
Bibliothèques numériques audiovisuelles

Des enjeux scientifiques et techniques

Bruno Bachimont

*Direction de la Recherche
Institut National de l'Audiovisuel
4, Avenue de l'Europe
94366 Bry sur Marne, Cedex
bbachimont@ina.fr*

RÉSUMÉ. L'audiovisuel connaît aujourd'hui deux mutations profondes : d'une part, les contenus audiovisuels sont de plus en plus reconnus être des documents possédant une valeur patrimoniale et culturelle, dont l'intérêt justifie la conservation et la diffusion ; d'autre part, les technologies numériques investissent le système technique audiovisuel et modifient considérablement les conditions d'instrumentation de ces documents, en particulier leur condition d'accès et de consultation. Au croisement de ces mutations émerge naturellement la problématique de constituer et gérer de vastes ensembles documentaires audiovisuels numériques. À bien des égards, cette problématique renvoie à des enjeux techniques et scientifiques encore non maîtrisés. Nous proposons d'en caractériser les principaux. En particulier nous tentons de montrer comment ils relèvent de difficultés propres à l'audiovisuel et comment les techniques numériques peuvent contribuer à les traiter.

ABSTRACT. Two major evolutions occurs today in audiovisual affairs : on one hand, audiovisual contents are more and more considered to be valuable documents of highest patrimonial and cultural interest. On the other hand, digital technologies are today pervasive in the technical audiovisual system. In particular access and consultation tools of audiovisual documents are deeply modified by those technics. As a result, we face now with the new problem of building and managing vaste bases made up of digital audiovisual documents. This problem motivates technical and scientific issues that are not yet well understood. This paper proposes some of these issues and shows how they come from the very nature of audiovisual and how digital technics can contribute to solve them.

MOTS-CLÉS : audiovisuel, document numérique, MPEG-7, sémiotique, SGML, extraction.

KEY WORDS: audiovisual, digital document, MPEG-7, semiotics, SGML, extraction.

Signature de l'article : nom de la revue. Volume 1 – n° 1/1998, pages 1 à x

1 Introduction

« Ceci tuera cela », on connaît cette sentence de Victor Hugo qui, dans *Notre Dame de Paris*, évoque l'émergence d'une culture reposant sur l'écrit et le livre venant se substituer à une culture fondée sur le symbole, qu'il soit consigné sur la pierre (les cathédrales) ou dans les images et enluminures. Elle est aujourd'hui passablement galvaudée pour évoquer l'émergence d'une culture multimédia venant se substituer à la culture de l'écrit, la télévision et le CD-Rom venant remplacer le livre. S'il est discutable, voire intenable de prétendre que la culture de l'écrit doive s'effacer devant l'émergence des nouvelles formes d'expression et de communication, il n'en demeure pas moins que notre siècle finissant a vu naître de nouveaux modes d'expression, en particulier la photographie, l'audiovisuel et le multimédia, pour respecter l'ordre chronologique de leur apparition.

La thèse opposant l'écrit à l'image ou l'audiovisuel, aujourd'hui le multimédia, se traduit souvent par une gigantomachie millénariste entre la Culture d'une part et la démagogie et l'abrutissement d'autre part, estimant que la prétendue facilité de consultation de l'image et de la séquence audiovisuelle en fait l'apanage des imbéciles. Cela a eu pour principal inconvénient d'empêcher ou de retarder la reconnaissance culturelle que ces nouvelles formes d'expression méritent. À l'exception notable du cinéma, qui a su très tôt se constituer comme un « septième art », ce n'est que récemment que l'on admet que les contenus audiovisuels possèdent, de manière générale une valeur culturelle (Denel, 1998). L'audiovisuel est une forme d'expression qui ne constitue pas seulement une singularité artistique comme le propose l'art cinématographique, mais un authentique mode d'expression pour la connaissance, la réflexion et la culture. Tandis qu'on ne remet pas en cause la valeur culturelle intrinsèque de l'écrit alors que des publications de mauvaise qualité prolifèrent, l'audiovisuel n'a conquis que tardivement le fait d'être un vecteur de connaissance. Si on lui reconnaît cette qualité, les contenus audiovisuels produits par des acteurs privés ou publics doivent par conséquent être conservés et maintenus accessibles en grandes bases documentaires, à la fois pour la constitution d'une mémoire collective que pour l'accès à la culture et à la connaissance.

Constituer de grands fonds documentaires pose de nombreux problèmes. Parmi eux, un problème récurrent est celui de la description et de la documentation du fonds. Les enjeux sont en effet nombreux :

- La description doit être globale pour rendre compte de la finalité du fonds et de sa structure. En effet, le fonds documentaire est une mémoire et représente une somme organisée et articulée de documents. La description documentaire doit rendre compte du contenu et de son articulation pour permettre son évolution et son exploitation. Par exemple, il importe d'avoir des critères de sélection pour agréger de nouveaux documents et trier les anciens. Une représentation globale et raisonnée du contenu du fonds est donc nécessaire pour rendre intelligible sa finalité et sa structure.

- La description doit être locale pour permettre l'accès et la consultation des unités documentaires composant le fonds. En effet, un fonds documentaire atteint rapidement un volume considérable posant le problème pratique de trouver un contenu donné. Il est nécessaire d'avoir une représentation de la nature des contenus et de leur localisation dans le fonds pour le rendre accessible et exploitable. Autrement dit, il faut en faire l'indexation. L'indexation s'effectue en fonction d'une stratégie dépendant de la nature du fonds et de sa fonction : la description globale prescrit ce qu'il faut représenter dans la description locale.

La description documentaire globale du fonds et locale des contenus correspond par conséquent à la problématique de l'indexation. Un fonds documentaire n'est pas seulement un stocks ou un ensemble de documents, mais une mémoire vivante en constante évolution. Ce qui fait d'un fonds une mémoire repose sur son instrumentation, c'est-à-dire son indexation.

Indexer des documents n'est pas tâche aisée, indexer des documents audiovisuels l'est encore moins. L'audiovisuel recèle en effet des spécificités qui rendent son indexation délicate. En particulier, contrairement aux documents textuels de la tradition écrite, les documents audiovisuels sont des documents temporels (ils se déroulent dans le temps) et ne sont pas construits de la combinaison d'un nombre limité de caractères. Autrement dit, les documents audiovisuels sont temporels et non alphabétiques et posent des problèmes spécifiques pour l'indexation.

Les manières de surmonter ces problèmes sont aujourd'hui reconfigurées par l'émergence des technologies du numérique et de leur présence de plus en plus massive dans le système technique de l'audiovisuel. En effet, cela fait désormais longtemps que l'on peut numériser images et sons. Mais l'audiovisuel connaît aujourd'hui ce qu'on peut appeler une seconde révolution numérique, celle du numérique intégral. Il ne s'agit pas tant de disposer d'images et de sons dans un format numérique que d'avoir désormais tout un système technique permettant de produire, constituer, transformer, diffuser, conserver des documents audiovisuels numériques. Le numérique peut être présent à toutes les étapes du cycle de vie des documents audiovisuels. Or le numérique implique par essence, pour son fonctionnement, une indexation des contenus numérisés. Il en ressort une modification importante du statut de la documentation et des conséquences pratiques pour son instrumentation. En particulier, si le numérique est présent à chaque étape du cycle de vie, la documentation doit y être également présente et intégrée à aux outils numériques. Autrement dit, à chaque étape du cycle de vie d'un document audiovisuel, les éléments de contenu qui le composent constituent une bibliothèque numérique. Au lieu d'être traditionnellement un lieu de conservation, on voit que la bibliothèque est un lieu de production : elle constitue une ressource éditoriale pour la production, la diffusion ou l'archivage de documents audiovisuels.

