier le contenu.

```
<Introduction>
Ceci est un segment de texte
</Introduction>
<chapitre>
Ceci est un autre segment
<paragraphe> encore un autre segment </paragraphe>
</chapitre>
```

Figure 5.2. *Balises et détermination d'unités de manipulation*

Cette possibilité permet d'envisager de manipuler directement le segment indexé à travers la manipulation des balises associées : les balises, posées par l'indexeur possèdent une syntaxe et une sémantique claires puisqu'elles sont créées par l'indexeur en vue de l'exploitation du contenu. Par conséquent, il est possible d'extraire toutes les introductions d'une collection pour constituer une anthologie simplement en repérant les segments textuels compris, dans chaque document, entre

la balise ouvrante <introduction> et la balise fermante </introduction>. Par le jeu des balises, toute unité de sémiotisation peut devenir une unité de manipulation.

La documentation en son ensemble subit donc une évolution fondamentale puisque la seule limite désormais à la pose d'index réside dans la capacité à repérer des unités de sens.

5.3.2.2. *Indexation par le contenu*

Si l'informatisation permet de redéfinir à volonté les unités de manipulation d'un document et d'articuler l'indexation du contenu sur la structure documentaire de ce dernier, l'informatisation a ouvert d'autres possibilités qui ont renouvelé également les travaux sur l'indexation.

En effet, traditionnellement, comme nous l'avons vu plus haut, le contenu documentaire s'indexe par des concepts exprimés linguistiquement dans un vocabulaire ou langage contrôlé. Le support numérique permettant de manipuler n'importe quelle représentation binaire, il n'est plus nécessaire de se restreindre à des index de nature linguistique. C'est ainsi que depuis de nombreuses années on étudie la possibilité de recourir à des index non linguistiques, par exemple des images ou des sons. La question est alors de savoir si l'on peut indexer un contenu documentaire par de l'image ou du son, en suivant l'intuition selon laquelle il serait plus naturel d'indexer de l'image par de l'image plutôt que par du texte.

Cette intuition se fonde sur une analogie avec la recherche d'information textuelle. En effet, le propre d'un texte est d'être composé de chaînes de caractères séparées par des blancs, correspondant peu ou prou à des mots. Cette propriété, anodine en apparence, est riche de conséquences. En effet, les mots possèdent la propriété d'avoir une signification hors contexte prescrite par le système fonctionnel de la langue³, et de posséder un sens en contexte qui correspond plus ou moins au sens hors contexte⁴. Par conséquent, en composant une requête de mots hors contexte, pour retrouver des segments textuels les utilisant (par une simple comparaison des chaînes de caractères séparées par des blancs avec les unités de la

3. Le système fonctionnel d'une langue correspond à l'ensemble des morphèmes munis de leur sèmes inhérents [RAS 94]. Un morphème correspond à la plus petite unité linguistique pourvue d'une signification. Les sèmes correspondent aux traits de signification caractérisant le signifié des morphèmes (le sémème) obtenus par une analyse différentielle avec les autres morphèmes voisins en langue. Par exemple, « bistouri » et « scalpel » se distinguent par le fait que le premier est pour les vivants et le second pour les morts. « Pour les morts » et « pour les vivants » sont par conséquent des sèmes, ici inhérents car caractérisant par défaut, avant leur usage en contexte, les morphèmes en question.

4. Il est d'usage en linguistique, depuis Beauzée notamment, de distinguer le sens de la signification. La signification correspond au sens idéalisé, reconstruit par le linguiste, abstrait à partir des différents usages du terme. Le sens correspond au sens en contexte.

requête), on obtient des parties du document qui ont potentiellement un lien avec la requête. C'est la base de la recherche en texte intégral. Ce schéma fondamental doit être évidemment amélioré pour pallier ses déficiences bien connues : puisqu'un mot en contexte ne signifie pas toujours la même chose que hors contexte (en fait dans le contexte d'une requête), on récupère des segments sans lien avec la requête, c'est le bruit ; puisqu'un mot peut avoir une signification exprimée par d'autres expressions linguistiques, on ne récupère pas les segments contenant ces expressions alors qu'il le faudrait, c'est le silence.

En évoquant une indexation par le contenu, on envisage ainsi deux choses :

- le fait que l'on dispose d'une requête sous la forme d'un son ou d'une image permet de retrouver des documents similaires au sens d'une distance entre images ou sons ; il s'agit alors d'une recherche d'information par similarité ;
- le fait que l'on puisse indexer un contenu documentaire par des index de nature imagée ou sonore.

Nous argumentons que si le premier cas renvoie à une possibilité effective mais peu utile, la seconde est impossible. Pour le dire en un slogan rapide, les descripteurs de l'indexation par le contenu ne sont pas des index au sens propre du terme. Il n'est donc pas correct de parler d'indexation par le contenu, mais simplement de recherche par similarité. Pour le comprendre, il faut revenir à la nature sémiotique de l'image et du son.

En effet, l'interprétation des images ne s'effectue pas selon le même régime que l'interprétation linguistique. Il convient par conséquent de préciser la nature sémiotique des images pour mieux appréhender les difficultés inhérentes de son indexation. Très grossièrement, il convient de distinguer au moins trois registres sémiotiques principaux : les images, les schémas et les concepts.

Les images sont des représentations analogiques du réel. A ce titre, ce sont des signes, en reprenant l'antique caractérisation du signe comme *aliquid stat pro aliquo*. Mais c'est un signe qui renvoie à ce qu'il signifie en le montrant à travers une forme perceptive analogue. C'est pourquoi on peut parler de « signe qui montre » pour les images. Les concepts sont des représentations arbitraires du réel : le lien qui unit un signe à ce qu'il signifie est arbitraire et conventionnel. On peut parler de « signe qui dit » pour caractériser ce type de signe. Enfin les schémas sont des signes qui illustrent par une forme perceptive le contenu d'un concept : un dessin vaut mieux qu'un long discours, comme l'évoque un célèbre adage. Dans ce cas, le schéma est le « signe qui illustre ».

