

# **Etat de l'art sur la lecture active de documents audiovisuels**

Bertrand Richard  
Bertrand.RICHARD@liris.cnrs.fr

23 mai 2006

## **Résumé**

L'objectif de ce document est de faire un point sur les différents travaux de recherche déjà effectués dans le domaine de la lecture active de documents audiovisuels. Ainsi nous rappellerons la définition de la lecture active et le rôle des annotations dans le cadre de cette activité. Nous présenterons ce que sont et permettent les hypervidéos et les hypermédias en général, comment elles sont construites et ce qu'elles apportent à la lecture de document. Enfin, nous mettrons en avant le rôle, les caractéristiques et les fonctionnalités des systèmes d'information audiovisuels ainsi qu'un cas particulier de ces systèmes que nous appelons systèmes d'information audiovisuels personnels pour la lecture active.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Lecture active et annotations de documents audiovisuels numériques</b>	<b>5</b>
2.1	La lecture active . . . . .	5
2.1.1	Définition . . . . .	5
2.1.2	Scenarii de lecture active . . . . .	6
2.1.3	Différents types de lecture active . . . . .	8
2.2	Lecture active et annotations de documents . . . . .	8
2.2.1	Lecture active et annotations de documents non numériques . . . . .	9
2.2.2	Lecture active et annotations de documents numériques . . . . .	10
2.3	Lecture active, annotations et documents audiovisuels numériques . . . . .	10
2.3.1	Documents audiovisuels numériques . . . . .	10
2.3.2	Lecture active et annotations de documents audiovisuels numériques . . . . .	11
<b>3</b>	<b>Hypermédias et hypervidéos</b>	<b>13</b>
3.1	Hypermédias . . . . .	13
3.1.1	Définition . . . . .	13
3.1.2	Modèles, définitions et exemples . . . . .	14
3.2	Hypervidéos . . . . .	19
3.2.1	Définition . . . . .	19
3.2.2	Exemples . . . . .	19
3.2.3	Discussion . . . . .	20
3.3	Caractéristiques des systèmes hypervidéos et modèles . . . . .	20
3.3.1	Systèmes pour la visualisation ou la navigation dans les hypervidéos . . . . .	23
3.3.2	Systèmes pour la création et l'édition d'hypervidéos . . . . .	25
3.3.3	Systèmes combinant édition et visualisation . . . . .	28
<b>4</b>	<b>Systèmes d'information audiovisuels</b>	<b>30</b>
4.1	Définition . . . . .	30
4.2	Indexation et description de documents audiovisuels . . . . .	33
4.2.1	Approche de descriptions . . . . .	33
4.2.2	Modèles d'indexation . . . . .	35
4.3	Recherche d'informations et exploration . . . . .	35
4.4	Exploitation . . . . .	36
4.4.1	Visualisation simple . . . . .	36
4.4.2	Réutilisation . . . . .	38

4.5	SIAM personnel pour la lecture active . . . . .	38
<b>5</b>	<b>Conclusion et perspectives</b>	<b>40</b>
<b>A</b>	<b>Outils d'annotation</b>	<b>48</b>
A.1	CLAN . . . . .	48
A.2	ELAN . . . . .	50
A.3	ANVIL . . . . .	52
<b>B</b>	<b>MPEG-7</b>	<b>54</b>

# Chapitre 1

## Introduction

Ce travail de recherche concerne la lecture active de documents audiovisuels et les outils associés à cette pratique. Le but est d'analyser et de formaliser l'activité de lecture active pour mettre en évidence les mécanismes nécessaires à cette activité. Les points principaux de ce travail seront l'étude et la conception de schémas pour structurer les annotations produites en cours de lecture active, la définition et la proposition d'interfaces pour la lecture active, permettant d'utiliser et de visualiser ces schémas et la façon de réutiliser les produits de la lecture active.

Pour cela, nous allons aborder dans ce rapport un état de l'art bibliographique concernant la lecture active de documents audiovisuels numériques et les domaines liés à celui-ci, tels que les hypermédias et hypervidéos, les systèmes d'information audiovisuels et les schémas de description de documents audiovisuels.

Il s'agit donc tout d'abord d'étudier et éclaircir la notion de lecture active de documents audiovisuels. Pour cela, nous rappellerons la définition de la lecture active dans un cadre général, puis nous proposerons quelques scénarii types de lecture active pour illustrer notre propos. Après avoir rapidement rappelé ce que nous considérons comme document audiovisuel numérique, nous verrons en quoi l'application de la lecture active à ce type de documents est spécifique. La pratique de la lecture active de documents audiovisuels s'effectuant par le biais d'annotations, nous préciserons ce que sont des annotations, leur nature et leur utilisation.

Le second axe d'étude qui se dégage du sujet est l'étude des hypervidéos, issues de l'activité de lecture active. Pour cela, nous partirons de la définition générale des hypermédias. Nous rappellerons les principaux modèles d'hypermédias, que nous illustrerons par des exemples, et nous montrerons de quelle manière notre définition des hypervidéos s'adapte à ce cadre. Nous en profiterons pour évoquer différents moyens de visualiser et créer des hypervidéos, et nous présenterons quelques systèmes permettant d'effectuer cette gestion d'hypermédias.

Nous montrerons ensuite comment la lecture active et les annotations s'insèrent dans le cadre particulier de systèmes d'informations audiovisuels et comment l'indexation et la recherche d'information dans de tels systèmes sont liés aux annotations et à leur structure.

Enfin, nous aborderons les différents types de structurations applicables aux annotations audiovisuelles. Nous mettrons en évidence la nécessité de structurer l'ensemble des annotations ainsi que leur contenu, mais aussi le besoin de laisser ces annotations le plus libre possible. Nous parlerons pour cela des schémas de description et de la façon de les construire pour qu'ils soient à la fois souples et rigides, en vue du partage et de

la diffusion des annotations.

## Chapitre 2

# Lecture active et annotations de documents audiovisuels numériques

Dans ce premier chapitre, nous commencerons par rappeler la définition de la lecture active. Nous présenterons ensuite quelques scénarii de pratiques de lecture active, dans des situations de recherche, d'enseignement ou de documentation. Nous évoquerons ensuite les différents types de lecture active existants. Nous présenterons la différence entre la lecture active de documents non numériques et celle de documents numériques, en nous basant principalement sur les annotations utilisées au cours de cette activité, et enfin nous recentrerons sur la lecture active de documents audiovisuels numériques.

### 2.1 La lecture active

#### 2.1.1 Définition

L'activité de lecture en tant que telle peut être perçue de nombreuses manières. Ainsi, Hochon et Évrard [Hochon 94], distinguent la lecture professionnelle, pouvant s'inscrire dans le long terme, une lecture attentive et pouvant engendrer une importante activité de rédaction, d'une lecture dont l'objectif est de goûter le texte en tant que tel. On parle généralement de lecture active dans le cadre d'une lecture professionnelle, lorsque l'on lit un document en assimilant et en réutilisant l'objet ou le résultat de notre lecture. « Lire en inscrivant sa lecture à même le texte lu est typiquement ce que fait un lecteur savant », ou lecteur actif [Stiegler 95]. D'après [Schilit 98] « Active reading is the combination of reading with critical thinking and learning, and is a fundamental part of education and knowledge work ». La lecture active est donc une activité intellectuelle importante axée sur le document, et faisant partie d'un travail d'appropriation de connaissances d'un document. Je propose donc la définition suivante pour la lecture active.

*Définition* : La lecture active est une activité composée d'une lecture critique et d'une production par rapport à un document. Un lecteur se retrouve en situation de lecture active dès lors qu'il effectue une lecture approfondie d'un document, pour s'en

appropriier le contenu, pour y rechercher des informations précises, ou même plus simplement pour en permettre une relecture simplifiée.

Ainsi la lecture active est perçue comme une activité autour d'un document, aboutissant à une production personnelle mesurable, pouvant prendre différentes formes, telles qu'un nouveau document, un document annoté, ou simplement la mémorisation de sa lecture.

### **2.1.2 Scénarii de lecture active**

Afin de préciser un peu cette définition et de montrer quelques cas concrets, je propose les quelques scénarii suivants, dans lequel le lecteur effectue une lecture active :

- Un chercheur étudiant les comportements et interactions dans une vidéo numérique
- Un étudiant réalisant une explication de texte
- Un professeur préparant un cours sur les techniques filmographiques
- Une documentaliste classant un nouveau document
- Un doctorant se documentant sur la lecture active

#### **Etude de comportements dans une vidéo numérique**

Prenons un premier exemple : un chercheur étudiant les comportements et interactions humaines par l'analyse de vidéos. Le chercheur commence par lire rapidement son document audiovisuel afin de repérer les parties de la vidéo intéressantes à étudier. Après avoir déterminé quelques passages pour son étude, le chercheur les lit en boucle, essayant de transcrire les paroles des protagonistes. Le son n'étant pas très clair, il est difficile de distinguer les paroles exactes, et plusieurs retours en arrière sont nécessaires pour transcrire correctement les échanges verbaux. Une fois ceux-ci listés, le chercheur essaye de repérer les différents gestes, et de les relier temporellement aux paroles, afin de pouvoir étudier les interactions entre les participants ainsi que les interactions entre le verbal et le gestuel. A la fin de son étude, le document produit est un ensemble de transcriptions textuelles des gestes et paroles d'un passage, permettant de mettre en évidence un mécanisme d'interactions dans une situation donnée.

Cet ensemble d'actions, de lectures et relectures de la vidéo, de transcriptions et de production, au final, d'un ensemble d'annotations en temps que nouveau document, nous permet d'après notre définition de parler de lecture active. L'objectif de cette lecture active était de produire une analyse comportementale basée sur des annotations séparées du support vidéo. Le lien entre les annotations et le flux temporel est constitué de marqueurs temporels indiquant les bornes des prises de paroles. Ce type de travail est effectué dans le cadre de l'équipe de recherche Incorporation du laboratoire ICAR du CNRS (UMR 5191) par l'utilisation d'outils de transcription tels qu'Elan [the M.I.P. 06] ou Anvil [Kipp 01] (voir l'annexe A pour plus de détails).

#### **Explication de texte d'un étudiant**

Dans le cas d'un étudiant devant rédiger un commentaire de texte, celui-ci commence par lire et analyser la structure du document. Il surligne et met en évidence les idées importantes, encadrant des termes, ajoutant des annotations à même le document. Il peut redéfinir certains termes dans la marge, faire des schémas pour représenter la structure d'argumentation, et par cette recherche d'information dans le document, extrait l'essentiel de son contenu de manière synthétique. Le résultat de son étude est

représenté par son commentaire, document présentant et critiquant les idées mises en évidence par ses annotations.

La lecture active est ici menée à l'aide des annotations du document textuel papier de différents types. Les annotations sont des traits, des schémas, des surlignages, des compléments d'information, et sont directement intégrées dans le support du document. Cette activité s'achève par la production d'un document en rapport avec le document d'origine, possédant des références vers celui-ci, et présentant le résultat de l'activité de lecture active.

### **Préparation de cours sur les techniques filmiques**

Un professeur décide d'enseigner les techniques filmographiques par l'analyse d'un document audiovisuel. Pour rendre cela plus efficace, il décide de présenter à ses étudiants la vidéo sous différents « points de vues ». Lors d'une première lecture de son document, il annote les différents plans de la vidéo. Il crée ensuite des annotations pour commenter et catégoriser les transitions entre les plans, et procède de même pour les différentes scènes <sup>1</sup>. Pour chaque plan, il décrit sous forme d'annotation la manière dont celui-ci a été filmé, ses caractéristiques techniques. Enfin, le professeur met en place des commentaires permettant d'expliquer l'impact des techniques filmiques détectées sur un spectateur (au cours d'une activité de lecture simple). Il crée ensuite différentes vues hypertextes pour présenter son analyse aux étudiants, avec des liens directs vers le flux vidéo pour illustrer directement son cours par le film.

La lecture active est ici présente sous deux aspects distincts. Tout d'abord, le professeur analyse le document, met en évidence certaines informations importantes pour son cours et ajoute ses commentaires au document. La lecture active s'appuie ici encore sur les annotations, qui permettent l'élaboration d'un nouveau document, une hypervidéo qui sera présentée pour le cours. D'autre part, les étudiants qui vont visualiser cette hypervidéo, pratiqueront eux aussi une lecture active du document. En effet, ils visualiseront des parties de la vidéo pendant leur lecture du cours, et agiront sur le document afin de s'en approprier le contenu.

### **Classification de documents**

Une documentaliste recevant un nouveau document à classer dans un système d'information documentaire va devoir l'indexer. Pour cela, elle effectue une lecture rapide de l'oeuvre ou du résumé et l'annote, la décrit, suivant certains critères, permettant de déterminer comment il sera classé.