Dans ce nouveau contexte, les bibliothèques audiovisuelles numériques occupent un rôle crucial dans la culture audiovisuelle, tant pour accéder à ses produits que pour en constituer de nouveaux. Les bibliothèques audiovisuelles numériques sont à

la fois un lieu de mémoire et de conservation et un lieu d'écriture et de production. Cette ambivalence, déjà présente dans le système audiovisuel contemporain, sera encore renforcée avec l'ère du numérique.

Mais, pour que ces promesses deviennent des réalités, de nombreux problèmes demeurent. En particulier, parmi les perspectives offertes par l'audiovisuel numérique, certaines renvoient à des problématiques qui appartiennent encore au domaine de la recherche. Ces problématiques se regroupent en thèmes assez bien délimités dont nous présentons une catégorisation. Ces thèmes vont des techniques d'extraction automatique aux problèmes de représentation des connaissances en passant par la présentation graphique des informations pour des outils de consultation et de lecture. Nous en présentons les thèmes nous semblant les plus importants et proposons leur enjeu et les hypothèses de travail nous semblant pertinentes.

L'article est structuré de la manière suivante. Dans un premier temps, nous précisons la problématique des documents numériques et de leur indexation. Dans un second temps, nous présentons les problèmes liés à l'indexation de l'audiovisuel avec ce que le numérique laisse entrevoir des transformations qu'il entraîne. Puis nous énonçons les enjeux qui nous paraissent émerger de ce constat pour nous concentrer ensuite sur les problèmes de recherche qu'ils impliquent. Nous concluons sur les perspectives ouvertes par les bibliothèques numériques audiovisuelles.

2 Numérique et indexation

Cette section revient sur des notions essentielles dont nous aurons besoin pour analyser les problèmes d'indexation de l'audiovisuel et les thèmes de recherche qu'ils suscitent. Premièrement, il faut définir ce que l'on entend par document et par documentation pour définir ce que recouvre la notion d'indexation. Deuxièmement, il faut considérer la mutation qu'entraîne dans la nature des documents leur numérisation.

2.1 Documents, bibliothèques et indexation

La notion de document est en soi banale et ne paraît poser de problème particulier. Il nous faut cependant préciser cette notion dans la mesure où nous aurons besoin de recourir souvent par la suite à ce terme.

Un document est tout d'abord un objet matériel exprimant un contenu. Il est indissociable d'un support matériel (un écran, une feuille de papier, etc.), support d'inscription où un contenu est exprimé. Ce contenu s'exprime en une forme interprétable (lettres, formes iconiques, etc.) pour un lecteur. Cette forme peut être dite sémiotique dans la mesure où elle fait signe pour un lecteur d'un sens qui lui est adressé. Ce qu'on appelle contenu est l'abstraction que l'on peut faire des différents documents exprimant la même chose : les différents exemplaires matériels d'une

même œuvre littéraire par exemple. Le support d'inscription et la forme sémiotique choisie contraignent l'expression du contenu et ses conditions de réception et d'interprétation. En particulier, deux types de contraintes émergent :

- Une structure *logique* qui impose un ordonnancement (une « mise ensemble ») d'éléments de contenu exprimés selon un mode d'expression donné. C'est le choix par exemple d'énoncer une argumentation selon un ordre plutôt qu'un autre. Les structures logiques se formalisent partiellement par les définitions des types de documents que l'on peut rencontrer dans des formalismes comme SGML.
- Une structure *matérielle* imposant une présentation (une « mise en forme ») de ces éléments conformément à l'ordonnancement prescrit. C'est le choix par exemple d'une typographie, d'une mise en page donnée, etc. Cette structure matérielle se formalise partiellement par la notion de feuille de style que l'on rencontre en traitement électronique des documents.

Un document n'est pas plus une structure logique qu'une structure matérielle, mais les deux à la fois. Le fait de caractériser un document comme un objet physique structuré logiquement et matériellement permet d'insister sur les conditions pesant sur la constitution du document et sur sa réception :

- Le document se constitue dans une tradition et selon un contexte de production ; ce contexte peut imposer une structure matérielle et/ou une structure logique ; par exemple les documents juridiques, les comptes rendus hospitaliers ont des structures logique et matérielle imposées par le contexte de production.
- Le document se consulte dans un contexte de réception. Ce contexte impose des critères d'interprétation des structures logiques et matérielles du document. Ainsi, un lecteur sait interpréter « spontanément » les caractères en gras comme des éléments importants. Il y a en général une continuité entre le contexte de production et le contexte de réception dans la mesure où un document est produit pour être consulté dans un contexte de réception donné : un compte rendu hospitalier est fait pour être lu par un autre médecin, une notice documentaire par une documentaliste. Cette continuité est rompue quand le contexte de réception n'est pas celui visé par le contexte de production. Il se pose alors moult problèmes d'interprétation et de compréhension comme en témoignent les problèmes d'interprétation historique et/ou philologique.

Par conséquent un document n'est pas la simple juxtaposition des éléments qu'il contient. Conçu dans un contexte de production donné, consulté dans un contexte de réception, réalisé conformément à des structures logique et matérielle, un document en tant que tel conditionne l'interprétation des éléments qu'il contient. Un document cristallise dans ses structures et contextes des connaissances et des prescriptions interprétatives pesant sur la lecture¹ à effectuer.

¹ Nous parlons de manière générique d'écriture d'un document et de lecture pour tous les documents, qu'ils soient textuels, audiovisuels, ou multimédia, sans entrer

En conclusion, un document, c'est un support matériel d'inscription, une forme sémiotique (ou plusieurs : documents texte et image), et des structures logique et matérielle.

Les documents sont généralement assemblés en des collections pour répondre à des objectifs pratiques déterminés. C'est par exemple la documentation d'un travail de recherche, la collection des documents d'un domaine de compétence particulier, etc. On obtient ainsi ce que l'on peut appeler, d'une manière trop rapide, bibliothèque. Il faudrait bien sûr distinguer ce qu'est une bibliothèque des autres collections de documents. Mais nous préférons adopter ici une caractérisation très générale des bibliothèques : collection de documents réunis pour un objectif donné.

Il s'agit donc d'exploiter les documents de la bibliothèque pour atteindre l'objectif donné. Cela implique qu'il faut être capable de déterminer la valeur de chaque document par rapport à cet objectif et de le décrire en fonction de cet objectif. Cela permet alors de retrouver le document en fonction de l'intérêt qu'il représente. Autrement dit, il faut paraphraser le document dans les termes pertinents pour son exploitation. On obtient ainsi une définition de l'indexation des documents dans une bibliothèque :

Une indexation est la paraphrase d'un contenu en une forme sémiotique interprétable permettant de rendre exploitable le contenu indexé dans le cadre d'une pratique donnée.