Le schéma montre dans un rapport analogique une réalité dont la complexité est réduite de manière à montrer perceptivement le contenu du concept. Le schéma

montre dans l'espace et le temps de la vision la signification du concept. Il n'en est pas l'explicitation ou l'explication, car le schéma interprète le contenu du concept pour l'illustrer⁵.

La principale difficulté présentée par l'image ou par tout signe qui montre est que l'image montre sans préciser ce qu'elle montre. L'image ne prescrit pas par elle-même une signification. Cela est abondamment illustré par la production télévisuelle où de mêmes images sont réutilisées dans des contextes différents et pour soutenir ou illustrer des significations différentes. C'est ainsi que l'image utilisée pour montrer la pollution pétrolière pendant la guerre du Golfe n'était pas une image filmée lors de ce conflit, mais l'image d'un cormoran mazouté de la marée noire provoquée par le naufrage de l'Amoco Cadiz dans les années 1970.

Il en est tout autrement pour les concepts : ceux-ci s'expriment par des mots empruntés à la langue et ils en héritent les propriétés de systématisme. La langue comme système est une conception linguistique théorisée par Saussure dans son *Cours de linguistique générale* [SAU 85]. Saussure pose que chaque unité linguistique signifie en fonction des identités et des différences qu'elle entretient avec les autres unités linguistiques, en particulier celles qui lui sont les plus proches dans la langue. La signification linguistique est donc *différentielle* : c'est la différence avec les voisins en langue qui constitue la signification d'un terme.

« Dans l'intérieur d'une même langue, tous les mots qui expriment des idées voisines se limitent réciproquement : des synonymes comme redouter, craindre, avoir peur n'ont de valeur propre que par leur opposition ; si redouter n'existait pas, tout son contenu irait à ses concurrents. »

La langue se présente donc comme un *système* où chaque unité voit sa signification déterminée en fonction de sa *position* dans le système. Toute modification d'une unité modifie en conséquence les autres et reconfigure leur signification :

« Quand on dit qu'elles [les valeurs] correspondent à des concepts, on sous-entend que ceux-ci sont purement différentiels, définis non pas positivement par leur contenu, mais négativement par leurs rapports avec les autres termes du système. Leur plus exacte caractéristique est d'être ce que les autres ne sont pas. »

5. On pourrait imaginer d'autres schématisations interprétant et illustrant dans d'autres registres perceptifs le concept. Cette notion de schématisme est complexe. Comme on le sait, elle remonte à Kant qui voyait dans les schèmes une figuration des concepts dans le cadre spatio-temporel d'une perception possible. Littéralement, le but du schème est de montrer à quoi pourrait ressembler, concrètement, le concept dont il est le schème. Un exemple kantien est que le concept de causalité se traduit dans le temps de la perception comme une succession ordonnée (si A cause B, alors dans les faits concrets, A précède B). Un exemple non kantien est le concept de démontage d'une roue et le schéma explicatif du mode d'emploi.

Le concept prescrit une signification car il appartient à un système d'oppositions et de différences qui lui donne hors contexte un contenu. La situation est donc très différente des images et des sons. Il n'existe pas de système fonctionnel pour les images et les sons, c'est-à-dire de système permettant de prescrire *a priori* une signification à une image ou l'un de ses segments. C'est pourquoi il faut donc soigneusement distinguer les *descripteurs* d'une image ou d'un son des index permettant de la qualifier :

- un descripteur est une information extraite d'un document ou d'un contenu par des moyens automatiques. L'information associée à un descripteur est donc proche du contenu physique du document ;

- un index est une position ou une valeur donnée dans un système d'interprétation associé à l'exploitation d'un ensemble de documents, c'est-à-dire un fonds documentaire ;

- un descripteur n'est pas un index dans la mesure où deux descripteurs possédant une information distincte ne renvoient pas nécessairement à deux interprétations différentes, c'est-à-dire à deux index différents. Par exemple, si l'information extraite est un histogramme de couleur, une différence entre deux histogrammes n'implique pas qu'il faille considérer différemment les images dont ils sont extraits. Le problème est que les descripteurs sont des informations extraites, mais pas catégorisées. Ils ne réduisent pas la complexité du réel comme permet de le faire un système conceptuel. En revanche, deux index ou concepts différents correspondent à deux usages ou exploitations distinctes du contenu. Les concepts catégorisent les contenus documentaires. Le schéma suivant résume l'argument en illustrant que les descripteurs sont de l'information précatégorisée et les index de l'information catégorisée. Autrement dit, si l'indexation a pour fonction d'appliquer un système de catégories sur un fonds documentaire pour en permettre l'exploitation, la description par le contenu n'est pas une indexation.



Figure 5.3. Iconicité de l'image : cette image montre quelque chose que l'on voit et reconnaît, mais que représente-t-elle ? Une tempête, un attentat, une émeute ?

La pratique documentaire confirme cette analyse. Dans le monde audiovisuel, les contenus sont accompagnés tout au long de leur cycle de vie de document textuel leur prescrivant une signification. Puisque les documents audiovisuels ne peuvent signifier par eux-mêmes, ils sont plongés dès leur conception dans un univers textuel qui construit le sens dont ils seront la manifestation audiovisuelle. C'est ainsi que lettres d'intention, projets, scripts, conducteurs, notes de montage, etc., entourent l'objet audiovisuel et en catégorisent le contenu. Lors de sa diffusion, les guides de programme et les journaux de télévision (par exemple *Télérama*) construisent un sens *a priori* pour le téléspectateur. Finalement, l'archivage documente les contenus pour leur associer une signification permettant de les retrouver et de les utiliser.