En lisant son document pour en extraire des points particuliers ou pour déterminer à quelle catégorie il appartient, la documentaliste effectue ici une lecture active. Elle résulte en une classification de l'oeuvre suivant des critères déterminés.

### **Documentation sur un sujet**

Pour réaliser un état de l'art sur son sujet de thèse, un doctorant doit lire et s'approprier un grand nombre d'articles et de documents divers. Celui-ci lit donc différents documents et prend des notes sur ceux-ci pour pouvoir effectuer des relectures plus rapides. Il peut aussi créer des fiches récapitulatives pour chaque article lu, résumant les points importants mis en évidence dans le texte par ses annotations. Celles-ci peuvent

---

<sup>1</sup>Un plan est une prise de vue sans interruption. Une scène est composée de différents plans, séparés par des effets de transition.

être publiées, par exemple sur un site internet, et être reliées aux documents d'origine par des liens vers ceux-ci.

L'activité de lecture active est ici évidente. Le doctorant cherche et extrait des informations dans les articles, les met en évidence par annotations, et produit en plus des documents annotés, d'autres documents en relation avec ceux-ci. La lecture active s'effectue à l'aide d'annotations, permettant de créer des documents annotés et de nouveaux documents.

### 2.1.3 Différents types de lecture active

Les scénarii précédents mettent en évidence plusieurs catégories de lecture active. En effet, plusieurs critères peuvent être dégagés des scénarii précédents :

- L'orientation de la lecture :
  - Recherche d'information dans le document
  - Appropriation du document
  - Indexation du document
  - Production en rapport avec le document
- La production en fin d'activité :
  - Le document d'origine annoté
  - Un nouveau document constitué par les annotations
  - Un nouveau document en relation avec le document d'origine
  - Une mémorisation du document
- Le moyen de la lecture active :
  - Des annotations
  - Des relectures approfondies

Ce que nous appelons « orientation » de la lecture correspond au but vers lequel le lecteur est tourné au cours de son activité. C'est la raison pour laquelle le lecteur étudie son document.

Ainsi, la recherche d'informations précises sur un sujet est un type de lecture active, de même que la lecture à des fins d'appropriation du document, ou encore la lecture dans le but d'indexer le document. Chacune de ces activités vise à rechercher, mettre en évidence, trier, ou aider à la mémorisation de l'information, en utilisant l'écriture et la lecture pour y parvenir. Elle vise à produire de la connaissance sur un document, sous forme d'un nouveau document, ou du document annoté, ou d'autres traces dans l'espace de travail.

## 2.2 Lecture active et annotations de documents

Comme nous l'avons vu dans les scénarii précédents, la lecture active se pratique essentiellement par l'annotation de documents.

*Définition* : Une annotation est une inscription sur un document ou liée à celui-ci, décrivant, mettant en évidence ou ajoutant de l'information à une partie ou à la totalité d'un document. Les annotations dépendent donc principalement du support du document (numérique ou pas), de son type (texte, son, vidéo) et de leur ancrage dans le document.

L'annotation a donc une forme de stockage, dans un format qui lui est propre, et une forme de restitution en rapport avec la nature du document. De plus, des ancres permettent de la situer dans le document, de manière spatiale et temporelle si cela

est nécessaire. Ces ancrs constituent des passages annotés du document, délimités spatialement et temporellement.

[Waller 03] définit une annotation numérique comme « A comment upon a digitally accessed resource as a whole, or the contents of a resource, and which itself can be digitally accessed as well as stored. » Autrement dit, *une annotation numérique est une information faisant référence à tout ou partie d'un document, accessible et stockable.*

Il est nécessaire de distinguer les ancrs qui lient le document à ses annotations des liens qui apparaissent pour la navigation dans les hyperdocuments ainsi formés, appelés hyperliens.

### 2.2.1 Lecture active et annotations de documents non numériques

Commençons par les documents non numériques. Lors de la lecture active d'un document textuel papier, le lecteur a recours à différentes formes d'annotations. Catherine C. Marshall présente dans [Marshall 97] une catégorisation de ces annotations sur du texte, qu'elle a pu mettre en évidence lors d'une étude universitaire. Il en résulte les différentes formes d'annotations suivantes :

Sous ou sur lignage d'éléments de structure
Sous ou sur lignage étendu
Surlignage court
Notes dans le texte
Notes dans la marge
Notes entre les lignes
Mots entourés
Symboles (astérisques)
Dessins dans la marge

TAB. 2.1 – Différents types d'annotations de texte

Les annotations peuvent être composées de textes, dessins, lignes, signes et symboles. Ces annotations peuvent être localisées sur le document, à l'intérieur même de son contenu, ou alors sur un support dissocié. Les annotations localisées sur le document permettent en général une explicitation de son contenu, ou une mise en valeur de celui-ci, tandis que les annotations extérieures permettent d'avoir une vue résumée du document, avec des liens vers celui-ci. Le pointage offert par ces liens est relatif à la nature du document. Souvent, ce sont des numéros de lignes et de paragraphes faisant référence au texte d'origine. Pour des documents non textuels, le choix des annotations est plus restreint. En effet, le lecteur n'a pas d'accès direct en écriture dans le document, et ne peut donc pas y inscrire ses annotations. Celles-ci seront donc forcément inscrites dans son espace de travail, extérieur au document. D'autre part, le lien entre l'annotation et le contenu annoté du document est lui aussi difficile à réaliser. Le repère à la temporalité est mis en place par positionnement relatif au début du document, mais le positionnement spatial dans l'image, dans le cas d'une vidéo, se fait par description textuelle ou schéma. Ainsi pour annoter un acteur jouant dans un film, il va être nécessaire de définir les bornes temporelles de début et de fin de l'annotation, mais aussi d'expliquer à quoi se réfère l'annotation dans le passage. La lecture active de ces documents est rendue plus accessible et plus pertinente dans le cas de documents numériques.

### 2.2.2 Lecture active et annotations de documents numériques

Un gros apport du numérique pour les documents audiovisuels est le fait de pouvoir annoter ces documents par des annotations de forme et de nature très variées, ainsi que la possibilité de stocker ces annotations de manière totalement indépendante du document objet de l'étude, sans pour autant perdre le lien entre eux. L'annotation de documents textuels peut toujours être composée de texte, de surlignage ou soulignage [Obendorf 03] (avec des systèmes comme Xlibris [Schilit 98]), mais peut aussi être composée de documents hypertextes (avec le système HATS par exemple [Kim 04]), ou encore des vidéos, des sons, ou tout autre type de document [Uren 05]. Ces dernières annotations sont situées à l'extérieur du document, mais lui sont reliées par des références à des ancres dans celui-ci. Les annotations numériques peuvent être de nature différente du document annoté, différenciant ainsi la forme de stockage de l'annotation et sa représentation. Par exemple, si l'on considère une vidéo sous-titrée ; lors de la visualisation de celle-ci, on pourra considérer les sous-titres comme faisant partie de la vidéo. Pourtant, en termes de stockage, les sous-titres seront disponibles sous forme de document texte. Le stockage des annotations peut aussi se faire hors du document annoté, comme avec Madcow [Bottoni 04], système pour lequel les annotations sont stockées sur des serveurs spécifiques. Il est donc nécessaire de bien distinguer forme de stockage et forme de visualisation dans le cas d'annotations numériques. Les annotations numériques peuvent être classées suivant différents critères :

- leur nature en tant que forme de stockage (texte, vidéo, image)
- leur lien avec le document annoté (ancrage spatial, temporel)
- leur forme de restitution (inclus ou non dans le document)
- leur complexité de contenu (simple lien, fragment de document ou document lui-même annoté)
- leur degré d'individualité (si elles ont un sens sans le document auquel elles font référence)

## 2.3 Lecture active, annotations et documents audiovisuels numériques

### 2.3.1 Documents audiovisuels numériques

Concernant les documents audiovisuels numériques, il est nécessaire tout d'abord de rappeler leur définition. Edmondson [Edmondson 04] propose une définition d'un document audiovisuel comme étant un ensemble de trois composantes : « Audiovisual documents are works comprising reproducible images and/or sounds embodied in a carrier whose

- recording, transmission, perception and comprehension usually requires a technological device
- visual and/or sonic content has linear duration
- purpose is the communication of that content, rather than use of the technology for other purposes »

La définition de B. Bachimont [Bachimont 99] du document audiovisuel, et celle que nous adopterons pour la suite, présente le document audiovisuel comme : *Définition* : 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.

Nous allons donc étudier la façon dont se traduit la lecture active dans le cas des documents audiovisuels numériques.

### **2.3.2 Lecture active et annotations de documents audiovisuels numériques**

La différence principale entre les documents textuels et audiovisuels réside dans le rapport au temps de l'activité de lecture. En effet, si on peut avoir une vision globale du document textuel au cours de la lecture, le document audiovisuel n'est pas abordable dans sa totalité à un instant donné. Il est nécessaire de respecter sa vitesse de lecture pour qu'il reste compréhensible par le lecteur. Les informations contenues dans les documents audiovisuels sont localisées spatialement et temporellement, et leur mise en évidence est donc plus compliquée. Les parties mises en évidence par la lecture active ne sont plus, dans ce cadre, des parties de texte ou des structures dans le texte, mais des éléments audiovisuels, tels que des parties d'images, des sons, des éléments de structure du document audiovisuel, des instants, *etc.*

L'annotation des documents audiovisuels peut se faire directement dans la vidéo, ou autour de celle-ci, dans ce qu'on pourrait appeler l'espace de visualisation. Cet espace correspond en fait à l'environnement autour de la fenêtre de visualisation de la vidéo. Comme pour les documents textuels papiers évoqués précédemment, on retrouve le fait qu'une annotation directement dans le document permet une mise en évidence efficace, tandis qu'une annotation « dans la marge » présente plutôt un complément d'informations liées au contenu. Dans le cas des documents audiovisuels, nous avons vu que les annotations doivent pouvoir être situées spatialement et temporellement par rapport au document. Dans le cas du numérique, le problème de la spatialisation est résolu puisque le lecteur peut accéder au document même pour inscrire ses annotations. Pour la temporalité, le lecteur peut choisir de faire apparaître ses annotations au moment auquel elles sont liées dans le document. Le lecteur retrouve donc pour les documents audiovisuels numériques un contrôle aussi précis que dans le cas des documents papiers textuels. Une étude sur les documents multimédias et les annotations [Geurts 05] met en avant différentes caractéristiques nécessaires pour une bonne utilisation des annotations par le lecteur. Ainsi, elle insiste entre autre sur la nécessité de pouvoir annoter de façon simple, d'utiliser un vocabulaire clairement défini et de structurer le contenu des annotations pour pouvoir les réutiliser.

Suite à cette présentation de la lecture active de documents audiovisuels numériques, nous pouvons émettre les quelques remarques suivantes :

Tout d'abord, une des raisons pour lesquelles un lecteur peut pratiquer la lecture active est de permettre une relecture fonction de ses annotations. Ainsi dans le cas d'un document textuel papier, le lecteur relit son document en tenant compte de ses notes et références, obtenant ainsi une lecture enrichie de celui-ci. Cependant, dans le cas des documents audiovisuels numériques, le document peut être directement modifié par les annotations, offrant une lecture bien différente de celle d'origine. Prenons l'exemple d'un film dont un lecteur fait l'étude. Celui-ci a décidé lors d'une première lecture de s'intéresser aux apparitions des protagonistes à l'écran, et a donc annoté les passages du document concernés. S'il s'intéresse par la suite aux interactions entre les acteurs, il peut bénéficier d'un gain de temps considérable en ne visualisant que les passages dans lesquels plusieurs acteurs sont présents.

D'autre part, on peut se pencher sur le cas d'un lecteur qui lit et relit en boucle un passage d'une vidéo à des fins d'appropriation visuelle. Ce type d'activité ne produit

aucune inscription physique, cependant le lecteur effectue cette lecture afin de s'approprier le contenu du document, et doit donc être considéré comme un lecteur actif. Il a donc été nécessaire de prendre cette situation en compte dans la définition précédente de lecture active, en considérant l'activité de mémorisation comme une activité évoluée d'écriture sur un élément de l'environnement de travail du lecteur, sa mémoire.

Enfin, on peut s'interroger sur le statut de ce document issu de la lecture active, si c'est le même document enrichi par ses annotations et présenté différemment, ou si justement par cette manipulation du document lors de sa restitution, nous avons affaire à un nouveau document.

La lecture active de documents est donc intimement liée à l'activité d'annotation. Celles-ci permettent de mettre en valeur les informations jugées importantes dans le document dans le cadre de l'étude menée, et sont donc un moyen d'écrire dans, ou autour, du document. L'activité de lecture active est caractérisée par plusieurs critères, tels que son orientation, le résultat produit et le moyen pour y parvenir. Il semble intéressant de penser que la lecture active ne s'accompagne pas seulement d'une production physique, mais peut aussi avoir pour résultat un approfondissement des connaissances sans support autre que la mémoire du lecteur. D'autre part, la possibilité de diriger la (re)lecture par les annotations amène à considérer la présentation des documents annotés produits comme des hyperdocuments, et dans le cas des documents audiovisuels, des hypervidéos.