L'indexation part d'un *document source*, appartenant à la bibliothèque, qu'il faut indexer en le paraphrasant. Cette paraphrase constitue un *document de description*. En effet, l'indexation d'un document produit un autre document. Le document de description est bien un document dans la mesure où il emprunte une forme sémiotique. Il possède une structure logique et une structure matérielle. Ses conditions de production et de réception sont bien sûr d'une part la tâche d'indexation elle-même et d'autre part la tâche de recherche et consultation des documents sources. De manière générale, un document n'est pas en soi un document source ou un document de description mais cela dépend du contexte de son usage. Un même document peut être considéré comme un document source ou comme un document de description selon les cas (nous y revenons plus loin). Par exemple, un article de la presse de télévision peut être lu pour lui-même, c'est un document source, mais utilisé également pour indexer l'émission audiovisuelle évoquée, et c'est alors un document de description.

L'indexation est toujours relative à une pratique donnée, c'est-à-dire à des usages qui gouvernent l'exploitation qui doit être faite des documents indexés. La pratique visée la plus habituelle est la recherche d'information : l'index permet de retrouver le contenu auquel il est associé. Mais les pratiques peuvent être plus générales et plus diversifiées. Ce peut être par exemple la possibilité d'indexer un

dans le débat sur la question de savoir s'il s'agit bien d'écriture et de lecture pour le son ou le multimédia. Nous précisons le type d'écriture ou de lecture quand cela est nécessaire.

contenu par les droits d'auteur qui y sont associés dans la perspective d'une exploitation commerciale ou de diffusion de ce contenu. Ce peut être aussi une exploitation particulière comme l'exploitation pédagogique : les contenus sont indexés par exemple par le programme scolaire auquel ils correspondent, permettant ainsi à un professeur d'accéder aux contenus relevant du programme qu'il doit enseigner. Ces usages peuvent évidemment se croiser : on peut rechercher des contenus pour lesquels les droits sont disponibles et renvoyant à un programme scolaire de Terminale.

L'indexation est donc une *interprétation* d'un contenu, une réécriture ou *reformulation* dans une forme propre à l'exploitation du contenu dans un contexte.

2.2 Document numérique

La notion de document numérique apparaît quand le support matériel de l'expression devient numérique. Un document numérique est donc un document inscrit sur un support informatique. D'un point de vue théorique, le numérique se caractérise par le fait de manipuler un système d'unités discrètes indépendantes les unes des autres par des règles formelles de type algorithmique. Par conséquent, un document numérique est une structure formelle constituée d'unités discrètes à laquelle est associée des règles algorithmiques formalisant ses propriétés documentaires (structures logique et matérielle), en particulier le fait le pouvoir être lu ou consulté de telle manière que son contenu soit accessible par un utilisateur-lecteur. Plus précisément, le document numérique met en jeu les distinctions suivantes :

- Le support d'enregistrement du document :
Le document numérique est codé par une structure discrète qui est enregistrée et conservée dans une mémoire informatique.
- La forme d'enregistrement :
L'enregistrement numérique se fait selon un format de codage, qui permet d'établir la correspondance entre le document tel qu'il doit être consulté ou lu, et sa structure interne.
- Le support d'appropriation du document.
Le document numérique doit pouvoir être lu ou consulté. Or, l'enregistrement numérique est en mémoire et ne peut être accédé directement. Il faut pour cela qu'un calcul produise à partir de l'enregistrement une représentation lisible ou intelligible pour un utilisateur. La représentation intelligible s'affiche ou se produit sur un support où l'utilisateur accède *directement* au document. Directement signifie ici que l'utilisateur peut consulter le contenu sans devoir faire appel à une médiation calculatoire, c'est-à-dire à l'exécution d'un programme. Ce support est un support d'appropriation, car l'utilisateur s'approprie le contenu du document. Les supports d'appropriation sont par exemple l'écran, des hauts parleurs, le papier.
- La forme 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'appropriier le contenu. Ce dernier accède au contenu sans le recours à la médiation du calcul, mais à la médiation d'un apprentissage de la forme d'appropriation. Par exemple, il a appris à lire, et la forme alphabétique d'affichage du document est une forme d'appropriation intelligible pour lui. Quand il y a plusieurs formes d'appropriation, 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 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.

L'audiovisuel connaît de semblables distinctions : de même que pour les documents numériques, la forme et le support d'enregistrement sont différents de la forme et du support d'appropriation. La représentation canonique que nous avons communément des documents correspond aux documents textuels consignés sur un support imprimé ou un support papier. Pour ces derniers, les supports d'enregistrement et d'appropriation sont confondus. De même, les formes d'enregistrement et d'appropriation. Autrement dit, ce qui est conservé est identique à ce qui est consulté, ce qui est enregistré est le même que ce qui est lu. Dès les origines, l'audiovisuel s'est constitué comme une séparation entre le document conservé et le document consulté. En effet, l'audiovisuel est un procédé d'enregistrement et de reproduction. Reproduction : reproduction de l'enregistrement en d'autres enregistrements (copie de film), mais aussi reproduction au sens où le document est produit à partir de l'enregistrement, il est projeté (projection d'un film) sur un support d'appropriation (« écran de projection »). Néanmoins, une différence demeure : l'enregistrement interne d'un document audiovisuel non numérique, le film, n'est pas un enregistrement calculable, mais un objet physique sans propriété computationnelle particulière. En numérisant l'audiovisuel, le statut de l'enregistrement deviendra très différent.

Nous avons défini plus haut le document comme étant une mise ensemble d'éléments de contenu dans le cadre d'une structure matérielle. Ces structures logique et matérielle potentialisent les effets de sens entre les éléments pour les utilisateurs. Puisque le document est une structure logique mise en forme matériellement, le document correspond nécessairement à ce qui est consulté dans le cadre d'une forme d'appropriation sur un support d'appropriation. Il ne correspond pas à l'enregistrement interne, contrairement à ce que la locution de « document numérique » pourrait laisser entendre.

Cela implique qu'un document numérique n'est un document que lorsqu'il est consulté sur un support d'appropriation à travers une forme d'appropriation.

Autrement, il n'est qu'un enregistrement. Le document numérique, considéré sur le support d'enregistrement, n'est pas un document mais une ressource à partir de laquelle peuvent être calculés autant de documents, c'est-à-dire de formes d'appropriation projetées sur un support d'appropriation. Cela est parfaitement illustré par la fragmentation d'un fichier sur les unités de stockages d'un ordinateur où pour l'afficher ou l'imprimer, des opérations sont nécessaires pour regrouper, selon les algorithmes de gestion de la mémoire, les fragments composant le fichier.

Ces distinctions permettent alors de définir les opérations suivantes de manières précises :

- La projection :
Il s'agit des calculs permettant de passer de l'enregistrement au document, c'est-à-dire de la forme interne à la représentation intelligible pour l'utilisateur. C'est par exemple l'affichage à l'écran. La projection est une opération du calculable à l'intelligible.
- L'abstraction :
Il s'agit des calculs permettant de passer d'une représentation intelligible à une forme interne d'enregistrement, pouvant donner lieu à d'autres projections. C'est par exemple l'opération de numérisation. L'abstraction est une opération de l'intelligible au calculable.
- La transformation :
Il s'agit des calculs permettant de transformer le format interne d'enregistrement en un autre enregistrement. Il s'agit en fait de tous les calculs possibles ne conduisant pas une forme d'appropriation. Parmi les transformations, on comptera la reconnaissance des caractères (après la numérisation, on utilise un algorithme de reconnaissance), la transformation SGML des documents structurés, etc. La transformation est une opération du calculable au calculable.
- La navigation :
Il s'agit des opérations permettant de passer d'une forme d'appropriation à une autre. Ainsi, en cliquant sur un contenu *affiché*, on obtient un nouveau contenu *affiché*. La navigation n'est pas obligatoirement le passage d'une forme projeté à une autre forme projeté du même enregistrement, mais également d'autres enregistrements. La navigation est une opération de l'intelligible à l'intelligible.