5.3.3. *Indexation : trois types et trois aspects*

Il est temps à présent de résumer nos propos sur l'indexation. Depuis sa tradition documentaire à sa mutation entraînée par le numérique, on peut distinguer trois grands types d'indexation :

- *l'indexation conceptuelle* : le contenu est qualifié par un concept qui décrit ce dont il est question dans le document. On dit en général que l'indexation conceptuelle est thématique. Les concepts appartiennent à des systèmes dont l'organisation est plus ou moins explicites, allant de simples index (au sens de liste de termes), à des ontologies en passant par des thésaurus ou des terminologies. L'organisation est toujours de nature linguistique ou logique : elle ne traduit pas comment sont articulés les segments décrits par les concepts, mais les relations logiques entre les concepts les indexant. De nombreux travaux portent à présent sur les possibilités d'utiliser les ressources de l'intelligence artificielle et de la représentation connaissance pour profiter dans le cadre de l'indexation des outils d'inférence ;

- *l'indexation structurelle* : les index ont pour fonction de décrire comment sont reliés entre eux les segments indexés. L'indexation structurelle porte donc sur la mise en forme du document et sa structure. Elle insiste particulièrement sur la manière d'articuler la qualification et la localisation ;

- *l'indexation par le contenu* : comme on l'a vu, cette dénomination est impropre, il conviendrait de parler de description par le contenu. Il s'agit d'extraire une information, une signature, permettant d'associer un contenu documentaire similaire au sens d'une métrique donnée. La métrique et la similarité peuvent prétendre à constituer une indexation si et seulement si elles sont capables d'offrir une catégorisation (par exemple, le contenu du concept « triste » correspond à un cluster constitué par similarité de contenus). Mais on sait que le problème de la catégorisation dépasse largement celui de l'indexation pour renvoyer à des questions passionnantes mais non résolues de sciences cognitives.

Par ailleurs, l'indexation se décompose en trois étapes importante qui retracent les fonctions que l'on assigne à une indexation :

- une étape de *qualification* : il s'agit de caractériser l'information contenue dans le document de point de vue de l'interprétation qu'elle possède pour une recherche d'information. Grossièrement, il faut savoir de quoi ça parle pour établir le lien avec ce que l'on cherche ;
- une étape de *localisation* : outre le fait de savoir de quoi ça parle, il faut savoir où on en parle. La localisation situe l'information recherchée en termes d'unités de manipulation ;
- une étape de *structuration* : les index, qualifiés et localisés, sont agencés et articulés entre eux. Cette articulation permet de croiser qualification et localisation au sein de structures d'index.

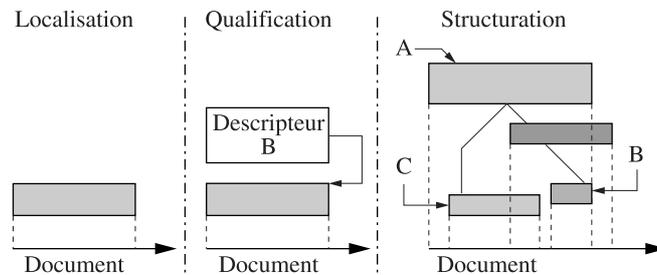


Figure 5.4. Les trois étapes de l'indexation

Ces trois étapes sont intimement liées et se font plus ou moins simultanément, car elles sont mutuellement dépendantes les unes des autres : pour localiser, il faut savoir que quoi l'on veut parler, et également pour savoir de quoi on veut parler, il faut déterminer où on en parle. L'indexation est par conséquent la production d'une description documentaire paraphrasant le contenu en index localisés et structurés.

L'indexation est par conséquent une interprétation du contenu. De manière fondamentale, l'indexation établit en quoi un document ou un objet est significatif, fait signe, pour une recherche d'information ou une autre exploitation. Etablissant la signifiante du document, l'indexation le sémiotise : littéralement, elle en fait un signe qui signifie pour son lecteur. En comprenant l'indexation comme une sémiotisation, c'est-à-dire un processus interprétatif tirant ses déterminations de la sémantique et de la sémantique, on peut en déduire des conséquences confirmées par la pratique professionnelle.

L'indexation n'est pas universelle, au sens où elle serait unique et valable une fois pour toute. Comme toute interprétation, elle est située dans un contexte qui en

fixe les limites de pertinence. Il n'y pas une objectivité documentaire permettant d'établir une indexation universelle, mais des points de vue interprétatifs plus ou moins stables et établis.

L'indexation n'est pas définitive : conséquence du point précédent, une indexation n'est jamais terminée. De la même manière qu'il est toujours possible de reprendre une interprétation, de la modifier, de la compléter, de la reformuler, une indexation est toujours bornée par un contexte et un acte interprétatif finis et limités.

L'indexation hérite de l'interprétation son caractère holiste, où le contexte global détermine le local. Comme le souligne F. Rastier [RAS 94], déterminer la signification d'un document, c'est-à-dire ce qui fait signe et comment, n'est pas un donné du document, mais le résultat d'une interprétation. L'établissement des signes est le résultat et non la condition du parcours interprétatif. Par conséquent, l'indexation qui est une sémiotisation, n'est pas un processus qui décrit ce qui est dans un document, ou ce qui est donné avec lui, mais une interprétation qui ajoute et enrichit le document. Elle est donc relative et amendable. Au lieu de décrire une réalité intrinsèque au document qui lui est immanente, elle apporte une interprétation extrinsèque, qui lui est transcendante. Extrinsèque ne veut pas dire arbitraire, gratuit ou aléatoire : l'indexation est une interprétation qui est *motivée* par le document indexé, fondée en lui, mais sans avoir de rapport de nécessité logique ou déductive avec lui.