Pour conclure, nous avons rappelé dans ce chapitre les définitions de lecture active de document et d'annotations. Nous avons aussi mis en avant les différences lorsque cette activité de lecture active est effectuée sur des documents numériques ou non, et nous avons précisé ses particularités dans le cas des documents numériques audiovisuels. Nous allons maintenant aborder le domaine des hypermédias et hypervidéos, ceux-ci pouvant être des produits issus de la lecture active.

## Chapitre 3

# Hypermédias et hypervidéos

Nous avons vu que l'activité de lecture active équivaut à une opération de production au cours de la lecture. Il peut s'agir d'une production strictement privée qui restera uniquement dans l'espace de travail, comme des annotations prises sur un bloc-note, mais cette production peut aussi être un document hypermédia, document d'origine enrichi par ses annotations permettant une navigation dans celui-ci. Nous présenterons dans ce chapitre la notion d'hypermédia ainsi que les deux principaux modèles dans ce domaine et quelques exemples pour illustrer nos propos. Nous aborderons ensuite la notion d'hypervidéo, et nous présenterons quelques systèmes de gestion, création et visualisation d'hypervidéo et leurs modèles.

### 3.1 Hypermédias

#### 3.1.1 Définition

La notion d'hypermédia est basée sur deux principes : l'hypertexte et le multimédia. Le premier définit des possibilités de navigation intra ou extra textuelle via des liens. Son application dans le domaine du multimédia, domaine du son, des images, de la vidéo, fait apparaître la notion d'hypermédia. Un hypermédia est donc un hyperdocument, document dont la structure permet de l'aborder de manière non-linéaire, de nature multimédia.

*Définition* : Un document hypermédia est un ensemble de ressources numériques intégrées sur un même support, constituées de données de nature différente (texte, son, image et fichier) avec des liaisons multiples qui donnent, à la consultation de ces ressources, une dimension réticulaire.<sup>1</sup>

Les hypermédias sont donc des documents représentables sous forme de graphes, dans lesquels les noeuds sont des fragments documentaires, et les arcs sont des hyperliens entre ces différents fragments. Les documents hypertextes sont des documents hypermédias particuliers, dans lesquels les noeuds sont principalement du texte et des images.

Dans la partie suivante, nous présenterons quelques exemples de modèles hypermédias. La documentation de ceux-ci étant souvent insuffisante, une reconstruction *a posteriori* de ceux-ci à partir des exemples d'utilisation a été nécessaire.

---

<sup>1</sup>Définition proposée par l'association pour le développement du Multimédia Informatisé Pédagogique francophone (MIP+) [MIP+ 06]

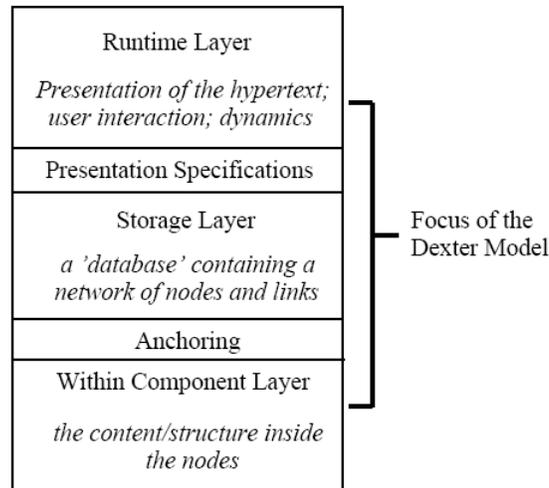


FIG. 3.1 – Les couches du modèle d’hypertexte Dexter

### 3.1.2 Modèles, définitions et exemples

Un modèle de document hypermédia permet de donner un cadre pratique à la définition de ce type de document. C’est un modèle de description de l’organisation des données contenues dans le document [Whitehead. 02]. Le but de ces modèles est de définir un cadre général applicable à tout type de document hypermédia. Il est donc nécessaire d’y définir ces différents éléments :

- La façon de stocker les éléments de l’hypermédia
- La façon de lier ces éléments de stockage
- La façon de visualiser l’hypermédia
- La façon de naviguer dans l’hypermédia
- La nature des éléments de l’hypermédia

#### Modèles

**Modèle Dexter** Le modèle Dexter [Halasz 94, Grnbaek 96] a pour but de rassembler les caractéristiques communes de nombreux systèmes Hypertextes, et de permettre d’établir des comparaisons, des échanges et l’interopérabilité entre ces systèmes. Il est organisé en trois couches (cf. figure 3.1) :

- « the storage layer », la couche de stockage, description du réseau de liens et de noeuds de l’hypertexte
- « the runtime layer », la couche d’exécution, description des interactions entre le lecteur et le document
- « the within-component layer », la couche décrivant le contenu et la structure interne des noeuds.

Ce modèle met plus l’accent sur la première couche et ses relations avec les deux autres. Il est important de noter que cette couche explicite la façon dont les noeuds sont reliés ensemble pour former le document hypertexte, tandis que la troisième couche traite de la structure interne des noeuds, et que celle-ci est donc dissociée de la structure du document hypertexte. Dans ce modèle, un hypertexte est considéré comme un

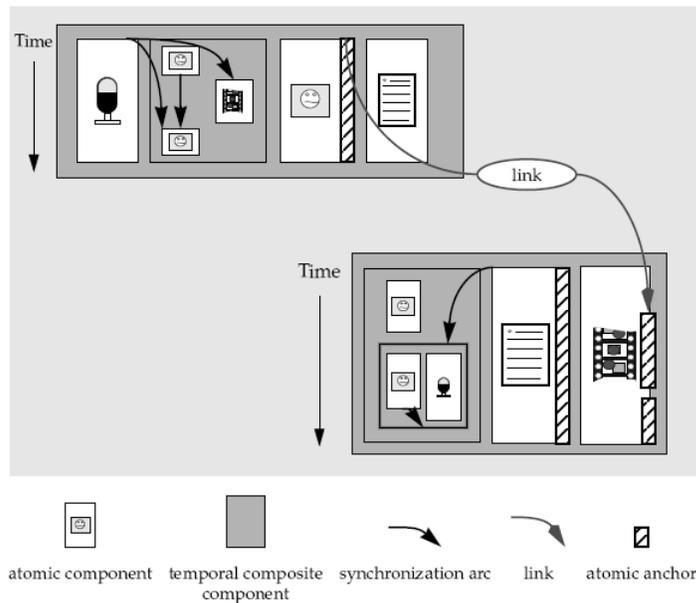


FIG. 3.2 – Une représentation du modèle AHM : deux éléments complexes unis par un lien disponible à sa source pendant toute la durée de l'élément, et pointant vers une ancre temporellement bornée du second.

ensemble de composants identifiés de manière unique. Parmi ces composants, les liens sont un cas particulier. Ceux-ci possèdent une caractéristique direction, indiquant s'ils sont bidirectionnels ou non. Ils ont en plus deux caractéristiques supplémentaires, une source et une destination, composants de type « ancre » de l'hypertexte. Enfin, les composants de base du document sont des « atomes ». Ce modèle d'hypertexte permet donc de spécifier, dans sa première couche, la façon dont les éléments de l'hypermédia sont structurés entre eux pour former le document. Cette couche décrit leur forme de stockage, les liens entre eux, la façon dont ces liens leur sont attribués, *etc.* La deuxième couche du modèle présente la façon dont le document sera montré au lecteur. En effet, on retrouve dans cette couche la description des interactions entre le lecteur et le document, la façon dont il pourra agir sur les liens, sur la structure et le contenu du document. Dans la troisième couche se trouvent les contenus des éléments du document et leur structure, ces éléments pouvant être des documents, des liens, des ancres, *etc.*

**Modèle Amsterdam (AHM)** Le modèle d'Hypermédia Amsterdam [Hardman 94] constitue en quelque sorte une extension du modèle précédent. En effet, il ajoute à celui-ci la prise en compte du contexte dans le suivi des liens, ainsi que la gestion de la temporalité, nécessaire pour les documents audiovisuels.

Ce modèle peut se décomposer en cinq types d'éléments distincts :

- des « media items », documents existants ou générés à la volée, dont la représentation spatiale et temporelle est connue.
- des « channels », zones de visualisation définies dans le document. Cela peut être une zone de l'écran, ou le flux sonore, *etc.*
- des « atomic components », formant les briques de base de l'hypermédia. Ceux-

ci contiennent des informations sur l'accès à un « media item », ainsi que des balises d'entrées dans cet objet. Ils sont basés sur les composants noeuds du modèle Dexter.

- des « composite component », regroupant des ensembles de composants atomiques, organisés spatialement et temporellement, avec des arcs de synchronisation interne des composants atomics.
- des « link components », liens entre les différents composants, définissant les contextes de départ et d'arrivée du lien ainsi que sa bidirectionnalité éventuelle.

Ces cinq composants se retrouvent à des niveaux différents du modèle Dexter. Ainsi, les objets media font partie de la couche de stockage, les composants atomiques et complexes de la structuration interne de l'hypermédia. Quant à la présentation et les interactions avec le lecteur, celles-ci sont définies dans chaque composant de l'hypermédia, et sont donc propres à chaque objet. De plus, l'objet « channel » permet de définir la façon dont la visualisation sera effectuée.

Suivant ce modèle, un hypermédia est composé d'éléments complexes reliés entre eux par des liens, eux-mêmes organisant spatialement et temporellement des éléments atomiques, donnant accès à des fragments de documents, visualisés dans des canaux. La figure 3.2 montre un exemple du modèle AHM. Elle représente un document hypermédia, composé de deux éléments complexes unis par un lien. Les éléments complexes sont des ensembles d'éléments simples organisés temporellement, représentant du son, de l'image et du texte et synchronisés entre eux. Le lien mène d'une balise présente tout au long du premier élément complexe vers une balise temporellement limitée du second.

## Exemples

Xanadu [Nelson 99] est le projet d'hypertexte de T. H. Nelson, inventeur des termes « hypertexte » et « hypermédia », démarré dans les années 60. Imaginé au départ uniquement pour du texte, le modèle s'est ensuite vu étendre pour prendre en compte un autre type de média, les images. L'idée originale consiste à pouvoir mettre en relation au cours de la lecture plusieurs textes, par l'intermédiaire de liens entre ceux-ci. La modélisation sous-jacente de Xanadu se résume donc à un objet document, ou hyperdocument, organisant un ensemble de fragments de documents et de liens, ces derniers permettant d'exprimer les relations entre les fragments (voir figure 3.3). Ce modèle se situe au niveau de la couche de stockage du modèle Dexter. Il met en place une façon de structurer et lier les éléments de base d'un hypertexte. Il évoque aussi la façon dont cette couche est liée à la couche de description du contenu des composants. Il ne précise aucunement la façon dont le document sera visualisé, ni les interactions avec le lecteur. Il s'agit donc d'un modèle focalisé sur l'organisation interne du document hypermédia.

A l'origine, le web actuel était un projet interne au CERN [Berners-Lee 89], mené par Tim Berners-Lee, visant à faciliter l'accès aux différentes ressources du centre de recherche. Le projet en lui-même est basé sur un modèle d'hypermédia assez simple, mettant en jeu :

- Des liens unidirectionnels, sous forme de balises sources et destination.
- L'inclusion de documents images
- L'inclusion de documents textes

A. Benel dans son travail de thèse [Bénel 03] présente un diagramme UML de rétro-conception du World Wide Web simplifié assez pertinent (cf. figure 3.4).



Ce modèle a ensuite évolué pour intégrer désormais d'autres types de médias tels que les documents audiovisuels, directement en tant que fragments de l'hyperdocument, à l'instar des documents textes et images. Il est de plus possible de n'inclure que des portions temporelles de documents audiovisuels, par l'intermédiaire de liens temporellement situés.

Le modèle du World Wide Web est encore en constante évolution, cherchant à intégrer et présenter des fragments de nature totalement différente au sein d'un même document hypermédia. Ce modèle s'insère aussi dans les trois couches du modèle Dexter. Tout ce qui concerne la structure interne des éléments de cet hypermédia est dissocié de la structure externe, composée de pages web et de liens. On peut rapprocher les « frames » du web des « channel » du modèle d'hypermédia Amsterdam. En effet, ceux-ci définissent un cadre de visualisation d'un contenu.