2.3 *Indexation numérique*

Quel rapport entre le numérique et l'indexation ? Si l'on considère la nature même des systèmes numériques : les contenus numérisés sont stockés en mémoire et ne sont accessibles que par leur adresse. Autrement dit, un contenu numérique n'est exploitable qu'à travers un index qui y donne accès. Le numérique, en son essence, articule contenu et index. Ce principe se prolonge jusque dans la pratique des systèmes numériques : les gestionnaires de fichiers habituels (ceux de Windows ou

de MacOS) proposent une organisation des fichiers et des systèmes d'index pour les retrouver et les consulter. L'indexation est au cœur du numérique et lui est consubstantielle. Tout système numérique est un système documentaire et mobilise une indexation.

Par ailleurs, que change le numérique dans la perspective de l'indexation ? De manière fondamentale, le numérique modifie le statut de l'indexation. Alors que l'indexation paraphrase un contenu pour le rendre accessible pour un usage donné, le numérique permet d'utiliser l'indexation pas seulement pour exploiter le document mais pour le transformer. En effet, le numérique éclate l'entité documentaire en un enregistrement interne d'une part et une forme d'appropriation recalculée d'autre part. En effectuant une indexation, on associe aux éléments composant l'enregistrement interne une description plus ou moins précise. Puisque l'on est au niveau de l'enregistrement interne, c'est-à-dire au niveau du calcul, on peut utiliser la description pour déterminer des calculs à effectuer sur le document source. Par exemple, les balises de SGML enrichissent un texte en explicitant sa structure logique : les parties du texte sont encadrées par des balises qui définissent le statut logique de ce qu'elles encadrent. <titre> un titre </titre> définit que « un titre » a la valeur logique de titre. Par conséquent, il devient possible de manipuler « un titre » en fonction de sa valeur logique et d'y effectuer toutes les opérations impliquées par cette valeur. Ainsi, on peut construire un nouveau document contenant tous les titres des documents composant la collection (une espèce de table de matières). L'indexation, principe du génie documentaire, est à la base d'un génie éditorial.

Par conséquent, on constate que la fonction désormais première de l'indexation n'est pas la recherche des documents mais leur réédition et leur transformation. Documenter, c'est instrumenter les documents sources pour de futures rééditions.

3 Audiovisuel et indexation

Nous nous intéressons ici aux documents numériques audiovisuels. Nous caractérisons ici ce que nous entendons par audiovisuel pour aborder ensuite ses spécificités et terminer sur les difficultés qu'elles entraînent.

3.1 *Qu'est ce qu'un contenu AV ?*

Nous nous intéressons aux documents audiovisuels. Audiovisuel signifie que les documents se composent d'images animées et de sons. L'ère des documents électroniques nous ont habitués à rencontrer des documents audiovisuels mêlés à d'autres types de documents, comme les documents multimédia. Cependant, audiovisuel et multimédia ne sont pas synonymes. On retiendra ici la distinction suivante :

- Audiovisuel :

Est audiovisuel un document composé d'images animées et/ou de sons se déroulant de manière linéaire selon un rythme temporel particulier le rendant intelligible pour un lecteur. C'est par exemple le cinéma, la télévision, la vidéo pour les documents audiovisuels, et la radio pour les documents purement sonores.

- Multimédia :

Est multimédia un document mêlant d'une part différents médias et d'autre part construit à partir de différents sous-documents unis par un réseau de liens. En particulier, les différentes vidéos qu'il comprend possèdent des temporalités différentes et ne constituent pas d'unité audiovisuelle. C'est par exemple un texte auquel à partir duquel on établit des liens vers des photographies ou des vidéos.

3.2 *Les spécificités de l'audiovisuel*

Quand elle s'applique à des contenus audiovisuels, la pratique de l'indexation rencontre des contraintes spécifiques à la nature audiovisuelle des contenus qui surdétermine son exercice. Ces contraintes sont diverses et induites par les différentes caractéristiques de l'audiovisuel :

- les contenus audiovisuel sont constitués *d'images* ;
- les contenus audiovisuels sont des *objets temporels* ;
- les contenus audiovisuels mêlent *image et son*.

Ces trois propriétés entraînent des difficultés qui surdéterminent la pratique documentaire.

3.2.1 *Les contraintes de l'image*

L'image constitue une forme sémiotique particulière dans la mesure où la forme signifiante n'est pas arbitraire par rapport au contenu signifié. Contrairement au signe linguistique pour lequel la relation entre forme signifiante et contenu signifié est arbitraire, il existe une relation forte entre la forme de l'image et le contenu qu'elle signifie. C'est ce que l'on évoque habituellement en parlant du caractère analogique des images (Metz, 1968).

Les images se caractérisent principalement par un aspect morphologique analogue à la réalité mondaine. Elles fonctionnent dès lors comme une pseudo-réalité qu'elles montrent en un rapport analogique et non comme une signification qu'elle signifie en un rapport arbitraire. Le signe qu'est l'image est une *signe qui montre* mais non un *signe qui dit*.

Or, à l'instar de la réalité qu'elle prétend représenter, l'image ne peut dire par elle-même ce qu'elle signifie et doit reposer sur une paraphrase langagière pour gagner l'intelligibilité qui lui manque. C'est pourquoi tout contenu visuel doit comporter un commentaire ou une légende. Contrairement au texte, l'image ne

possède pas de niveau zéro² de sa signification où la langue, via son système fonctionnel³, prescrit une signification. Les images ne prescrivent pas de sens car les images ne constituent pas de système fonctionnel que viendraient déterminer les normes culturelles de lecture. Plus exactement, l'image peut s'interpréter si l'on fait appel aux conditions qui ont présidé à sa conception et aux conditions de son interprétation ou lecture. Mais la constitution technologique de l'image ne permet pas d'objectiver dans sa structure des prescriptions interprétatives pesant sur toute lecture comme un a priori dont il faut tenir compte par la suite : par exemple, l'image n'est pas (encore) construite comme la combinaison d'objets dont le découpage objectif correspond à la détermination de ce qui fait sens, d'un certain point de vue, dans l'image.

La comparaison de l'image et du texte peut être éclairante sur les difficultés propres à l'image. D'un certain point de vue, le texte est sa propre indexation : il est constitué d'unités segmentées par des signes de ponctuation ou des blancs si bien que ces unités peuvent dès lors être utilisées comme index ; c'est l'indexation en texte intégral. Le texte ne détermine pas bien sûr quelle interprétation il doit recevoir, il peut susciter un nombre potentiellement infini d'interprétations plus ou moins précises, plus ou moins redondantes ou cohérentes entre elles. Ce qui fait signe dans un texte, les unités linguistiques signifiantes, ne sont pas données a priori mais sont déterminées à la suite d'une première interprétation : le signe est le résultat d'un parcours interprétatif et non sa condition (Rastier et al., 1994). Mais la mise en forme graphique du texte prescrit une interprétation selon laquelle les unités signifiantes sont les suites de lettres séparées par des blancs, *i.e.* les mots. Cette prescription fonctionne fort bien, et c'est un préjugé fort tenace du sens commun et

² Cela ne signifie pas qu'un texte possède un sens « littéral ». Le fait qu'il y ait des prescriptions sémiotiques inhérentes au système fonctionnel de la langue n'implique en effet pas qu'un texte signifie d'abord, par son sens littéral, ce que ce système prescrit, pour ensuite, par son sens figuré, signifier un sens prescrit par le contexte. Au contraire, les prescriptions fonctionnelles sont renégociées par le contexte de leur emploi dès lors qu'elles sont utilisées, si bien qu'il est impossible de dégager un sens littéral. Le degré zéro du texte n'est donc pas un degré zéro de l'écriture. Autrement dit, ce que le système fonctionnel propose, ce sont des significations, c'est-à-dire des contenus idéalisés et rationalisés par les linguistes et lexicographes, et ce qu'un texte signifie, c'est un sens, c'est-à-dire la contenu contextuel. La signification n'a pas de sens, mais idéalise ce dernier en figures non contextuelles, comme tentent de le faire nos dictionnaires habituels.