5.4. Indexation multimédia

Le numérique permet d'envisager des documents multimédias, c'est-à-dire mobilisant des formes sémiotiques d'appropriation différentes. Comme dans le cas du texte, le numérique permet de considérer autant d'unité de sémiotisation que nécessaire. Le multimédia introduit deux innovations essentielles qui modifient considérablement la problématique de l'indexation : la prise en compte des objets temporels et des hyperdocuments.

5.4.1. Les objets temporels

Les objets temporels sont des contenus dont la forme sémiotique de restitution est temporelle ou dynamique. Ce sont par conséquent des objets qui prescrivent l'ordre et le rythme de la lecture, par apposition aux documents aux formes sémiotiques de restitution spatiales qui n'imposent ni ordre ni rythme de lecture. Les objets temporels possèdent une durée qui est constitutive : il est impossible de les considérer globalement dans un seul instant, ce sont par essence des contenus pour

lesquels il faut du temps pour les consulter⁶. Les objets temporels posent classiquement de nombreux problèmes à l'indexation qui sont principalement liés à leur temporalité et leur caractère non alphabétique. Dans un contexte technologique numérique, ces problèmes peuvent être levés mais de nombreuses difficultés doivent alors être résolues.

Dans un contexte analogique, les systèmes de lecture sont des dispositifs physiques ayant pour fonction de restituer la forme temporelle du document à partir de son enregistrement spatial codé. Ces dispositifs sont dédiés, c'est-à-dire qu'ils sont spécialement conçus pour reconstruire la linéarité temporelle de la lecture. Cela a pour conséquence d'interdire quasiment l'accès direct ou aléatoire à une partie quelconque du document indépendamment du reste. Quand quelques outils le permettent, ils restent d'un maniement malaisé et garantissent rarement l'exactitude requise, comme par exemple le magnétoscope.

Par ailleurs, la forme temporelle des documents ne se prête pas à une lecture rapide, un feuilletage du document comme nous sommes accoutumés le faire pour les documents papiers. L'accès à l'information doit donc passer soit par une lecture entière du document, jusqu'à ce que l'on rencontre l'information recherchée, soit par son repérage explicite, permettant de la localiser sur l'enregistrement pour y accéder directement. Mais le repérage explicite retombe sur le problème du caractère non alphabétique des images. En effet, puisque les images montrent sans signifier, il faut expliciter dans un registre de type linguistique l'information recherchée et lui associer une localisation dans le document. Mais dans ce dernier cas, on retombe sur le problème précédent d'un accès direct techniquement malaisé.

Le numérique permet de réviser la perspective de l'indexation en offrant de multiples possibilités. En effet le numérique intervient, concernant les objets temporels, à deux niveaux :

- la numérisation, qui consiste à discrétiser les contenus en un format numérique, et à leur appliquer des traitements (compression, protection, etc.) et analyses (détection, extraction de descripteurs, etc.) ;
- l'informatisation qui consiste à intégrer les contenus temporels dans un système d'information pour l'échange et l'exploitation. Plus récente que la numérisation, l'informatisation est responsable de la révolution numérique dont on parle tant depuis quelques années.

6. Les objets temporels ont été particulièrement étudiés dans la tradition phénoménologique. Husserl a ainsi consacré de célèbres analyses au son et aux mélodies dans ses *Leçons sur la conscience intime du temps* ([HUS 64], p. 128). Ces études ont connu des prolongements dans la postérité phénoménologique ([GRA 68], p. 69] ou analytique ([MIL 84], p. 108) mais ont également été exploitées pour étudier les objets audiovisuels ([STI 96, p. 97]).

Alors que le problème-clé de la numérisation est la *transformation* du contenu, en appliquant des techniques issus du traitement du signal, reconnaissance des formes, plus généralement de mathématiques appliquées, celui de l'informatisation est l'échange et la manipulation du contenu, en appliquant des techniques issus du génie documentaire, génie logiciel, plus généralement de l'informatique symbolique. Pour la numérisation, le contenu est un objet physique (la lumière encodée dans les pixels, une onde sonore) que l'on traite par des mathématiques du continu opérationnalisées par le calcul numérique, pour l'informatisation le contenu est une information, un objet informatique que l'on traite par des processus symboliques. C'est essentiellement l'informatisation qui modifie la perspective de l'indexation pour les objets audiovisuels. Ces modifications sont les suivantes.

Intégration de la documentation et de la chaîne documentaire

La documentation, dans un contexte analogique, est nécessairement sur un autre support que l'objet temporel, dont le support est dédié au codage de l'enregistrement du contenu. Par ailleurs, l'objet temporel étant un objet techniquement et éditorialement complexe se produit en plusieurs étapes distinctes possédant chacune leur propre documentation : production (scripts, storyboard, etc.), diffusion (magazines de télévision, conducteurs, etc.), conservation (notice documentaire). La documentation est donc éclatée sur différents supports selon les différentes étapes de son cycle de vie.

Le numérique permet d'une part d'intégrer le document et la documentation sur un même support, et d'autre part de rassembler et d'échanger la documentation tout au long du cycle de vie. Le contenu se construit, s'échange, se diffuse, s'archive en emmenant avec lui sa documentation qui s'enrichit et s'adapte en fonction de son cycle de vie.