Enfin, le système Opales [Nanard 01] de l'Institut National de l'Audiovisuel (INA) est un système d'information basé sur un modèle hypermédia. En effet, chaque élément du système est considéré comme une unité informationnelle. Cela va du document indexé à sa description en passant par les liens entre eux, leur propriétaire, *etc.* Chaque unité informationnelle est décomposée en deux parties : son contenu et son descripteur. Le premier représente le contenu même de l'unité, données structurées propres à l'élément. Le second est un ensemble composé d'un identifiant et de quatre liens particuliers typés, associant l'unité à son propriétaire, son rôle, son type et son associé. Les liens entre les unités, étant eux-mêmes décrits, sont des éléments à part entière du système, peuvent être interrogés et suivis. On retrouve donc un modèle hypermédia comme base de gestion de documents dans un système d'information.

Au delà de ces modèles, un modèle de construction d'hypermédia [Romero 03] a été conçu, intégrant des représentations d'éléments réels dans un environnement 3D. C'est donc une conception élargie de l'hypermédia, faisant intervenir, à travers des représentations, des éléments de la réalité dans le document. Le principe de l'hypermédia permet d'intégrer ces éléments réels dans un espace virtuel constitué d'entités ayant des liens entre elles. L'utilisateur d'un système basé sur un tel modèle se retrouve plongé dans un espace virtuel en trois dimensions, dans lequel des opportunités d'interactions suivant des liens hypermédiés sont proposés au fil de l'exploration de cet espace. Le modèle sous-jacent est composé d'objets et de liens entre eux, les objets pouvant être des objets réels ou virtuels, ces derniers étant des documents multimédias. De tels systèmes peuvent être utilisés pour des visites guidées de musées, des jeux à réalité augmentée, de l'assistance médicale. Un jeu d'enquête policière a ainsi été développé, dans lequel le personnage doit relever des indices dans un espace hypermédia. Le modèle proposé par les auteurs de cet article permet donc de créer des systèmes hypermédiés pour raconter ou préparer une histoire, comme dans hypercafé [Sawhney 96], mais dans un environnement virtuel. Le « lecteur » interagit donc suivant les liens hypermédiés, dans les limites de ce qui a été préalablement défini par les auteurs. Ce modèle peut lui aussi être considéré comme une extension des modèles Dexter et Amsterdam, ajoutant à celui-ci la dimension « réelle », par l'intermédiaire d'une représentation à l'interface entre les objets réels et le monde virtuel. La différence se situe alors dans la nature des éléments composant l'hyperdocument, ainsi que dans les interactions avec le lecteur.

## 3.2 Hypervidéos

### 3.2.1 Définition

Une hypervidéo est un cas particulier d'hypermédia, dont le média principal est celui de l'audiovisuel. Il existe là encore, de nombreuses définitions de ce terme. Celle que je présente ci-après a été définie dans [Aubert 05b] :

*Définition* : Une hypervidéo peut être définie comme une vue d'un document audiovisuel annoté, permettant d'une part, d'enrichir le document audiovisuel par ses annotations, et d'autre part de contrôler directement le flux audiovisuel à partir de celles-ci.

Un *document audiovisuel annoté* est dans ce cas, un document audiovisuel auquel on a ajouté une structure d'annotations liée spatio-temporellement au flux de ce document.

Une *vue* d'un tel document est une manière de visualiser celui-ci. Elle définit la proportion d'informations du document d'origine et de ces annotations qui seront présentées au lecteur. En général, une vue offre au lecteur un mélange du flux d'origine avec ses annotations, mais elle peut aussi simplement proposer une visualisation réorganisée du document suivant ses annotations. Une telle définition de l'hypervidéo s'inscrit parfaitement dans le cadre du modèle AHM. En effet, on trouve d'une part la couche de stockage dans l'organisation du document et de ses annotations à travers des schémas organisant celles-ci et définissant les relations entre elles. D'autre part, le contenu des annotations et du document audiovisuel est aussi structuré de manière indépendante. Le document audiovisuel annoté est alors l'ensemble vidéo + annotations uni par des liens ancrés dans ces documents. Enfin, la couche de présentation est assurée par les différentes vues du document, qui s'appuient sur ces liens et annotations pour les présenter de manière compréhensible au lecteur. Les vues présentent un point de vue particulier du document audiovisuel [Costa 02].

Cette définition met en avant le fait qu'une hypervidéo est, du point de vue stockage, un ensemble composé de différents médias (vidéo, hypertexte, annotations, etc.). Une hypervidéo peut être représentée par un graphe dont certains des noeuds sont des fragments de documents audiovisuels. D'autre part, une hypervidéo offre une appréhension différente de ces médias, en fournissant un moyen de s'affranchir de la temporalité linéaire de ceux-ci [Aubert 04c, Aubert 03]. *Une hypervidéo est donc un document issu d'une activité de lecture active, permettant une visualisation enrichie et guidée par les annotations de la vidéo.*

### 3.2.2 Exemples

Reprenons l'exemple d'une étude de film. Le lecteur a d'abord visualisé son film en annotant tous les passages avec les noms des acteurs présents dans la scène. A partir de ces annotations et du document vidéo, une hypervidéo pourra être créée, affichant par exemple l'ensemble des acteurs par ordre d'apparition sous forme de page HTML, et permettant l'accès direct au flux à chacun de leurs apparitions dans la vidéo. L'ensemble formé par ce document HTML, la vidéo du film, les annotations et leur lien au flux forme une hypervidéo. De même, considérons un sommaire construit sur les différentes scènes d'un film et donnant accès à la temporalité du flux. L'ensemble composé

du sommaire, de la vidéo qu'il concerne et des hyperliens entre eux formera une hypervidéo. D'autres exemples d'hypervidéos sont présentés dans la suite du document.

### **3.2.3 Discussion**

Il semble intéressant de préciser qu'une hypervidéo est un hypermédia basé sur un document audiovisuel. Il existe dans ces documents, ce que l'on peut appeler un « document central », qui est celui qui a permis d'élaborer l'hypervidéo. Ainsi, il est à noter par exemple, qu'une page HTML résumant le contenu d'une vidéo et fournissant ne serait-ce qu'un hyperlien permettant de jouer celle-ci, est une hypervidéo.

On peut souligner aussi que la nature des hyperliens est plus complexe qu'il n'y paraît au premier abord. Il est nécessaire pour les hyperliens de navigation dans des documents audiovisuels de définir la disponibilité spatio-temporelle de ceux-ci. Dans [Chambel 02], Chambel établit une classification des liens dans les hypervidéos :

- Lien inconditionnel, toujours présent quelque soit la position temporelle dans la vidéo.
- Lien spatial, localisé spatialement sur la vidéo, mais indépendant de la position temporelle.
- Lien temporel, disponible pendant un certain intervalle de temps
- Lien spatio-temporel, valide pendant un certain temps, et disponible sur une région donnée de l'image.

## **3.3 Caractéristiques des systèmes hypervidéos et modèles**

Nous considérerons comme systèmes d'hypervidéos, tout système permettant de construire, modifier, visualiser des hypervidéos à des degrés variables. Ainsi, un outil permettant de visualiser un document audiovisuel avec des annotations fournissant un contrôle sur le flux sera un système d'hypervidéos. De même, un système implémentant un langage permettant d'accéder au flux d'une vidéo depuis un lien sur une page web sera considéré comme un système d'hypervidéo. Bien souvent, de tels systèmes se focalisent sur un point précis de l'exploitation ou de l'édition des hypervidéos. Nous allons d'abord présenter deux formats particuliers utilisés pour la présentation et la visualisation d'hypervidéos puis nous aborderons quelques uns de ces systèmes et les repositionnerons par rapport au cadre théorique de la définition.

### **Techniques de visualisation et présentation d'hypervidéos**

On retrouve dans cette catégorie les formats Macromédia Flash et SMIL (Synchronized Multimedia Integration Language) [Bulterman. 03] pour générer des documents hypermédiés à partir de descriptions au format MPEG-7 [Nack 99a, Nack 99b]. SMIL est un langage permettant principalement de lier le flux d'une vidéo à une description web et de les synchroniser. MPEG-7 est un format de description de vidéos, permettant d'ajouter des annotations suivant un schéma précis sur tout ou partie d'un document vidéo.

Flash est un format permettant d'associer des documents multimédias dans une présentation interactive.

Nous décrirons dans la suite du document les langages SMIL et Flash utilisables pour la présentation de documents annotés.

**SMIL :** Le langage SMIL du W3C est un langage XML, et possède donc une structure permettant des traitements. Ce langage utilise des espaces de nommage, permettant ainsi de créer de manière souple et propre des extensions de ce langage et de le rendre extensible. Un des avantages de SMIL est qu'il ne nécessite aucune modification du média à annoter. Ce langage permet de coder des annotations à propos d'un document audiovisuel sans avoir à toucher à ce dernier. De plus il possède des options de contrôle intéressantes sur la temporalité, les liens et la présentation des annotations et du document.

Un document SMIL se compose des mêmes parties « head » et « body » qu'un document HTML simple. Cependant, la définition de zones de visualisation est effectuable dans la partie « head » grâce aux éléments <layout> et <region>, déterminant des zones rectangulaires distinctes dans la page, repérées par les coordonnées de leur point supérieur gauche, leur hauteur et leur largeur. Il est aussi possible de définir si l'annotation doit s'afficher par dessus d'un autre objet, ou dans une fenêtre « pop-up », ou bien encore de la faire bouger avec l'objet qu'elle annote si celui-ci est dynamique.

Dans le corps du document, les objets à annoter et leurs annotations sont encapsulés dans une balise <par> afin de présenter en parallèle différents objets. Celle-ci permet de déterminer l'ensemble des objets qui nécessiteront une synchronisation, et par rapport à quel objet celle-ci devra être effectuée. Pour chaque objet présent dans ces balises (<img>, <video>, <ref>), il est alors possible de définir certains attributs :

- la zone dans laquelle il apparaîtra (correspondant à une région définie dans la partie entête)
- un identifiant pour l'objet
- la source du média
- le temps à partir duquel l'objet devra être présenté
- la durée de la présentation de l'objet

D'autre part, certaines annotations nécessitent de faire une pause dans la présentation d'un autre média, lorsque l'on veut présenter une image agrandie d'une vidéo par exemple. Cela est possible dans SMIL, en utilisant à la place de la balise <par> une balise <excl>, signifiant que les contenus sont exclusifs, et que donc un seul sera actif à la fois. Ainsi, dans le cas de la présentation d'une image au cours d'une vidéo, il sera possible de faire une pause de la vidéo à l'apparition de l'image pendant un certain temps, puis reprendre la lecture du document. Il est aussi possible grâce aux balises <priorityClass> de définir quel sera le média prioritaire pour la lecture.

Ce langage offre des possibilités de contrôle de contenus, tels que choisir quel objet faire apparaître en fonction de la langue du système. Les tests effectuables pour déterminer quel contenu utiliser sont personnalisables. On pourra ainsi présenter un contenu personnalisé en fonction du lecteur. Cette fonctionnalité est implémentée sous deux formes dans le langage SMIL : une balise <switch> permettant d'exprimer différents choix, et des attributs customTest ajoutable à chaque élément du document, et dont la valeur évaluée vaut vrai ou faux.

Enfin, SMIL offre des fonctionnalités étendues de liens entre les annotations et les documents. Il permet entre autres de définir une période pendant laquelle le lien pourra être activé, en ajoutant des attributs temporels à l'élément <a>. D'autre part, les sources des liens peuvent être définies finement. Ainsi, des zones peuvent être construites avec des balises <area> pour définir quelle partie d'une image sera considérée comme source du lien, et donc activable.

Pour recentrer sur le modèle d'hypermédia, SMIL est un langage permettant de visualiser et concrétiser les hyperliens entre la vidéo et une page Html. Il se situe donc à l'interface entre la couche de visualisation et la couche de stockage.

**Flash :** Flash est un langage de programmation développé par la société Macromédia qui offre la possibilité aux concepteurs de créer des présentations interactives avec le lecteur. Les projets réalisés en flash peuvent inclure des animations, du contenu vidéo, du texte animé, du son, *etc.* Les éléments de contenus créés avec Flash sont appelés généralement « applications ».

Les documents flashs sont particulièrement utilisés dans des présentations web, du fait de leur légèreté. Un document flash se compose de trois parties :

- La bibliothèque
- Le scénario
- Le code « ActionScript »

La bibliothèque regroupe l'ensemble des objets multimédia composant le document flash. Il peut s'agir d'images, de sons, de textes ou de vidéos. Ces objets peuvent être répartis dans une fenêtre de visualisation afin de préparer leur présentation finale.

Le scénario permet de définir les moments d'apparition des objets multimédias dans la présentation. Le scénario fonctionne par systèmes de calques disposés suivant une ligne temporelle. Les objets multimédias sont disposés suivant la ligne temporelle pour régler leur moment d'apparition, et sur des couches de calques, pour déterminer leur profondeur d'apparition (devant ou derrière).

Enfin, le code ActionScript va permettre de définir les interactions avec l'utilisateur et le comportement de l'application flash.