³ Le système fonctionnel d'une langue correspond aux significations et aux unités linguistiques qu'un locuteur trouve disponible dans sa langue pour s'exprimer et les modifier selon ses usages. Les normes culturelles imposent certains emplois et certaines inférences ; elles sont cependant distinctes du système fonctionnel dans la mesure le système fonctionnel permet d'inférer *par défaut* un sens alors que les normes culturelles doivent être explicitement requises par le contexte ; voir sur ces points (Rastier, 1987; Rastier, Cavazza, & Abeillé, 1994).

de certaines linguistiques que les mots sont des unités signifiantes données a priori, existant comme des réalités objectives de la langue. Or on sait comment la ponctuation et les blancs séparateurs sont des acquis technologiques de l'écriture relativement tardifs (Ilich, 1991) (Guglielmo & Chartier, 1997). Le texte est sa propre indexation car l'interprétation graphique effectuée par son écriture prescrit une sémiotisation⁴, *i.e.* la détermination d'unités signifiantes faisant sens, qui peuvent dès lors être reprises comme index.

Pour l'image, la mise en forme matérielle ne correspond pas à une sémiotisation. En effet, l'image ne repose pas sur une discrétisation mobilisant différentes unités signifiantes enrôlées dans un système fonctionnel et culturel. Puisque l'image ne prescrit pas par elle-même une interprétation, il faut lui adjoindre une prescription interprétative explicite, une sémiotisation explicite, dès lors que l'on veut exploiter les images en fonction d'une certaine valeur ou signification. Ainsi la pratique actuelle de la documentation à l'INA repose-t-elle sur la description linguistique puis en mots clés des programmes audiovisuels. Une documentaliste visionne un programme et décrit de manière linguistique en temps réel le contenu. On utilise la langue comme système sémiotique pour rendre intelligible et accessible le contenu audiovisuel.

Tout cela n'est toutefois pas tout à fait exact : les conditions technologiques de production (le système technique), les conditions herméneutiques de production (le contexte culturel de l'auteur et de son temps) sont autant de grilles de lecture de l'image et fournissent autant de prescriptions interprétatives et de sémiotisations potentielles. C'est ainsi qu'une philologie de l'image est possible et permet de décrire l'image pour elle-même, c'est-à-dire dans son contexte de production. Par ailleurs, on assiste également à des mutations technologiques qui viennent sensiblement modifier les conditions sémiotiques de l'image. En considérant l'image comme étant construites de plusieurs objets, les normes récentes⁵ proposent une sémiotisation a priori qui peut fonctionner comme base pour l'interprétation par le lecteur. Non seulement comme une base, mais comme une contrainte, car cette sémiotisation est décodée par les « players » et imposée à toute lecture.

En conclusion, la contrainte exercée par l'image est que le contenu audiovisuel doit se paraphraser textuellement pour renvoyer à une interprétation et une pratique

⁴ La sémiotisation est le processus interprétatif selon laquelle une certaine unité est définie comme faisant sens ; cette unité est reconnue fonctionner comme un *signe* et comme telle, posséder une valeur *sémiotique*.

⁵ La norme MPEG4 propose de structurer le flux audiovisuel en objets reliés entre eux par des relations spatio-temporelles. Il en résulte que l'image et le flux AV sont désormais discrétisés en objets, en unités permettant de coder l'image et le flux. Les objets MPEG4 codant un flux AV ne sont pas en soi des unités de sens ; mais ils proposent des unités pour une sémiotisation a priori d'origine technologique.

donnée. *Toute pratique documentaire de l'audiovisuel doit donc comprendre une réécriture textuelle de l'image pour rendre possible son accès.*

3.2.2 Les contraintes des séquences temporelles

Les contenus audiovisuels sont des séquences temporelles d'images et de sons. La nature temporelle de ces séquences est particulièrement importante et conditionne elle aussi la pratique de l'indexation. Les séquences audiovisuelles sont des « objets temporels », en entendant par cette locution deux notions qui se recourent :

- une acception *empirique* selon laquelle le rythme de lecture d'une séquence audiovisuelle est imposé par le document lui-même et non par le lecteur, contrairement à d'autres types de documents, comme les documents textuels ou les images ;
- une acception *phénoménologique* selon laquelle le flux audiovisuel coïncide avec le flux de conscience ; cette coïncidence permet, à travers le flux documentaire d'accéder au flux de la conscience et à ses structures ;

La temporalité des séquences audiovisuelles a pour conséquence qu'il est impossible de consulter rapidement un document, de le « feuilleter » pour trouver un élément que l'on recherche : pour feuilleter une heure de vidéo, il faut une heure. En conséquence, l'usage documentaire est de retarder le plus longtemps possible la consultation directe du document audiovisuel pour consacrer davantage à la consultation indirecte, reposant sur la consultation de description documentaire du contenu audiovisuel. La description documentaire permet de ne consulter qu'un segment pertinent vis-à-vis de la recherche de l'utilisateur.

La temporalité des contenus audiovisuels modifie sensiblement le problème de l'indexation tel qu'il se pose pour les images. Le principal problème rencontré avec l'indexation des images est que l'image ne propose pas de sémiotisation *a priori* empruntée à un système fonctionnel permettant une première indexation où chaque unité signifiante déterminée par cette sémiotisation peut servir d'index. Or, les contenus AV ne sont pas purement temporels, au sens d'une continuité temporelle. Ils résultent d'une succession d'images, procédant d'une discrétisation du temps. Dès lors, on comprend que le contenu AV se constitue à partir d'une sémiotisation correspondant à la discrétisation du flux : le découpage en image prescrit quelles sont les unités signifiantes. C'est pourquoi tant d'analyses du cinéma ne considèrent que les images : ce sont les unités signifiantes proposées par le système technique, à l'instar des linguistiques qui ne considèrent que les mots, unités prescrites par le découpage technologique de l'écriture. Le paradoxe, c'est que ces analyses ignorent la temporalité alors que c'est un travail technique sur cette dernière qui permet leur possibilité.

Les unités signifiantes sont définies à différents paliers : l'image, le plan, la séquence. Le plan correspond aux blancs de l'imprimerie : prescrit et proposé *a priori* par le monteur, leur interprétation peut modifier leur pertinence et redéfinir

leur extension temporelle⁶. Toute segmentation est relative et ne fonctionne au plus que comme prescription (et non obligation) interprétative. La sémiotisation implique, par la discrétisation en image, plans, séquences, que l'indexation doit étudier particulièrement la structure temporelle des contenus AV et déployer la même finesse d'analyse (et non la même analyse) que pour l'étude de la structure logique et documentaire des textes.