Alignement de la documentation

Dans un contexte analogique, la documentation, séparée du document, est difficilement couplée à ce dernier, comme on l'a dit. Cela implique en particulier que la structure de la documentation n'est pas articulée à la structure du document. Dans le contexte numérique où document et documentation sont réunies, les différentes parties de la documentation peuvent être alignées, référencées sur des segments du document.

Intégration des documents et de la documentation

La documentation, intégrée et alignée au document, rassemble les informations permettant d'accéder et d'exploiter le contenu. La documentation est par conséquent une condition *sine qua non* de l'exploitabilité du contenu et ne peut en être séparée :

la documentation, indissociable du contenu, devient une partie du document⁷. C'est le célèbre slogan *metadata is data*. On connaît bien ce principe à présent que les contenus numériques se répandent : l'information documentaire, présente sur la jaquette des CD musicaux est désormais intégrée dans les DVD sans laquelle ils sont difficilement consultables. Le format du DVD permet d'intégrer document et documentation.

En conclusion, la principale innovation apportée par le numérique est un fait simple mais fondamental : la coexistence sur un même support du contenu et des informations documentaires. Mais cette coexistence n'implique pas une inclusion : si les métadonnées sont intégrées et alignées au contenu, elle ne sont pas incluses dans le contenu.

En effet, dans un document temporel numérique, les balises permettant de localiser un segment indexé ne sont plus dans le document lui-même, ce qui permettait de déterminer implicitement que le segment textuel indexé par deux balises correspond au texte qu'elles encadrent. Du fait qu'il s'agit d'un flux binaire temporel, les informations documentaires engendrées par l'indexation ne sont pas dans le document, mais dans une notice ou une description matérialisée en une structure numérique distincte du document indexé. Il se pose alors le problème de faire l'articulation entre le document indexé et son indexation de manière explicite, et de permettre une localisation spatio-temporelle des segments indexés.

Cela entraîne quelques difficultés. L'ordre de la description ne coïncide pas nécessairement avec l'ordre du décrit : l'ordre dans lequel se succède les descriptions documentaires ne correspond pas nécessairement avec l'ordre dans lequel se succède les segments documentaires décrits. Par exemple, dans la figure suivante, la description documentaire commence par la description d'un morceau d'anthologie qui succède dans le document décrit à des segments qui ne sont décrits qu'après lui. Par ailleurs, il faut adopter des méthodes pour référencer le contenu tant dans l'espace que dans le temps.

Pour toute indexation d'un document temporel, il faudra par conséquent adopter un format permettant d'une part d'exprimer une localisation et d'autre part de l'articuler effectivement au contenu. Ce sera notamment repris par des normes comme MPEG-7 que nous présenterons brièvement à titre d'exemple.

7. On connaît ce phénomène dans le monde du texte. En effet, la table des matières est avant une métadonnée permettant de naviguer dans un contenu et de le consulter. Cette information est tellement indispensable que tout ouvrage est muni d'une table des matières, qui fait désormais partie du contenu à part entière. D'ailleurs, une table des matières peut se lire pour retracer par exemple l'argumentation de l'auteur.

5.4.2. Les hyperdocuments

Par hyperdocument, nous entendons la notion désormais banale d'hypertexte généralisée à des documents quelconques. Bien que les hyperdocuments existent depuis longtemps dans la tradition écrite, le numérique altère sensiblement leur nature et par conséquent l'indexation qui peut en être faite.

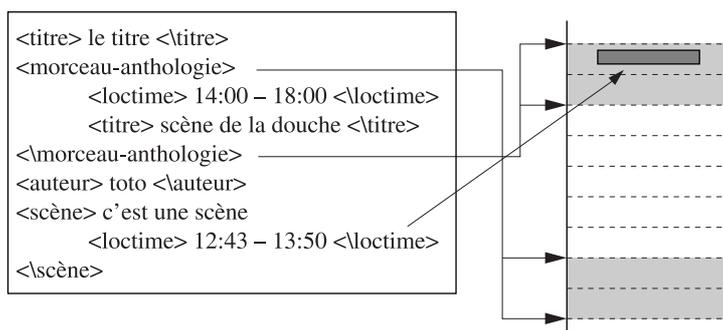


Figure 5.5. Indexer un objet temporel : l'information documentaire n'est pas dans le document, et son organisation n'est pas celle du document décrit ; l'ordre de la description n'est pas celui du décrit

On a vu plus haut qu'un document était un contenu publié, délimité spatialement et temporellement. Or les hyperdocuments numériques que l'on peut trouver sur le web en particulier, ne vérifient aucune de ces propriétés :

- un contenu sur le web ne possède plus de délimitation spatiale, les hyperliens interdisant de fixer une frontière. Quand on le fait, c'est que l'on réplique sur le web des critères documentaires hérités du monde du papier et de l'imprimerie ;

- un contenu sur le web ne possède plus de délimitation temporelle : puisque la publication est implicite et assumée par la mise en ligne, il est impossible de savoir si le contenu a changé depuis une dernière consultation ou non. Le contenu n'est pas figé en une version de référence, ce qui est du ressort de la publication éditoriale. Par conséquent, le web fait bien une mise à disposition du public un contenu, ce en quoi il s'agit bien d'une publication, il ne publie pas les contenus au sens du monde de l'édition. C'est lors de la lecture que l'on peut fixer une version de référence (un contenu et une date) de manière à distinguer lors des prochaines consultations si l'on a affaire à une nouvelle publication ou non. On a donc un déplacement de la publication éditoriale de l'auteur vers le lecteur : ce n'est plus un auteur qui s'entend avec un éditeur pour avoir une version de référence, mais un lecteur qui organise ses consultations et retient des versions de référence, publiées, à partir des mises à jour du contenu qu'il peut constater.