Les applications flash séparent en trois parties distinctes le contenu, sa présentation et les interactions et hyperliens disponibles pour le lecteur. Les documents flashs sont stockés sous différentes formes : un document .fla contiendra toutes les informations ci-dessus, un .swf est une version compressée d'un .fla (pour permettre l'affichage dans une page web), les .as sont des fichiers de code « ActionScript » et les .swc sont des composants flashs réutilisables directement dans d'autres applications flashs.

Pour conclure, flash est un langage permettant de créer des présentations interactives de contenu. Une présentation basée sur un document audiovisuel et fournissant via des hyperliens un contrôle sur ce document, permettant de naviguer entre celui-ci et les autres documents de la bibliothèque de la présentation est considéré d'après notre définition comme une hypervidéo.

Dans [Zhou 05], les auteurs proposent un système permettant de générer des vues hypertextes contenant des détails d'une vidéo sur demande. Cela permet de créer des pages de descriptions d'une vidéo, liées directement avec le flux temporel, ce qui est donc une hypervidéo. Ce système est basé sur la transformations des descriptions des segments de vidéo sur la norme MPEG-7 en langage SMIL afin de les intégrer à des pages web (figure 3.5). Il permet donc ensuite une navigation dans la vidéo à partir des pages Web générées, mais le « lecteur » n'a pas possibilité de modifier sa navigation. Ainsi, les pages générées sont statiques du fait qu'elles sont basées sur des annotations de fragment pré-générées. Le langage MPEG-7 est utilisé dans beaucoup d'applications travaillant sur les annotations de documents audiovisuels, comme par exemple dans Videozapper [Boavida 04], une application permettant d'adapter la vidéo au profil du lecteur.

En appliquant la définition des hypervidéos, on constate que la page HTML utilisant le langage SMIL, ou une application flash, est une vue d'une hypervidéo. Celle-ci est composée des descriptions MPEG-7 et du document vidéo, et un ensemble de règles définies par SMIL permet de générer la vue nécessaire. Le stockage des annotations est assuré par le langage MPEG-7, ainsi que celui de la vidéo. Celui-ci assure la structuration externe et interne des annotations. Du point de vue présentation et navigation,

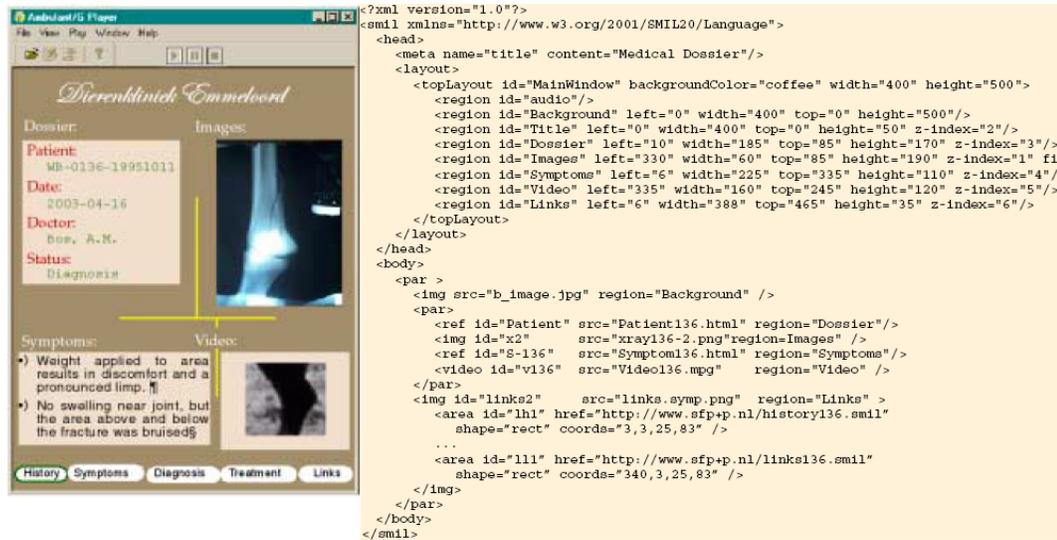


FIG. 3.5 – langage SMIL permettant de lier le flux à une page web.

celles-ci sont assurées par la vue web et les liens disponibles via celle-ci.

Rejoignant un peu le mode d'utilisation de SMIL, Annodex [Pfeiffer 03] et le langage CMML permettent d'accéder au contenu d'une vidéo via des URI augmentées d'une dimension temporelle. Ces langages se situent donc au niveau du stockage du document annoté, proposant une solution pour lier les annotations au flux vidéo d'une manière simple. Le langage CMML définit des balises à l'intérieur du flux vidéo, accessibles directement via son URI. La navigation dans les hypervidéos basées sur ces langages est donc définie à travers ces liens. Cependant, aucune spécification sur la visualisation n'est proposée par ce système. Le langage Atlas [Bird 00] établit lui aussi une interface entre les annotations de documents, leur utilisation et les documents eux-mêmes.

### 3.3.1 Systèmes pour la visualisation ou la navigation dans les hypervidéos

Certains systèmes facilitent plus la navigation et la visualisation des hypervidéos que leur création. Parmi ceux-ci, Hypercafé [Sawhney 96] est un prototype de système hypervidéo qui plonge le lecteur à l'intérieur de l'hypervidéo.

#### Exemple : Hypercafé

Dans ce système, l'accent est mis sur les opportunités de liens proposés au lecteur. Il place l'utilisateur dans un café virtuel dans lequel ont lieu plusieurs conversations, et lui offre la possibilité de naviguer entre celles-ci. Le spectateur peut donc suivre une conversation (vidéo) puis en suivant un lien hypertexte temporellement situé, continuer vers une autre vidéo. L'utilisateur n'a aucun contrôle sur le flux temporel de la vidéo, il se contente de le suivre et de suivre ou non les liens qui lui sont proposés. Les liens interviennent comme des propositions événementielles (figure 3.6). Si l'utilisateur souhaite les suivre, il se retrouve dans une autre situation, suivant une autre conversation,



FIG. 3.6 – Hypercafé : différentes opportunités de conversation.

sinon le lien disparaît au bout de quelques instants. Entrer dans hypercafé est donc similaire à pénétrer dans un café et tenter de suivre les conversations aux différentes tables. Ce système permet donc de naviguer différemment dans un flux vidéo, mais le lecteur n'est que consommateur du flux. Les activités de modification de l'hypervidéo et de lecture de celui-ci, comme pour Hyper-Hitchcock sont dissociées.

**Modèle :** Dans hypercafé, les hypervidéos sont formées d'un ensemble de documents audiovisuels et de liens entre eux. Les éléments atomiques de l'hypervidéo sont des documents audiovisuels distincts, mis en relation par des liens spatio-temporellement situés. La présentation peut être multiflux : le flux principal est présenté au lecteur, ainsi que les flux cibles des liens disponibles au cours de la lecture. Ces flux sont plus ou moins actifs, suivant qu'ils sont l'objet central de la lecture, la cible d'un lien actif, la cible d'un lien passé ou futur ou encore la source d'un lien suivi. La définition d'un flux peut être rapprochée de celle des canaux de visualisation présentés dans le modèle AHM. Les liens de ces hypervidéos peuvent être localisés temporellement, mais aussi spatialement. Ainsi le lecteur peut explorer les détails de la vidéo. Plusieurs liens peuvent donc être actifs en même temps dans une hypervidéo. Le choix de les suivre est laissé au lecteur. Ce système est donc composé de plusieurs ensembles :

- Des documents audiovisuels.
- Des liens bidirectionnels spatialement et temporellement situés.
- Un ensemble de flux pour les visualiser.

Les documents audiovisuels sont présentés suivant différents flux, mais ceux-ci ne peuvent être interrompus par le lecteur. Celui-ci n'a aucun contrôle sur le déroulement temporel des vidéos. Les liens lui sont présentés de manière opportuniste en ajoutant un flux et en augmentant son volume sonore pour exprimer le fait qu'un lien entre le flux principal et celui-ci est disponible. Si le lecteur décide de suivre ce lien, l'ancien flux principal devient inactif, c'est à dire qu'il continue de se dérouler dans le temps mais son niveau sonore baisse et il devient moins important, tandis que la cible du

lien devient à l'inverse, active. Le seul contrôle que le lecteur possède sur la lecture de l'hypervidéo est la possibilité de naviguer dans celle-ci.

Ce système est donc composé d'une couche de stockage comprenant les différents documents audiovisuels et les liens entre ceux-ci. L'information concernant le contexte des liens, temporellement et spatialement situés, est elle aussi stockée, pour pouvoir être présentée au lecteur. La visualisation est directement liée à la navigation. Elle n'est possible que suivant la vue prédéfinie par l'ensemble des liens et documents audiovisuels, mais elle est multiflux.

**Utilisation :** Dans hypercafé, le lecteur est directement plongé dans l'hypervidéo. Il n'a aucun contrôle sur le flux vidéo, mais peut, en fonction des liens disponibles à chaque instant, recentrer son attention sur le sujet proposé qui l'intéresse. Le fait de proposer en multiflux les différentes parties de l'hypervidéo permet au lecteur d'être conscient des différentes opportunités de navigation proposées.

### 3.3.2 Systèmes pour la création et l'édition d'hypervidéos

D'autres systèmes sont beaucoup plus orientés en vue de la création des hypervidéos que pour leur visualisation. Ils permettent, à partir de différentes ressources (vidéos et autres documents) de créer des hypervidéos mais offrent peu de moyens de visualisation. Ils sont donc beaucoup plus liés à la couche de stockage que les précédents systèmes.

#### Exemple : Hyper-hitchcock

Hyper-Hitchcock [Shipman 05, Shipman 03] est un système d'édition d'hypervidéos à partir de fragments de vidéos réorganisés (figure 3.7). Le but de ce projet était de permettre l'exploitation de vidéos de manière non linéaire, non bornée à la ligne de temps de la vidéo. Le principe de construction d'une hypervidéo avec hyper-hitchcock est assez simple. Une hypervidéo est constitué d'un ensemble de fragments de vidéos, contenant des liens pour naviguer vers d'autres fragments. Ainsi un lecteur pourra approfondir sa lecture en suivant les liens vers des vidéos plus détaillées vers le sujet qui l'intéresse. Chaque fragment de vidéo ne peut cependant contenir qu'un seul lien vers une autre vidéo à un moment donné. Hyper-hitchcock offre un espace de travail pour organiser les fragments de vidéos afin de composer une hypervidéo, ainsi qu'une interface sommaire de visualisation de celle-ci. De plus, il offre la possibilité de générer automatiquement des sommaires pour les hypervidéos ainsi créées. Cependant, les activités de lecture et de constitution de l'hypervidéo sont ici dissociés, et l'accent est porté sur l'édition plus que la visualisation.

**Modèle :** Comme nous l'avons expliqué précédemment, le système d'Hyper-hitchcock se base sur un ensemble de fragments de vidéos, qui seront reliés entre eux par des liens, pour former une hypervidéo. Les éléments atomiques de cette hypervidéo sont donc des fragments de vidéos et des liens. Un fragment peut être la cible de plusieurs liens, mais ne peut pointer que vers un seul autre fragment à un moment donné. Les liens sont fixés temporellement sur un intervalle pendant lequel ils sont actifs. A chaque lien est associé une description textuelle, ainsi qu'un mode de retour, définissant si après avoir fini de lire le segment cible du lien, l'hypervidéo doit revenir à son point de départ, ou à une coordonnée temporelle spécifique. Le lien possède donc une information sur son

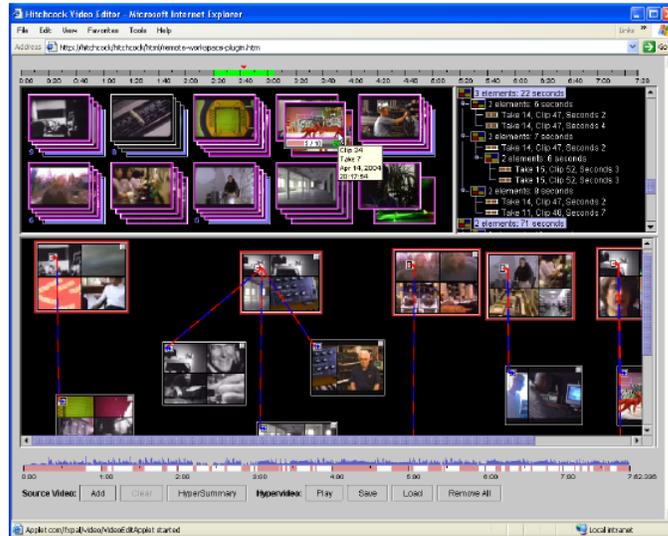


FIG. 3.7 – Interface d’hyper-Hitchcock

contexte d’origine. L’hypervidéo obtenue est une vidéo à partir de laquelle des liens vers d’autres fragments de vidéo sont disponibles au cours de la lecture.