La possibilité de l'accès direct à un segment particulier apporté par le numérique permet en principe de ne consulter directement que les segments pertinents : la seule limite rencontrée est celle de la précision des descriptions documentaires. Dans le domaine de l'image, la description documentaire permet d'être suffisamment précis pour retarder le plus longtemps possible d'avoir à consulter le contenu décrit, ce qui requiert par définition un temps aussi long que le contenu lui-même. Mais il n'en est pas de même pour le son : étant donné la difficulté à décrire un contenu sonore, il est parfois plus simple et plus rapide d'écouter les premiers instants d'une séquence sonore que de parcourir des descriptions dont la terminologie est parfois floue ou alors trop particulière au concepteur de la description.

3.2.3 *Les contraintes de l'audiovisuel*

L'audiovisuel est, comme son nom l'indique, un mélange de son et d'image animée. Cela pose la difficulté que le contenu véhiculé par une séquence audiovisuelle n'est pas purement visuel ni sonore. Cela se traduit par les difficultés suivantes :

- le contenu sonore n'est pas le contenu visuel : la description de ce qui est dit ou entendu n'est pas la description de ce qui est vu à l'image ; souvent, en particulier dans le flux télévisuel, la bande son est une interprétation abstraite de ce qui est montré à l'image qui est alors plus factuelle ;
- le contenu d'une séquence est produit par la synthèse son et image ; il est difficile de séparer les composantes si bien que la description documentaire parvient rarement à être pertinente et surtout à savoir repérer une *unité documentaire* qui soit une *unité de sens* dépendant du contexte.

Les contraintes propres à l'audiovisuel comme unité constituée d'image animée et de son renvoient à la difficulté de décrire une unité de sens qui n'est ni une unité visuelle ou une unité sonore, mais un mixte des deux (Chion, 1994; Chion, 1998).

3.3 *Audiovisuel et indexation numérique*

Comme on l'a vu plus haut, l'accès au contenu audiovisuel repose en grande partie sur la documentation qui l'accompagne. Cet accès doit être compris au sens

⁶ Ainsi, un plan peut posséder une extension telle qu'il doit être découpé en unités de sens inférieur, comme dans le cas des « plans séquences » ; au contraire, les plans doivent s'agréger en unités plus vastes pour prendre leur sens et fonctionner comme signe.

large : il s'agit à la fois de l'accès physique ou matériel au document lui-même, comme on le voit dans les bibliothèques ou dans les bases documentaires. Il s'agit aussi de l'accès au contenu, c'est-à-dire de la compréhension et de la réception du document. Pour retrouver physiquement et comprendre un contenu audiovisuel, il faut une documentation. Or, le numérique, comme système technique, reconfigure les relations entre les documents audiovisuels et leur documentation. En particulier, nous faisons l'hypothèse que le numérique permet la mise en œuvre de quatre principes fondamentaux pour la documentation :

- l'intégration de la documentation ;
- l'alignement de la documentation ;
- l'intégration de la chaîne documentaire ;
- l'intégration de la documentation et des contenus.

3.3.1 *Intégration de la documentation*

La documentation, dans le contexte analogique, est spécifique à chaque usage. Dans le contexte numérique, il est possible d'intégrer les différents niveaux de description documentaire en une même documentation structurée, mise en correspondance avec le contenu audiovisuel. C'est par exemple la possibilité d'intégrer la description du contenu et la description des droits attachés au contenu de manière à pouvoir accéder, exploiter, utiliser un segment audiovisuel en fonction de son interprétation en termes de contenu et son interprétation en termes de droits.

3.3.2 *Alignement de la documentation*

La documentation peut désormais être alignée sur le flux audiovisuel. On entend par alignement le fait que tout segment ou structure de la description documentaire renvoie à un segment audiovisuel. Réciproquement, tout segment audiovisuel peut être plongé dans une unité documentaire possédant une description permettant de l'exploiter. C'est par exemple le fait de pouvoir considérer un extrait donné, de prendre la plus petite unité documentaire le contenu (une séquence) et consulter la documentation associée, concernant par exemple les droits associés et les autres contenus corrélés. En conséquence, on assiste à une autonomisation des contenus locaux ou segments des documents audiovisuels. Ils deviennent accessibles et manipulables indépendamment du document global dont ils sont issus. Cette tendance ne va pas bien sûr sans poser de nombreux problèmes sur le statut de l'œuvre.

3.3.3 *Intégration de la chaîne documentaire*

La documentation est habituellement spécifique à chaque moment de la chaîne documentaire : production, diffusion ou archivage. La chaîne numérique documentaire permet d'intégrer et de rendre interopérables les éléments techniques composant ces différents moments. Ainsi, les documents de la production (EDL,

projet du montage) peuvent en principe être conservés pour être utilisés pour l'archivage. En conséquence, le cycle de vie de la documentation est parallèle à celui des documents audiovisuels. Au lieu d'être à chaque fois réinventée, la documentation est reformatée et rééditée à chaque étape du cycle de vie à partir de la documentation produite lors des étapes précédentes. Cette tendance a des conséquences importantes sur la nature du travail de documentation : de la même manière que le numérique modifie le statut des archives en ressources éditoriales, la documentation devient une ressource pour la constitution de documentations adaptées à d'autres usages et fonctions.

3.3.4 *Intégration de la documentation et des contenus indexés*

La conception d'une nouvelle chaîne documentaire numérique rend possible d'accéder aux documents audiovisuels munis de leur documentation puisque désormais la documentation peut être inscrite dans le flux ou bien fortement articulée ou alignée à ce dernier. Cela signifie que la description documentaire ne sera plus seulement une documentation ne possédant pas d'autre valeur que celle de décrire un contenu audiovisuel, mais devient un document à part entière, consultable pour lui-même, à côté du document audiovisuel qu'il décrit.

Le problème n'est plus désormais de décrire un contenu audiovisuel mais d'articuler des documents audiovisuels, sonores, textuels, iconiques, etc., entre eux de manière à ce qu'ils puissent constituer chacun une description documentaire des autres, en y donnant accès selon une certaine interprétation. Par exemple, un journal télévisé peut être documenté par un article de presse qui peut à la fois permettre d'accéder au contenu télévisuel et en être un commentaire. C'est là tout le sens qu'il faut donner à ce nouveau slogan que l'on rencontre désormais à propos des nouvelles normes de descriptions et de structuration documentaires : *metadata are data*.

Autrement dit, la documentation au sens classique du terme ne sera plus qu'un élément parmi d'autres de la documentation des documents audiovisuels numériques. La documentation devra permettre la navigation et la consultation de tout un ensemble documentaire fortement structuré et interrelié. C'est par exemple le fait d'avoir une documentation alignée sur une segmentation des contenus audiovisuels permettant d'articuler ces derniers aux programmes scolaires, à d'autres documents audiovisuels associés, à des documents de référence, des commentaires, etc. Ces documents sont tous des data, c'est-à-dire qu'ils peuvent se lire pour eux-mêmes, indépendamment du fait qu'ils soient ici des descriptions documentaires. Cependant, ils sont lus ici dans la perspective de l'usage d'un contenu audiovisuel dont ils constituent les conditions d'intelligibilité.