Cela amène à réviser la conception classique de l'indexation documentaire. En effet, quel est l'objet de l'indexation ? Comment identifier le contenu, le délimiter ? Ces problèmes se posent de manière particulièrement aiguë pour archiver les contenus du web ou d'un intranet, et les indexer. Ils devront pourtant être résolus pour aborder le dépôt légal du web comme semble le préconiser les projets de lois sur la société d'information, et par ailleurs pour que le web soit autre chose qu'un simple instrument de diffusion de contenus publiés ailleurs, mais un lieu de publication original, où l'on trouve ce qui n'existe pas ailleurs, et par conséquent exigeant une mémoire.

5.4.3. Un cahier des charges pour l'indexation multimédia

Face à ces transformations induites par le numérique et aux difficultés soulevées, une indexation multimédia et hypermédia doit par conséquent proposer des solutions aux différents problèmes suivants :

- gérer des contenus possédant différentes formes sémiotiques, en particulier les formes sémiotiques temporelles ;
- permettre de représenter une information qualifiant le contenu ; cette information peut être de différents formats et niveaux d'abstraction : descripteurs linguistiques constituant des index, descripteurs physiques, descripteurs iconiques, etc. ;
- articuler l'information qualifiant le contenu au contenu lui-même : il doit être possible de repérer n'importe quelle partie d'un hypermédia (segment d'un texte, segment spatiotemporel d'une vidéo, d'un son, fragment hypermédia d'un hypermédia) ;
- articuler la structure de l'information décrite à celle de l'information décrivant : comment articuler la structure de la description à la structure spatiotemporelle d'une vidéo ?

Ces points ne sont actuellement traités par aucun outil ou formalisme. Cependant, de nombreux travaux sont en cours qui permettent d'envisager des solutions prometteuses. Nous concluons ce chapitre avec un exemple représentatif : MPEG-7.

5.5. MPEG-7

MPEG est un comité d'expert (*motion picture expert group*) qui s'est constitué pour proposer des normes technologiques pour l'audiovisuel et aujourd'hui l'hypermédia. Les travaux de ce comité ont essentiellement porté jusqu'à présent sur la compression des objets sonores et vidéos. Les normes MPEG-1, MPEG-2 et MPEG-4 permettent de transmettre sur les réseaux des contenus audiovisuels dont le

volume non compressé rendait cette opération impossible. MPEG-7 est la dernière née des normes MPEG. Contrairement aux précédentes, MPEG-7 ne porte pas sur la compression mais la description des contenus. Avec MPEG-7, MPEG change de paradigme et élargit son horizon : il ne s'agit plus de numériser un signal et de le transformer, mais d'échanger les contenus en communiquant pour cela les descriptions documentaires permettant leur identification et exploitation. En reprenant les notions présentées plus haut, MPEG passe de la numérisation à l'informatisation.

MPEG-7, de son nom complet *multimedia content description interface*, est une proposition de normalisation portant sur la description des contenus. Mais, au lieu de normaliser ce qu'il faut dire à propos d'un contenu, MPEG-7 porte sur la manière de le dire. MPEG-7 a donc pour ambition de permettre l'échange des métadonnées, mais non de prescrire leur nature. A ce titre, l'objectif de MPEG-7 est fort proche de XML, et on peut résumer lapidairement MPEG-7 comme le XML pour l'audiovisuel. A ce titre, MPEG-7, comme XML, ne traite pas des points suivants :

- MPEG-7 ne prescrit pas *comment obtenir* les métadonnées ;
- MPEG-7 ne prescrit pas *comment utiliser* les métadonnées ;
- MPEG-7 ne prescrit pas *comment interpréter* les métadonnées.

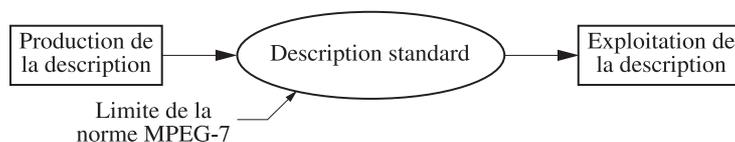


Figure 5.6. Limite de la norme MPEG-7

Autrement dit, MPEG-7 n'est pas une norme pour les outils d'extraction (obtenir les métadonnées), pour les moteurs de recherche (exploiter les métadonnées), pour les ontologies (interpréter les métadonnées).

5.5.1. Structures de MPEG-7

MPEG-7 propose plusieurs structures d'information permettant d'exprimer les descriptions. Ces structures sont les suivantes :

- les notions ou *features*, qui correspondent à tout ce que l'on peut avoir besoin d'exprimer pour décrire un contenu. On pourra avoir ainsi besoin de la notion d'auteur, de plans, de scène, de segment, etc., pour exprimer une description ;
- les descripteurs ou *descriptors* qui correspondent la représentation formelle et informatique d'une feature ;

- les schémas de description ou *schema description* qui correspondent à des structures articulant entre eux différents descripteurs et d'autres schémas de description. D'une certaine manière, le schéma de description adapte au contexte de MPEG-7 la notion de DTD de XML ou de schéma de XML/schema ;

- un métalangage, le DDL (*description definition language*), permettant de créer des descripteurs et schémas de description. Le DDL est fondé sur XML-Schema qu'il étend quelque peu.

On appelle *description* la méta-information créée sur un contenu et exprimée en suivant la norme MPEG-7. MPEG-7 propose *a priori* des descripteurs et des schémas de base permettant d'exprimer de manière standard l'information documentaire. D'une certaine manière, MPEG-7 va au-delà de la simple spécification d'un métalangage en intégrant dans la norme des structures de base formatant la description. Pour reprendre un exemple issu du monde du texte, MPEG-7 intègre XML et la TEI qui est, comme on le sait, une DTD proposée à la communauté des chercheurs en sciences humaines pour exprimer et échanger leurs corpus textuels. Le cœur de MPEG-7 est donc constitué des différents schémas de base proposés. Ils se répartissent en fonction de la nature du média : on compte donc les schémas pour le son, la vidéo et le multimédia.