**Utilisation :** Ce système permet la création de présentation interactive de clips vidéos, sous forme d’hypervidéos. L’interface de création permet la manipulation directe des fragments de vidéo, et permet de manière très intuitive d’associer ces fragments entre eux pour former une hypervidéo. A chaque lien peut être associée une description qui est ensuite affichée pendant la visualisation, fournissant ainsi un complément d’information sur la cible du lien. Le fait de pouvoir choisir différentes modalités de retour permet aussi de rendre l’hypervidéo plus intuitive. Prenons l’exemple d’une hypervidéo présentant une recette de cuisine. A chaque étape, un lien est disponible pour visualiser de manière plus détaillée, ce qui est expliqué dans la vidéo de base. A la fin de la visualisation du détail, il est plus judicieux de revenir à la fin du passage explicite, plutôt qu’au moment où on a décidé de suivre le lien. Le système d’hyper-hitchcock permet aussi de visualiser les hypervidéos créées, et de naviguer dans celles-ci en suivant les liens, mais cette vue fournie par le système est relativement restreinte.

Dans l’exemple de la figure 3.8, l’hypervidéo est composée de quatre vidéos distinctes. Dans la première, un lien est proposé lors de la lecture du premier et du troisième fragments. Si le lecteur désire suivre par exemple, le lien du troisième fragment, sa lecture de cette vidéo va être interrompue et remplacée par celle de la vidéo 3. Lors de cette lecture, un lien va lui être proposé au cours du deuxième fragment de la vidéo. Là encore, s’il le suit, sa lecture de cette vidéo va être remplacée par la lecture de la vidéo 4. A la fin de cette lecture, le lien va le ramener directement au début du fragment d’où il venait. S’il termine alors la lecture de la troisième vidéo, le lien qu’il avait suivi pour y arriver le ramènera à la fin du troisième fragment de la première vidéo.

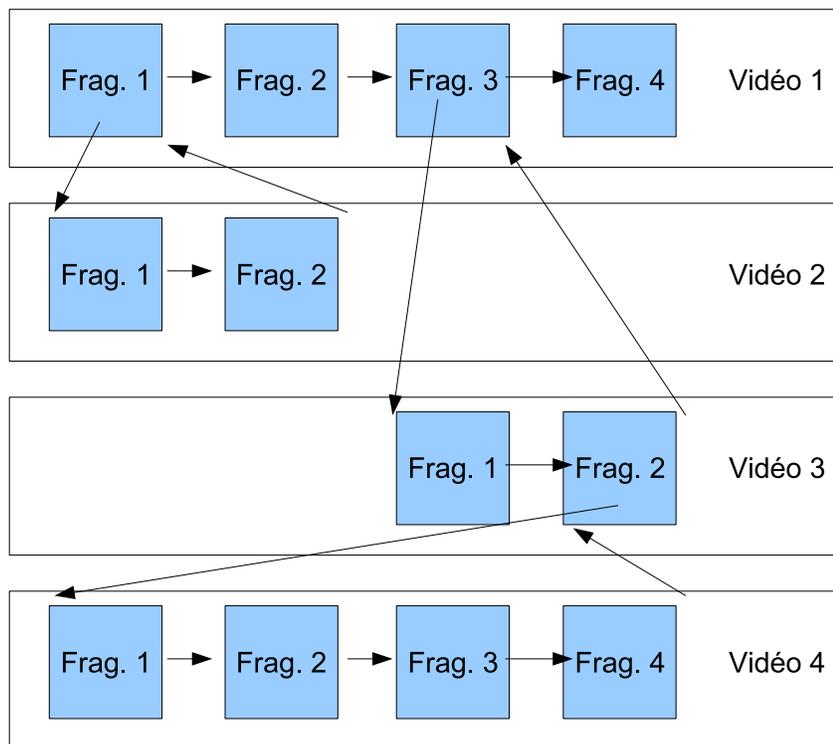


FIG. 3.8 – Exemple de structure d'une hypervidéo d'Hyper-hitchcock

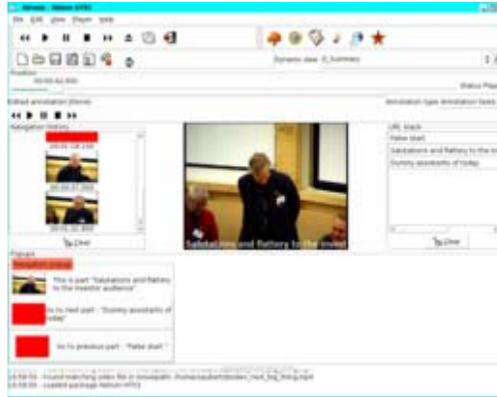


FIG. 3.9 – Interface principale d’advene, en cours de visualisation d’une hypervidéo

### 3.3.3 Systèmes combinant édition et visualisation

D’autres systèmes combinent création et visualisation d’hypervidéo, permettant tout en lisant la vidéo d’annoter celle-ci et de construire l’hypervidéo sur le document et ses annotations. En effet, les deux premiers types de systèmes d’hypervidéos proposent la création et la visualisation, mais sont manifestement plus destinés à l’une qu’à l’autre de ces deux fonctions.

#### Exemple : Advene

Dans [Aubert 04b, Aubert 05a], les auteurs présentent un système de création et gestion d’hypervidéos basées sur les annotations : Advene. Le système permet à un lecteur d’annoter une vidéo, puis d’utiliser ces annotations pour contrôler le flux temporel de la vidéo et le présenter suivant différentes vues. Ainsi une indexation des parties de la vidéo sous forme de sommaire HTML peut être créée, ou encore une boucle de lecture de fragments de la vidéo. Le lecteur peut dans ce système jouer en même temps le rôle d’acteur, créateur de l’hypervidéo. En effet, l’interface lui permet d’annoter le document, activité indispensable dans la lecture active, et de créer des vues de l’hypervidéo basées sur ses annotations, puis de les visualiser (figure 3.9). Il est même possible d’annoter le document tout en le visualisant à partir d’une vue spécifique, basée sur d’autres annotations.

**Modèle :** Advene est une plateforme permettant d’une part l’annotation de documents audiovisuels suivant des schémas d’annotation et d’autre part la visualisation sous forme d’hypervidéos de ces documents annotés [Aubert 04a]. La notion centrale du modèle d’Advene est celle de recueil : Un recueil est le regroupement de l’ensemble des annotations d’un document audiovisuel, classées suivant des types et organisées entre elles par des relations, ainsi que des vues. Celles-ci représentent des ensembles de fragments du flux annotés et organisés pour produire des hypervidéos. L’ensemble formé du document audiovisuel, des annotations et des relations entre celles-ci constitue la couche de stockage. La partie visualisation est assurée par les vues qui définissent un angle de perception de l’hypervidéo.

La figure 3.10 représente en UML le modèle Advene.

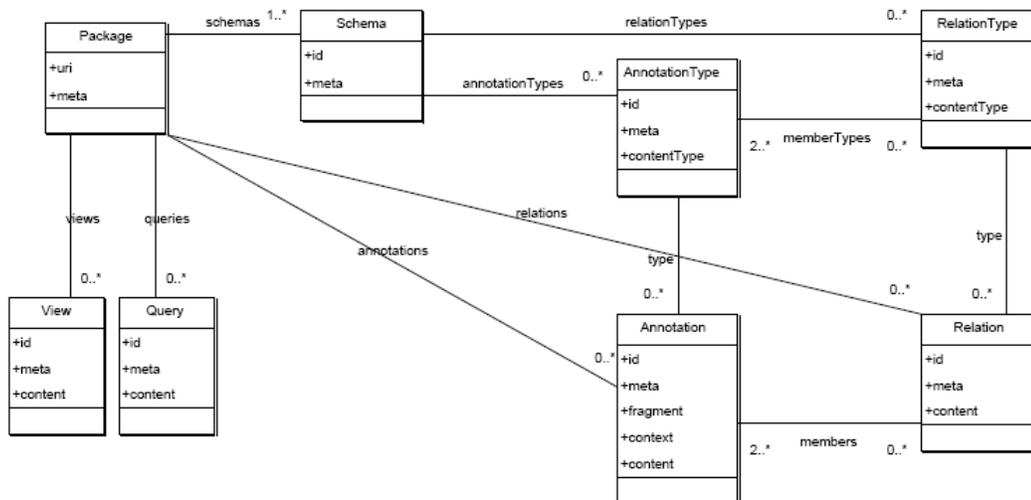


FIG. 3.10 – Représentation UML du modèle d’Advene

Les annotations sont classées suivant des types définis par l’annotateur, pouvant être contraints par des schémas d’annotation, et sont des données attachées à des fragments du flux audiovisuel. D’autre part, les annotations peuvent être en relation entre elles. Un contenu spécifique peut être associé à ces liens entre les annotations. Enfin, les vues sont des manières de présenter le contenu du document audiovisuel annoté au lecteur. Celles-ci sont aussi définies par l’utilisateur, et permettent donc de générer des hypervidéos. Celles-ci peuvent être globalement statiques, offrant une présentation HTML par exemple, avec liens directs sur le flux audiovisuel, ou globalement dynamiques, offrant une présentation du flux audiovisuel avec de manière événementielle des liens vers d’autres contenus ou la présentations d’annotations au lecteur.

**Utilisation :** La plateforme Advene permet donc les deux types de manipulations des hypervidéos vues précédemment. D’une part, elle permet de créer et d’éditer des hypervidéos, en annotant la vidéo d’origine et en définissant des vues basées sur ces annotations, mais elle permet aussi de visualiser les hypervidéos ainsi créées, composées du document et de ses annotations présentés suivant ces vues. Dans une présentation « dynamique » d’une hypervidéo, les contenus des annotations peuvent être présentés de plusieurs façon. Cela peut aller du sous-titre de la vidéo à l’ouverture de liens dans des navigateurs web ou la présentation de ces liens dans une « pile d’URL ». Le lecteur garde toujours le contrôle sur le flux, et peut même, dans le cas où la vue définie le permet, naviguer en suivant les relations entre les annotations. Enfin, la présentation « statique » d’une hypervidéo peut permettre de présenter plusieurs accès à des fragments de document audiovisuels à partir d’un sommaire HTML par exemple. Ce type de présentation est assez semblable à la présentation des résultats de recherche des systèmes d’information audiovisuels tels que Internet Archive [internet Archive 04].

Nous avons donc précisé dans ce chapitre les notions d’hypermédias et d’hypervidéos. Nous y avons exposé les principaux modèles d’hypermédias existants, ainsi que quelques exemples. Enfin nous avons présenté quelques systèmes de gestion d’hypervidéos actuels, leurs modèles et leur utilisation.

## Chapitre 4

# Systemes d'information audiovisuels

Dans ce chapitre nous aborderons la notion de système d'information audiovisuel. Nous fournirons tout d'abord une définition de celle-ci, puis nous présenterons les principales phases composant le cycle de vie d'un SIAV : indexation, recherche, exploitation. Nous présenterons pour cela différents modèles d'indexation de documents audiovisuels, les différentes méthodes de recherche d'information et quelques possibilités d'exploitation. Enfin nous présenterons un cas particulier de SIAV auquel nous nous intéresserons plus particulièrement par la suite, les SIAV personnels pour la lecture active.

### 4.1 Définition

*Définition :* Les systèmes d'information audiovisuels (AVIS) fournissent des moyens d'indexer, de rechercher et d'utiliser des documents audiovisuels.

Ce domaine de recherche est très actif [Slaughter 00, Hauptmann 02], même si des applications de celui-ci sont déjà utilisées couramment [autonomy Inc. 04]. L'utilisation d'un système d'information audiovisuel peut se décomposer en cinq parties (voir figure 4.1) : création, stockage, indexation, recherche d'information et exploitation.

Toutes les étapes de ce cycle ne sont pas présentes dans chaque système d'information audiovisuel. Les trois étapes clés, que nous décrirons plus précisément dans la suite de ce chapitre, sont l'indexation, la recherche et la visualisation. L'étape de stockage du document n'est pas nécessaire. En effet, les documents peuvent très bien être indexés sans pour autant être stockés par le système, avec juste une référence vers le lieu où le document est effectivement stocké. Cela pose tout de même un problème : comment être sûr que le document n'aura pas changé de place, que celui-ci sera toujours disponible. Ce problème est commun à tous les systèmes d'information.

L'étape d'édition du document et de création est, elle aussi, particulière. Peu de systèmes permettent à l'heure actuelle de modifier directement le document stocké. Cependant, certains systèmes comme JumpCut dont nous parlerons plus loin permettent de créer de nouveaux documents audiovisuels à partir de documents existants dans le système. Ces nouveaux documents donnent lieu à une nouvelle étape de stockage et une indexation dans le système.

L'indexation est la base des SIAV.