4 Les enjeux du numérique AV pour les bases documentaires

Quels enjeux se dégagent de cette reconfiguration des conditions de l'indexation par le numérique ? Les remarques précédentes permettent de conclure que l'indexation de documents audiovisuels numériques nécessite la résolution de deux problèmes fondamentaux :

- la localisation : il s'agit de localiser dans le document une unité dont l'indexation fera la paraphrase. C'est le problème de l'alignement de la documentation sur le document décrit.
- la qualification : il s'agit de qualifier cette unité. C'est le problème d'articuler les différents niveaux de description d'une part et les différents documents de description d'autre part.

La localisation définit une unité signifiante, la qualification détermine son interprétation, c'est-à-dire une espèce de signifié. En fonction de ces deux problèmes fondamentaux, les enjeux pour la constitution et l'exploitation d'une bibliothèque numérique audiovisuelle sont les suivants :

- Comment définir la localisation dans un document audiovisuel ?
- Comment définir la structure d'un document audiovisuel ? C'est le problème de qualifier en termes de structure les localisations effectuées.
- Comment établir un format d'échange des documents de description pour permettre l'intégration documentaire de la chaîne de production audiovisuelle ?
- Comment dégager des documents audiovisuels des unités pouvant constituer une première indexation. ? Comment dégager des objets du flux audiovisuel pouvant constituer autant d'index de ce flux ?
- Comment qualifier les unités significatives déterminées et les localisations effectuées ? Quelle terminologie utiliser ? Quels notions ou concepts y correspondent ?
- Comment utiliser l'indexation pour rechercher et consulter une base de documents audiovisuels ?
- Comment exploiter le contenu de cette base et se l'approprier ?

Chacun de ces points renvoie à une problématique de recherche.

5 Les problèmes de recherche

5.1 Structuration et description documentaire

L'enjeu de cette recherche est de disposer d'un moyen de décrire le contenu d'un document audiovisuel. Il faut distinguer deux niveaux :

- Le moyen d'exprimer une description : c'est le problème de définir un langage de description tel que chaque unité de la description puisse être associée à une partie localisée du document audiovisuel ;
- Les termes de la description : c'est le problème de savoir quels concepts ou quels objets peuvent être utilisés pour qualifier ou indexer le contenu.

La recherche sur la structuration et la description documentaire ne porte ici que sur le premier point. Inspirée très fortement des acquis documentaires du domaine textuel, cette recherche vise à définir un méta-langage (comme SGML pour les documents textuels) permettant de déclarer le type de description que l'on veut effectuer. Cette recherche est très active ces dernières années et suscite un intérêt très vif dans le cadre de MPEG-7. Les principaux problèmes sont les suivants :

- Contrairement aux formats de description des documents textuels, l'indexation ne peut s'effectuer par des balises qui enrichissent le texte. Les balises sont dans le texte si bien que la localisation ne pose pas de problème particulier. Pour l'audiovisuel et le multimédia, la documentation n'est pas dans le document. Il faut donc explicitement relier la documentation à aux parties décrites du document.
- Les formats habituels de description documentaire prescrivent des structures arborescentes. Cela n'est pas adapté en général aux documents audiovisuels dans la mesure où par exemple la structure de la bande son n'est pas celle de la structure image.

L'hypothèse de travail que l'on peut proposer est de définir des structures génériques permettant d'articuler des éléments de description aux éléments décrits. Ces structures formalisent des propriétés génériques des documents audiovisuels comme par exemple le fait qu'une description renvoie généralement à un segment temporel du document. Cette hypothèse est développée dans le langage AEDI développé à l'INA dont on trouvera une description dans (Auffret et al., 1998).

5.2 *Segmentation et extraction AV*

L'enjeu d'une telle recherche est de disposer d'outils d'analyse du flux audiovisuel permettant d'extraire plus ou moins automatiquement des objets ou des unités pouvant constituer des index, ou dont on peut tirer des index. Il s'agit principalement des algorithmes de segmentation de plans et de séquences, et des algorithmes de détection de mouvement, visages, objets, etc. Le principal problème rencontré est que

- les descripteurs extraits automatiquement du flux d'image et de sons sont « *pixel-based* » et reflètent davantage les propriétés du signal visuel et sonore véhiculant le contenu que les propriétés du contenu lui-même ; les descripteurs extraits sont des descripteurs physiques et concrets du flux AV qui résultent d'une analyse et non d'une interprétation du contenu.
- les descripteurs utiles d'un point de vue documentaire, c'est-à-dire correspondant à une paraphrase ou réécriture utile du contenu dans la perspective d'un usage donné, sont de nature symbolique et abstraite ; ces descripteurs résultent d'une interprétation du contenu et non d'une analyse.

La description symbolique peut ne pas correspondre à des éléments factuels du contenu AV, et dans ce cas, la situation est sans espoir pour l'extraction automatique ; mais il se peut que la description abstraite ne soit qu'une abstraction symbolique d'une représentation physique de l'image ;

Il faut donc définir les descripteurs symboliques utiles au niveau documentaire possédant une traduction au niveau physique pour déterminer quelles méthodes d'analyse et d'extraction permettent de les détecter automatiquement.

Il s'agit d'une problématique de *mapping* entre le symbolique et le physique ou, en termes plus philosophiques, d'une problématique de la schématisation des concepts de la description documentaire en modèles physiques du flux AV. Ce problème ne résulte pas seulement d'une insuffisance de l'état de l'art que le progrès technologique pourrait combler. Il s'agit au contraire d'une différence de principe et de nature opposant *l'analyse* d'image de *l'interprétation* d'image, les descripteurs physiques extraits et les descripteurs symboliques abstraits. Le but de la recherche est alors :

- détermination des descripteurs symboliques adéquats selon les types d'images et de flux ;
- pilotage des algorithmes d'extraction et de segmentation en fonction de la nature des descripteurs symboliques recherchés.

Par exemple, on sait bien détecter automatiquement la présence des boutons sur les vêtements, mais difficilement les contours d'un costume. On veut extraire les descripteurs symboliques abstraits « costume croisé » ou « costume droit ». Il suffit alors de décrire ces descripteurs en termes non de forme de costume, ce que laisse pourtant entendre leur définition même (croisé ou droit), mais de nombre et de disposition des boutons. Ainsi, on peut reconnaître automatiquement ce descripteur (Cavazza & Green, 1998).

5.3 *Structuration conceptuelle et terminologique*

Outre les objets qui peuvent être obtenus par l'analyse physique de l'image, il est nécessaire de déterminer en quels termes la description documentaire doit s'effectuer. L'enjeu est de disposer d'un ensemble structuré de termes et de concepts pouvant être utilisé dans le langage de description. La difficulté est de savoir quel concept utiliser pour qualifier un contenu donné. Le but est de fournir un environnement au documentaliste qui lui permette, quand il sait ce qu'il veut exprimer, de trouver le concept et un libellé le plus pertinent.

Une hypothèse prometteuse est d'articuler la structuration conceptuelle et la description terminologique autour des ontologies. Une ontologie est la description formalisée des notions élémentaires utilisées pour une modélisation. Elle permet donc de relier d'une part des préoccupations logiques de modélisation et de raisonnement et d'autre des préoccupations terminologiques d'appellation et de désignation. En effet, l'objectif est de pouvoir raisonner sur une description à partir moment où elle est formulée. Il faut donc les mots pour décrire, et les concepts correspondant pour raisonner. Les ontologies constituent un domaine actif de la représentation des connaissances et de l'intelligence artificielle (Gruber, 1993). L'approche évoquée ici est développée dans (Bachimont, 1999).