Il reste bien sûr possible de déclarer de nouveaux schémas en fonction de ses propres besoins mais, ces nouveaux schémas n'appartenant pas à la norme bien qu'exprimés par son moyen, rien ne garantit que les informations formatées par ces schémas puissent être échangées et exploitées par d'autres utilisateurs : il faut pour cela d'une part transmettre la structure du schéma lui-même (comme il faut communiquer la DTD pour analyser un document SGML ou XML), et d'autre part transmettre la sémantique de cette structure pour permettre son interprétation. En effet, l'intérêt d'utiliser les schémas exprimés dans la norme est que leur sémantique est définie dans les documents normatifs auxquels chacun a accès.

Il est donc possible de réutiliser l'information transmise dans le format MPEG-7. A présent nous donnons quelques exemples de schémas de descriptions, renvoyant le lecteur aux documents normatifs pour plus de détails.

5.5.2. Les schémas pour le son

Les schémas pour l'audio sont conçus pour permettre d'exprimer des informations sur le signal sonore, telles que les outils actuels permettent de l'extraire. Ces schémas portent sur :

- les effets sonores ;

- la description du timbre des instruments. Le timbre correspondant à ce qui peut différer quand deux sons possèdent la même hauteur et la même intensité. Le schéma décrit les propriétés perceptuelles avec un nombre réduit de descripteurs comme « richesse », « attaque », etc. ;
- la parole : elle est décrite par une combinaison de sons et de mots, permettant ainsi de retrouver les mots inconnus d'un vocabulaire par les sons associés ;
- les mélodies : le schéma est conçu pour permettre des requêtes par similarité, notamment entre un air sifflé ou fredonné par un utilisateur et une mélodie stockée ;
- des descriptions de bas niveau sur le son : enveloppe temporelle, spectre, harmonies, etc. ;
- un descripteur « silence » permet finalement de décrire un contenu comme silencieux !

5.5.3. Les schémas pour la vidéo

Les schémas consacrés à la description des contenus vidéo et visuels sont complexes et élaborés. Ils portent principalement sur les aspects suivants :

- des descripteurs basiques permettant de localiser selon une précision variable une partie dans un contenu visuel ou vidéo. Cela renvoie par exemple un système de coordonnées 2D qui peut être propre à chaque frame ou commun à plusieurs, permettant dans ce dernier cas de comparer les positions entre différentes frames ;
- la couleur. On retrouve le savoir faire habituel de l'extraction automatique. On dispose ainsi de descripteurs et schémas pour déclarer l'espace de couleur, la quantization, la couleur dominante, etc. ;
- les textures : représentation d'une image comme partition de zones de texture homogène, des histogrammes, etc. ;
- les formes : analyse en région, représentation des contours, des formes tridimensionnelles ;
- le mouvement : mouvement de caméra, trajectoire, analyse du mouvement dans l'image, etc. ;
- la localisation : en particulier une localisation spatiotemporelle permettant de définir dans une séquence vidéo un « tube » ou région spatio-temporelle, qui peut être connexe ou non d'ailleurs ;
- enfin, la reconnaissance des visages.

Comme on le voit, les schémas pour la vidéo sont essentiellement conçus en fonction de l'état de l'art de l'extraction. Autrement dit, si MPEG-7 ne normalise pas les techniques d'extraction, il normalise la manière de représenter les résultats de ces techniques. On peut regretter que la normalisation ait été plus pilotée par ce que

l'on sait faire que par ce qui est utile. L'hypothèse implicite est que l'information extraite par les algorithmes d'extraction est utile et qu'il est inutile de rechercher ce qui est utile indépendamment de savoir comment l'extraire, manuellement ou automatiquement, car les algorithmes pourront à terme tout extraire. Cette hypothèse ne nous semble pas fondée car elle suppose pour être vérifiée que l'on sache combler le fossé sémantique entre l'information physique d'un contenu vidéo et l'interprétation conceptuelle qui en est faite, ce qui n'est pas envisageable pour le moment.

5.5.4. Les schémas pour le multimédia

Cette dernière ensemble de schémas est le plus riche et le plus complexe. De plus, il prend explicitement en compte des informations d'ordre documentaire, éditorial, et conceptuel, nécessaires pour la description hypermédia. MPEG-7 élargit son intérêt pour l'information documentaire en général au-delà des résultats d'algorithmes d'analyse. Les principales structures proposées concernent les points suivants :

- le management du contenu : les éléments MPEG-7 décrivent ce qui concerne la création du contenu, sa production, son encodage, les formats de codage et de fichier ;
- la description du contenu : cela concerne les aspects structurels (structure du contenu d'un point de vue spatial, temporel) et les aspects conceptuels (description du contenu du point de vue d'une sémantique « fondée sur le monde réel » et mobilisant des notions conceptuelles) ;
- la navigation et l'accès au contenu : des structures de résumé sont proposées, ainsi que des partitions ou décompositions et également des « variations » : description de différentes ressources représentant des variantes d'un même contenu, que l'on peut choisir en fonction du contexte ; par exemple des variantes linguistiques d'un même programme audiovisuel ;
- l'organisation du contenu : ces descripteurs et schémas permettent de décrire l'organisation en collections des objets, événements ou segments du contenu ;
- l'interaction avec l'utilisateur : cela concerne les préférences et le profils des utilisateurs pour la consultation des contenus. Ces descripteurs peuvent être utilisés avec ceux concernant les « variations » pour proposer les contenus les plus adaptées à un utilisateur (ceux de sa langue par exemple).