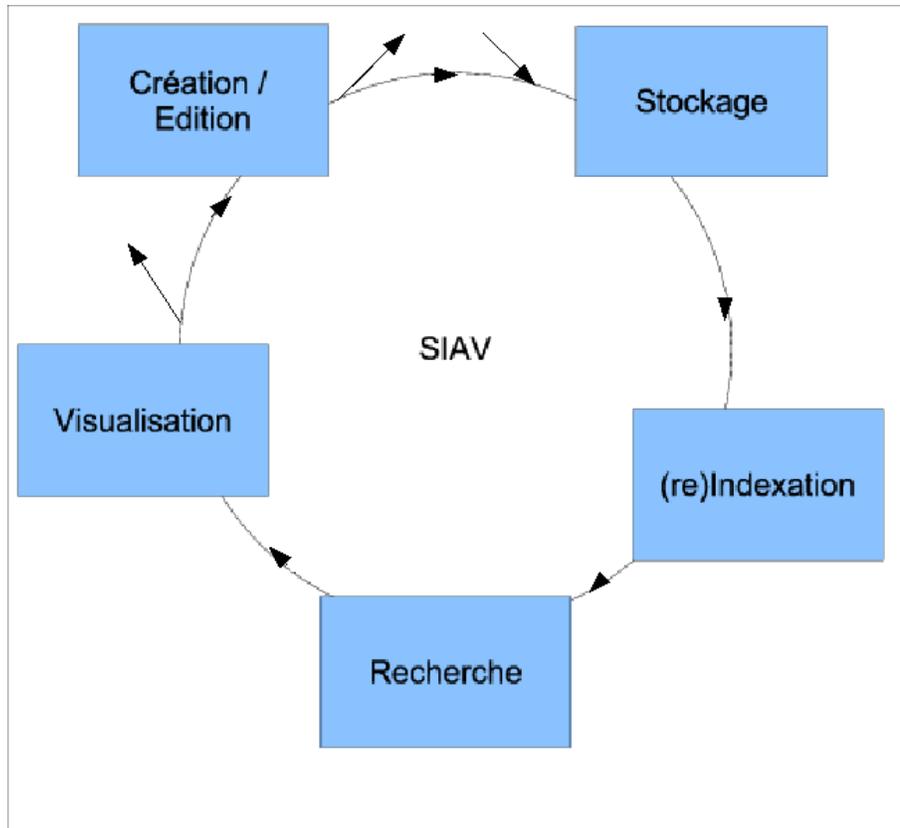


FIG. 4.1 – Cycle d'utilisation d'un SIAV

## Search Results

Results: 1 through 50 of 276 (0.19 secs)

You searched for: collection:opensource\_movies AND firstTitle:O



### [O Christmas Tree](#)

Charlie Brown ain't got nothin on our family tree.

Keywords: [christmas](#); [tree](#); [san antonio](#); [geroianni](#)



### [O Fortuna - War, Bombs, Death, and Destruction Archetype Splice](#) - Max Tohline

In what he dubs an "Archetype Splice," Max Tohline uses instantly recognizable archival footage from archive.org and other websites to communicate vague themes of societal collapse and explosion, using footage from Edison's The Great Train Robbery and music by Carl Orff to bind the work together. As of 27 January, 2006: Due to copyright concerns, this film has been indefinitely removed from archive.org...

Keywords: [collapse](#); [explosions](#); [bombs](#)



Downloads: 1,823 **Batting Average:** 10.61%

Average rating: ★★★★★ (2 reviews)

FIG. 4.2 – Présentation des résultats d'une recherche dans Open Archive, avec alternance de plans fixes.

*Définition* : Indexer un document audiovisuel, ou un corpus de documents audiovisuels signifie proposer une description du document ou du corpus de documents. Il existe plusieurs manières d'indexer un document. On peut l'indexer dans sa globalité, en tant que document audiovisuel, ou l'indexer sur son contenu précis, en se basant sur une description de chaque fragment du document. L'indexation est en général réalisée dans deux optiques distinctes : la première orientée « documentaliste » vise à indexer précisément et classer dans une catégorie chaque document, la seconde, orientée « chercheur » au contraire ne cherche pas à indexer exclusivement un document, mais plutôt à l'exploiter sous tous les angles possibles. Il semble donc nécessaire de pouvoir fournir à l'utilisateur des systèmes d'information un moyen de se construire sa propre indexation des documents. Un moyen pour cela semble être de baser l'indexation sur les annotations de l'utilisateur du document. Ces annotations de haut niveau sémantique [Christel 05] peuvent être créées manuellement en phase de visualisation par exemple.

La phase suivante est la recherche d'information. L'utilisateur interroge le système d'information audiovisuel, en formulant une requête par les moyens fournis par le système. Il peut s'agir d'une recherche booléenne, d'une recherche vectorielle, ou d'une exploration.

Une fois les informations relatives à la requête de l'utilisateur trouvées, il faut pouvoir présenter les différents documents à celui-ci. Or, contrairement à des documents textes ou images, simples à présenter, les documents audiovisuels possèdent un facteur temporel difficile à gérer. La présentation de ces collections de vidéos se fait généralement en présentant des plans fixes de celles-ci dans des documents hypertextes (figure 4.2), parfois en permettant des lectures rapides de parties des vidéos [Graham 03]. Enfin, l'exploitation du résultat final de la requête se fait en général à travers la visualisation du document choisi [internet Archive 04]. Cependant, des outils permettant d'autres types de manipulation telles que l'édition ou la création d'hypervidéos sont des sujets de recherche actifs.

## 4.2 Indexation et description de documents audiovisuels

Comme nous l'avons expliqué dans le paragraphe précédent, l'indexation est la base de fonctionnement des systèmes d'information documentaire. Afin de pouvoir faire des raisonnements, des recherches ou tout autre traitement sur le corpus de documents audiovisuels disponible, il est nécessaire d'indexer ceux-ci avec des descripteurs organisés suivant une structure prédéfinie. Il est nécessaire de noter que les documents audiovisuels n'offrent pas *a priori* de décomposition en éléments discrets signifiants pouvant être indexés et qu'il n'est pas possible de décrire objectivement l'ensemble image et son composant ces documents. Il est donc nécessaire, comme l'explique [Auffret 99], d'indexer le document audiovisuel totalement, pour décrire chaque partie de celui-ci. Nous allons donc présenter les différentes méthodes de structuration des descripteurs de documents audiovisuels.

### 4.2.1 Approche de descriptions

#### Schémas de données

La méthode la plus classique de structuration provient du domaine des bases de données. Les différents descripteurs utilisés pour décrire le contenu du document ou celui-ci dans son ensemble sont organisés suivant un schéma de données. Celui-ci définit des classes d'objets regroupant des ensembles d'attributs communs. Les modèles sous-jacent sont les modèles Entité-association ou UML par exemple. Les langages d'interrogation de base de données sont utilisés pour exécuter les requêtes et renvoyer les résultats.

#### Schémas de connaissances

Une autre façon de décrire les documents audiovisuels est de le faire suivant une ontologie. On définit dans celle-ci tous les termes et éléments permettant d'annoter les documents. Dans [Bertini 05], les auteurs proposent d'utiliser une ontologie enrichie du domaine concerné par le document pour structurer les annotations. Ainsi, ils utilisent une ontologie du football enrichie par des images pour former le contenu des annotations de leur vidéo (voir figure 4.3). Les documents sont annotés suivant les termes définis par l'ontologie. Cela permet de cadrer plus précisément l'indexation et aussi de fournir des capacités de raisonnement sur celle-ci. Ainsi, comme l'explique B. Bachimont dans [Bachimont 00], les descriptions pourront être complétées par des inférences logiques sur les descripteurs déjà associés au média. [Saarela 98, Srikanth 05, Isaac 04] sont d'autres exemples d'utilisation d'ontologies pour structurer des annotations à des fins d'indexations.

#### Schémas documentaires

Un troisième moyen utilisé pour annoter les documents audiovisuels est de les décrire suivant des schémas documentaires. Ainsi le format MPEG-7 [Tran-Thuong 03] décrit en annexe B fournit des schémas, dérivés de XMLSchema [W3C 06a, W3C 06b], permettant de décrire des documents ou des fragments de documents audiovisuels. Le document et son contenu sont décrits suivant une structure arborescente. Le format MPEG-7 veut permettre la standardisation des descriptions de contenu à tous les niveaux (structure, sémantique, bas niveau), pour faciliter la gestion des documents audiovisuels. Le langage permet d'associer des descripteurs aux éléments de la vidéo,

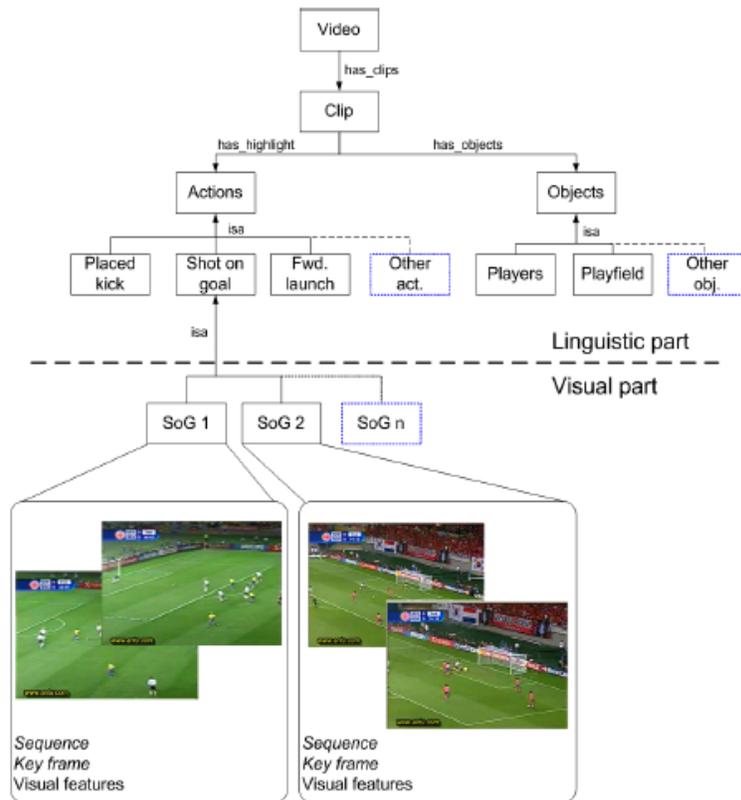


FIG. 4.3 – Ontologie enrichie par des images

organisés suivant des schémas eux-mêmes définis selon un modèle de description issu de XMLSchéma.

D'autre part, il existe actuellement différents travaux pour mixer les approches documentaires et cognitives. Ainsi des équipes travaillent sur l'utilisation d'ontologies pour ajouter des connaissances et possibilités de raisonnements sur des schémas documentaires type XML.

#### **4.2.2 Modèles d'indexation**

Nous présenterons brièvement dans cette partie les modèles principaux d'indexation. Le premier est le modèle à description structurée, le second est la classification hiérarchisée.

##### **Modèle à notices**

Dans ce modèle, à chaque document audiovisuel est associé une notice descriptive. Celle-ci regroupe un ensemble de champs permettant de décrire le document et son contenu. Chaque champ possède une valeur, qui peut être issu d'un vocabulaire prédéfini ou non. L'indexation se fait donc suivant les différents champs des notices.

##### **Modèle à classification hiérarchisée**

Dans ce modèle le document est classifié suivant un ensemble de concepts hiérarchisés. Le document peut appartenir à plusieurs branches de la classification s'il est concerné par différents concepts. C'est donc plus le document qui est associé aux mots-clés prédéfinis que le contraire.

D'autres méthodes d'indexation existent, elles aussi basées sur des mots-clés associés au document. Ces mots clés peuvent être des descripteurs ou des non-descripteurs, et peuvent appartenir à un vocabulaire prédéfini (par un thésaurus ou un index par exemple) ou être libres.

### **4.3 Recherche d'informations et exploration**

Les problèmes de la classification et de la recherche de documents audiovisuels pertinents pour une requête utilisateur sont nombreux. En effet, la recherche dépendant principalement du mode d'indexation choisi, toutes les méthodes de recherche ne sont pas applicables à tous les systèmes. Plusieurs approches ont déjà été proposées pour améliorer et faciliter cette démarche, basées sur différents types de recherche :

- recherche booléenne
- recherche vectorielle
- recherche exploratoire

La recherche booléenne s'effectue en particulier sur les documents indexés par des mots-clés. L'utilisateur émet une requête en demandant que tel ensemble de mots-clés soit présent, et tel autre absent. Il est aussi utilisé sur les modèles structurés, pour lesquels les requêtes deviennent des listes de champs et valeurs souhaités (comme nom de l'auteur = B. Richard). Si les requêtes s'effectuent sur des descripteurs appartenant à un vocabulaire prédéfini, celui-ci peut être proposé à l'utilisateur pour faciliter ses recherches. Les solutions proposées correspondent exactement à la requête.