5.4 *Représentation des documents et de leur méta-données*

Le numérique reconfigure le problème de l'accès à l'information. En effet, deux caractéristiques semblent spécifiques au numérique :

- La massification de l'information : le numérique permet la constitution en une même structure intégrée de très grandes bases documentaires.
- La virtualisation de la masse d'information ; l'ensemble des informations traitées par un système numérique est devenu virtuel : la masse n'est plus appréhendable physiquement et matériellement. La masse virtuelle des informations n'est plus accessible que par le calcul, elle n'est plus appréhendable que par la médiation des procédures calculatoires manipulant l'ensemble des informations.

Dans le cas des supports analogiques, l'appréhension de la globalité de la masse d'information repose sur une appréhension matérielle, physique et perceptive : par exemple, un dossier de documents papiers peut se manipuler, s'étaler sur une table, etc., et se traiter rapidement car, appréhendé globalement par un utilisateur, ce dernier peut s'orienter dans cette masse et définir une stratégie d'accès à l'information. Dans le cas des supports numériques, l'appréhension de la globalité n'est plus directement possible et doit passer par la médiation d'une représentation calculée de la masse. En apportant la possibilité de traiter exhaustivement la masse d'information, le numérique fait perdre l'appréhension de la globalité que permettaient les supports analogiques. Le problème de la représentation de l'information est donc le problème de représenter localement, sur l'écran, la globalité de la masse informationnelle. Il faut restituer, par l'intermédiaire de l'exhaustivité du calcul, l'appréhension de la globalité des supports physiques (Bachimont, 1997).

La nature audiovisuelle des documents traités leur confère le statut d'*objets temporels*. Il se pose alors le problème spécifique à l'audiovisuel de parcours et de navigation dans un objet temporel. Comme les notions de « parcours » et de « navigation » l'implicitent, il s'agit d'un problème de délinéarisation et de spatialisation du document audiovisuel pour le rendre manipulable dans le cadre d'une exploitation donnée. Il s'agit en effet d'une part de décomposer le document en *unités* pertinentes entre lesquels bâtir un parcours ou une navigation et d'autre part donner un point de vue global sur le document et le représenter synoptiquement. La temporalité du document soit rapportée à une vue simultanée des différentes unités qui le composent et permettent une appréhension globale permettant l'orientation et la lecture.

5.5 Outils de lecture et d'appropriation

Dès lors que l'on reconnaît aux documents audiovisuels un véritable contenu culturel, ils justifient une consultation active où le lecteur travaille le contenu consulté. Il s'agit de toutes les activités de reformulation et d'annotation où le lecteur annote, surcharge, commente, griffonne, etc., le document consulté pour mieux se l'approprier. En y apportant sa propre marque, le lecteur devient l'auteur de sa propre lecture dans la mesure où il reformule dans ses propres termes ce qu'il

voit : en quelque sorte, il écrit le document qu'il aurait voulu consulter s'il avait été l'auteur. Ce constat, fait de manière générale pour toute lecture (Stiegler, 1996), a été déjà activement étudié pour les documents textuels (Chahuneau, Lécluse, Stiegler, & Virbel, 1992).

L'enjeu d'une telle recherche est alors de concevoir pour le support audiovisuel les outils permettant l'annotation et l'appropriation par le lecteur. Or on constate qu'en déterminant les fonctionnalités que doivent posséder ces nouveaux outils, on retrouve celles des outils d'indexation. Cela signifie que l'annotation est une indexation personnelle faite par le lecteur lui-même. Mais les outils sont les mêmes. On retrouve alors ce que nous avons évoqué plus haut : la problématique de l'indexation est générique sur toute la chaîne documentaire de l'audiovisuel, de la production à la consultation. Et en particulier les mêmes outils sont requis.

5.6 Ajustement et usages

L'objet de la recherche est double. D'une part, il s'agit de pouvoir décrire quel sens peut posséder un document audiovisuel. Notre discussion plus haut a montré combien ce problème restait difficile et peu instrumenté. Pourtant, il est nécessaire de pouvoir décrire quel sens peut avoir un document audiovisuel pour un utilisateur ou lecteur donné si l'on veut indexer ce document en fonction de ce sens. Autrement dit, pour instrumenter l'indexation, il faut d'abord caractériser le sens que doit posséder cette indexation. D'autre part, il s'agit de définir comment les utilisateurs s'approprient les outils d'indexation et de consultation et parviennent à maîtriser la délinéarisation et la spatialisation des documents audiovisuels.

Cette recherche est complémentaire des précédentes et s'inspire davantage des sciences humaines et sociales.

6 Conclusion

Le numérique modifie profondément les conditions dans lesquelles les documents audiovisuels se constituent et s'utilisent. En particulier, les bibliothèques numériques audiovisuelles prennent une ampleur particulière dans la mesure où non seulement elles deviennent technologiquement faciles à réaliser mais surtout elles constituent une ressource éditoriale indispensable à toutes les étapes de conception, production et utilisation d'un document audiovisuel.

Nous avons proposé ce qui nous semblait constituer les principaux enjeux pour la constitution de telles bibliothèques. Ces enjeux donnent lieu à des recherches dont certaines sont très actives aujourd'hui.

7 Bibliographie

- Auffret, G., Carrive, J., Chevet, O., Dechilly, T., Ronfard, R., & Bachimont, B. (1998). Audiovisual Event Description Interface (AEDI): a model and an exchange format for the description of audiovisual documents content. Paper presented at the ACM Hypertext'99.

- Bachimont, B. (1997). La numérisation du dossier patient. *Annales des Mines*, 49-57.
- Bachimont, B. (1999). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances. In J. Charlet, M. Zacklad, & G. Kassel (Eds.), *Ingénierie des connaissances*: Eyrolles.
- Cavazza, M., & Green, R. (1998). *Multimedia Semantic Features and Image Content Description*. Paper presented at the MMM'98 Conference, Geneva.
- Chahuneau, F., Lécluse, C., Stiegler, B., & Virbel, J. (1992). *Prototyping the Ultimate Tool for Scholarly Qualitative Research on Texts*. Paper presented at the Proceedings of the 8th Annual Conference of the UW Center for the New Oxford English Dictionary and Text Research, Waterloo.
- Chion, M. (1994). *L'audiovision : son et image au cinéma*. (2e édition ed.). Paris: Nathan Université.
- Chion, M. (1998). *Le son*. Paris: Nathan Université.
- Denel, F. (1998, 21 janvier). *La télévision objet d'étude*. Paper presented at the Conférence Ina: la dimension audiovisuelle de l'information et de la communication.
- Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5, 199--220.
- Gugliemo, J., & Chartier, R. (Eds.). (1997). *Une histoire de la lecture*: Seuil.
- Illich, I. (1991). *Du lisible au visible : la naissance du texte. Sur l'art de lire de Hugues de Saint-Victor*: Le Cerf.
- Metz, C. (1968). A propos de l'impression de réalité au Cinéma, *Essais sur La signification au cinéma* (Vol. Tome I, pp. 13--24): Klincksieck.
- Rastier, F. (1987). *Sémantique Interprétative*. Paris: Presses Universitaires de France.
- Rastier, F., Cavazza, M., & Abeillé, A. (1994). *Sémantique pour l'analyse*. Paris: Masson.
- Stiegler, B. (1996). *La technique et le temps ; Tome {II} : la désorientation*: Galilée.