Ces schémas et descripteurs sont donc fort riches et complets. Il semble, au premier abord que la plupart des informations sur un contenu puissent être décrites par leur intermédiaire. Mais seul l'usage futur de la norme pourra le confirmer.

Pour conclure cette section sur MPEG-7, nous donnons un exemple de spécification d'un schéma de description portant sur les transitions entre des segments vidéos montés, c'est-à-dire établies lors du montage. Il faut d'abord établir la syntaxe. Les déclarations suivantes, à l'aide du DDL fondé sur XML/Schéma permettent de le faire.

```

<!-- #####-->
<!-- Definition of Transition DS -->
<!-- #####-->
<complexType name = "TransitionType">
  <complexType>
    <extension base = "mpeg7 :VideoSegmentType">
      <sequence>
        <element type = "GradualEvolution"
          type = "mpeg7 :GradualEvolutionType"
          minOccurs = "0"/>
        <element name = "SpatioTemporalLocator"
          type = "mpeg7 :SpatioTemporalLocatorType"
          minOccurs = "0"/>
      </sequence>
      <attribute name = "editingLevel" use = "optional">
        <simpleType base = "string">
          <enumeration value = "global"/> <!--Or InterShot-->
          <enumeration value = "composition"/>

```

Par ailleurs, la sémantique est précisée de la manière suivante.

TransitionType	<p>Describes a transition realized between two edited video segments during an editing process.</p> <p>Three different types of transitions are distinguished based on the value of the attribute editingLevel : global transitions, composition transitions, and internal transitions. Three different types of transitions are also distinguished based on the value of the attribute evolution : cuts, analog cuts, and gradual transitions.</p>
GradualEvolution	<p>Describes the transition when it is a gradual (optional).</p> <p>It does not apply to transitions that are cuts and analog residual cuts.</p>

On constate ainsi que tout utilisateur de MPEG-7 se voit proposer une syntaxe et une sémantique pour déclarer l'information documentaire sur le contenu. MPEG-7 est donc bien un format général pour l'indexation hypermédia.

5.6. Conclusion

Du fait du numérique, l'indexation est désormais hypermédia : les contenus sont multimédias et organisés par des hyperliens ou par une juxtaposition dans un même espace de présentation (un écran, une page, etc.). L'indexation a pour but de produire les informations documentaires permettant de rendre exploitables, pour des buts donnés, des contenus hypermédias. Plusieurs enjeux se dessinent ainsi que quelques verrous à lever.

L'interéchangeabilité des données et des métadonnées

L'informatisation des contenus permet d'envisager l'intégration des outils et informations participant aux différentes étapes du cycle de vie des contenus. Fluidifiant la chaîne de production, diffusion, conservation des contenus, l'échange des informations documentaires permet d'intégrer à chaque étape du cycle de vie les connaissances et informations engendrées lors des autres étapes. L'enjeu désormais est autant de créer l'information documentaire que de la rééditer pour l'adapter aux utilisations propres à l'étape envisagée (des métadonnées pour produire, diffuser ou conserver et réexploiter).

L'interopérabilité des outils d'exploitation

L'intégration informatique implique que désormais de mêmes postes informatiques permettent d'effectuer toutes les tâches se rapportant à un contenu, qu'il s'agisse de le créer, de le produire, de l'éditer, archiver, etc. Au lieu d'avoir des stations spécialisées, on s'oriente vers des stations génériques, banalisées, accueillant les applicatifs nécessaires. Mais cette convergence des outils sur une plate-forme générique ne pourra être pleinement effective que si les applications sont interopérables et permettent à l'utilisateur non seulement d'échanger ses informations et métadonnées mais également de passer librement d'une application à une autre pour travailler le contenu.

L'intercompréhension des données et métadonnées

Echanger les informations, dialoguer entre applications ne portent tous leurs fruits qu'à la condition que chaque application recevant des informations puissent disposer de leur sémantique pour optimiser leur exploitation. Or, pour un outil, disposer de la sémantique signifie avoir une représentation formelle du sens qui paramètre son fonctionnement. Différentes solutions sont à l'étude, en particulier autour des ontologies et des formats de représentation comme RDF.

L'intercompréhension des données et métadonnées constitue l'enjeu, comme on le sait, des travaux orientés autour du web sémantique. En effet, puisque l'on sait désormais échanger des informations grâce aux outils fondés sur XML, il s'agit à présent de partager une compréhension commune de ces informations pour

permettre leur exploitation. Sans nul doute le web sémantique deviendra le problème par excellence pour conceptualiser l'indexation hypermédia et expérimenter des solutions⁸.

5.7. Bibliographie

[BAC 00] BACHIMONT B., « L'archive numérique, entre authenticité et interprétabilité », *Archives*, vol. 32, p. 3-15, 2000.

[RAS 94] RASTIER F., CAVAZZA M., ABEILLE A., *Sémantique pour l'analyse*, Masson, Paris, 1994.

[SAU 85] SAUSSURE F., *Cours de linguistique générale*, Payot, Paris, 1985.

8. Ces études ont eu le support de la Région Rhône-Alpes (projet SHIVA de type Emergence). Nous remercions également Brigitte Meillon pour le développement des logiciels de capture et de mesure oculaires ainsi que Michel Nicolaï qui a contribué à une partie des analyses et des résultats dans le cadre de son DEA en sciences cognitives. Pour plus de détails sur les travaux de S. Hollard, B. Meillon et M. Nicolaï, voir le site web du laboratoire Clips : <http://www-clips.imag.fr/>.