La recherche vectorielle a pour principe de considérer que les documents disponibles et le document recherché appartiennent à un même espace vectoriel. Chaque

descripteur valable représente une dimension de l'espace vectoriel. Ce type de recherche ne peut s'effectuer que sur des ensemble de descripteurs mesurables, et non sur du texte libre. La recherche consiste à calculer la distance entre le document recherché projeté dans l'espace vectoriel et les documents indexés. Les solutions proposées ne sont pas forcément exacte, mais sont les plus proches possibles de la requête.

La recherche par exploration consiste à fournir une vision explorable de l'espace de recherche à l'utilisateur. Ce type de recherche est particulièrement adapté aux modèles à hiérarchie. L'utilisateur choisit le domaine correspondant le plus à ce qu'il recherche, et affine en descendant dans les profondeurs de la hiérarchie.

Nous pouvons présenter quelques travaux illustrant ces concepts. Dans [Christel 05], l'auteur propose d'annoter les documents par différents concepts, pondérés sur une échelle de 1 à 5. La recherche est ensuite basée sur cette liste de critères. Les résultats présentés sont encourageants, mais il est cependant contraignant de devoir baser sa recherche sur des mots-clés prédéfinis.

Dans Vox populi [Bocconi 05], un autre principe de recherche est abordé. Le système présenté permet de visualiser des documents audiovisuels centrés sur un thème particulier, en créant une sorte de documentaire. L'utilisateur doit donc d'abord choisir un thème dans une liste de mots-clés, comme pour le système précédent, mais peut aussi définir s'il souhaite que la vidéo présentée soit pour, contre, ou neutre par rapport à celui-ci. Ensuite, la recherche peut être complétée par d'autres documents exprimant une opinion sur le sujet.

Enfin, dans [Hauptmann 02], l'auteur présente différents descripteurs sémantiques pouvant être associés aux documents audiovisuels pour faciliter leur recherche. Ainsi, il propose d'associer à chacun de ces documents des concepts définissant si l'action se passe en intérieur, extérieur, en ville, *etc.* De plus il s'appuie sur des descripteurs bas niveau des vidéos pour améliorer son indexation. La recherche est donc basée sur ces descripteurs, et la présentation du résultat se fait sous la forme d'un ensemble de segments vidéos (figure 4.4).

La présentation des résultats de la recherche à l'utilisateur pose aussi certains problèmes. Si seuls des plans fixes des vidéos sont présentés, ceux-ci peuvent ne pas être très représentatifs du contenu du document. D'autre part, présenter directement l'ensemble de flux vidéos est extrêmement coûteux. La méthode intermédiaire présentée en figure 4.2 paraît être un bon compromis. Une alternance de plans fixes représentatifs des différentes parties du contenu des vidéos permettent de se faire une meilleure idée de celui-ci sans toutefois alourdir le système. La présentation de l'ensemble des descripteurs ayant permis de les sélectionner peut aussi être affichée. Certains systèmes comme Grouper <sup>1</sup> permettent de visualiser l'ensemble des résultats sous forme d'une mosaïque d'images, et de zoomer dessus pour choisir laquelle doit être visualisée.

## 4.4 Exploitation

### 4.4.1 Visualisation simple

Dans les systèmes d'information audiovisuels actuels, l'exploitation des résultats se traduit bien souvent par une simple visualisation de celui-ci. Une fois que l'utilisateur a choisi quel document correspond à sa requête, il peut visualiser celui-ci directement. C'est le cas par exemple de « google video » <sup>2</sup>, qui offre la possibilité de lire la vidéo

---

<sup>1</sup><http://www.grouper.com/>

<sup>2</sup><http://video.google.com/>



désirée parmi les résultats obtenus. Les résultats de la requête sont affichés sous forme de vignettes accompagnées de leur titre. Lire la vidéo ouvre un lecteur multimédia intégré à la page consultée, fournissant comme seul contrôle sur le flux la possibilité de faire des pauses dans la lecture. D'autres SIAV comme celui de la CNN <sup>3</sup> ou « Youtube » <sup>4</sup> permettent eux aussi simplement de visualiser les documents audiovisuels. La possibilité est laissée à l'utilisateur de les regarder dans la fenêtre de visualisation prévue ou en plein écran.

#### 4.4.2 Réutilisation

D'autres fonctionnalités d'exploitation commencent à apparaître. Ainsi on trouve des systèmes permettant l'édition des documents audiovisuels, la publication de ceux-ci directement dans d'autres documents via un lien. Grouper<sup>5</sup> par exemple, permet de lire la vidéo obtenue, de la télécharger sur son ordinateur, ou de la lier depuis un autre site internet. La vidéo peut donc être intégrée dans ses propres pages web, et il est même possible de lier la collection de l'utilisateur sous forme de vignettes. D'autre part, il est possible dans Grouper de visualiser le résultat des requêtes sous la forme d'une mosaïque d'images, de zoomer sur les différentes images, et de visualiser la vidéo correspondante d'un clic. « MTV overdrive » <sup>6</sup> ou « Youtube » permettent aussi de constituer une liste de lecture de différentes vidéos.

De nouvelles exploitations de ces documents audiovisuels commencent à être intégrés aux SIAV, comme le montage de vidéo. « Jumpcut » <sup>7</sup> permet de créer ses propres clips vidéos à partir de l'ensemble des vidéos présentes sur le site, en les organisant, insérant des transitions et changeant la bande sonore. Enfin, « ClickTv » <sup>8</sup> permet d'annoter le document vidéo et de visualiser ces annotations au cours de la lecture. Cliquer sur une annotation positionne le document vidéo au début de celle-ci, et lire la vidéo séquentiellement fait défiler la liste des annotations simultanément pour garder la correspondance. Le fait d'ajouter des commentaires à des parties de la vidéo permet de faire des recherches plus précises sur le contenu de celle-ci. En effet, plutôt que de rechercher un tag dans un ensemble de tags associés à une vidéo, la recherche s'effectue sur le texte décrivant son contenu.

### 4.5 SIAV personnel pour la lecture active

La conception des systèmes d'information audiovisuels peut tirer des bénéfices de la lecture active. Ainsi, dans le cas des SIAV utilisés pour gérer des petits corpus de documents pour des chercheurs, ce que nous appellerons les Systèmes d'information audiovisuels personnels, les documents ont besoin d'être en constante réindexation en fonction de l'avancement des travaux de recherche.

Ces systèmes d'information audiovisuels ont pour caractéristiques :

- d'être centrés sur un usage
- de gérer de petits corpus de documents
- d'être très fréquemment exploités

---

<sup>3</sup>[www.cnn.com](http://www.cnn.com)

<sup>4</sup><http://www.youtube.com/>

<sup>5</sup><http://www.grouper.com/>

<sup>6</sup><http://www.mtv.com/overdrive/>

<sup>7</sup><http://www.jumpcut.com/>

<sup>8</sup><http://www.click.tv/>

C'est souvent ce type de SIAV qui est utilisé pour gérer des corpus en cours d'étude, et pour lesquels la lecture active permettrait de les exploiter plus facilement. En effet, ces SIAV nécessitent d'être fréquemment mis à jour et réindexés. La lecture active est une activité fournissant de nombreuses informations sur le contenu des documents. Celles-ci peuvent donc être ensuite exploitées pour fournir une indexation basée sur la description du contenu du document. Ces informations sont, de plus, généralement très précises et adaptées à l'activité en cours. Elles permettraient donc un enrichissement de l'indexation, des recherches plus approfondies du document, et en phase de visualisation, des fonctionnalités supplémentaires telles que l'annotation.

Nous avons rapidement présenté dans ce chapitre les systèmes d'informations audiovisuels et leurs principales caractéristiques. Nous avons abordé les différentes techniques d'indexation et de recherche utilisés à ce jour. Enfin nous avons présenté les nouvelles possibilités d'exploitation du résultat de ces recherches, de visualisation et de réutilisation des documents obtenus.

## Chapitre 5

# Conclusion et perspectives

Dans ce document, nous avons présenté un état de l'art des différents domaines liés à la lecture active de documents audiovisuels. Nous avons d'abord rappelé les différents aspects de la lecture active et les avons adaptés aux documents audiovisuels. Puis nous avons évoqué les travaux sur les hypermédias et hypervidéos, produits issus de la lecture active de documents et de l'exploitation de leurs annotations. Enfin nous avons rappelé les différentes fonctionnalités des systèmes d'information audiovisuels et exposé de quelle manière la lecture active commence à prendre place dans leur cycle de fonctionnement.

Au cours de ce travail, nous avons pu mettre en avant certains points importants concernant la lecture active de documents.

Tout d'abord, la lecture active de documents audiovisuels est une activité complexe, s'exprimant généralement par la production d'annotations sur ces documents. Le document annoté lors du cycle de lecture active constitue donc un enrichissement considérable du point de vue sémantique du document d'origine.

Il est nécessaire de laisser une liberté aussi grande que possible dans les annotations employées pour mener cette lecture active, afin que le lecteur puisse exprimer tout ce qu'il souhaite. Cela s'applique aussi bien à la forme des annotations, qu'à leur contenu ou qu'au fragment de document qu'il peut être possible d'annoter.

Cependant, afin de pouvoir réutiliser ces annotations, pour présenter sous forme d'hypervidéo une vue du document annoté par exemple, il est aussi nécessaire de leur fournir une structure de description suffisamment rigide pour être réexploitées.

Enfin, ces annotations produites lors de la lecture active peuvent être réutilisées dans le cadre des systèmes d'informations audiovisuels pour fournir une description plus poussée du document en vue d'une indexation par le lecteur. Le document est alors soumis à une réindexation constante dans le système.

La suite du travail consistera à effectuer une théorisation de l'activité de lecture active de documents audiovisuels. Pour cela, nous commencerons par étudier la théorie de l'activité afin de fournir un cadre théorique solide à notre proposition. Nous effectuerons ensuite la modélisation des différentes phases composant la lecture active de documents audiovisuels ainsi que des relations entre elles. Nous nous pencherons aussi sur les différentes opérations possibles pendant cette activité, et nous étudierons l'évolution du document et des annotations au cours du cycle de lecture active. Les éléments principaux à prendre en compte pour cette théorisation seront le document, les annotations et le lecteur.

Nous nous intéresserons ensuite aux modèles de description permettant de fournir une structure à la fois souple et rigide pour les annotations, ainsi qu'aux différentes façons de les construire et les faire évoluer pour satisfaire le plus possible aux besoins d'un lecteur actif. Pour cela, nous ferons un point sur les travaux effectués dans le domaine de la structuration de données et/ou d'idées, afin de déterminer quel type de structuration est la plus adaptée pour ce type d'activité. Il est en effet nécessaire de trouver une façon souple de structurer les annotations des documents audiovisuels, comme expliqué dans [Nack 01]. De même, une étude des différentes façons de construire ce type de schémas est nécessaire, ceux-ci pouvant être construits *a priori*, extraits *a posteriori* ou même directement construits et modifiés en cours d'utilisation comme les « Mindmaps » [Shum 03], ou carte heuristique.

Nous définirons donc un modèle de description adapté aux annotations de la lecture active, pour lequel nous fournirons un formalisme. De plus nous étudierons les différentes opérations possibles et nécessaires sur ces schémas et les formaliserons. Celles-ci pourront aller de la création de schémas à l'instanciation, en passant par l'ajout d'éléments dans le schémas, la fusion d'éléments, la modification à différentes échelles... Afin de garantir une évolutivité importante pour les schémas, un grand nombre d'opérations seront à modéliser.

Pour compléter cet axe de recherche, nous effectuerons une étude sur les interfaces nécessaires pour manipuler ces schémas dans un cadre général et nous l'appliquerons au projet Advène. Nous nous baserons sur les principes des interfaces évolutives afin de pouvoir faire évoluer celles-ci en même temps que les schémas. Il sera nécessaire de définir des interfaces de création de schémas, mais aussi de visualisation de ceux-ci. De plus, il sera sans doute intéressant de pouvoir visualiser les schémas en parallèle avec les annotations de manière à pouvoir avoir une vue globale et faciliter l'évolution de ceux-ci.

Enfin, l'utilisation de la lecture active dans le cadre des SIAV personnels sera à étudier. Nous avons évoqué dans ce rapport le fait que la lecture active permet une ré-indexation constante des documents par le lecteur. Cette évolution offerte par la lecture active permet au lecteur d'indexer ses documents selon ses besoins. Nous nous rapprocherons aussi du domaine du web sémantique, en essayant d'établir les interactions possibles entre nos schémas et les ontologies de manière à fournir un cadre consistant pour des traitements logiques.

L'objectif, à terme, est de pouvoir proposer un système complet permettant une lecture active aisée et efficace de documents audiovisuels, ce travail de recherche se poursuivra par l'étude du partage des schémas utilisés pour les annotations. En effet, l'intérêt de tels schémas, outre le fait de structurer l'ensemble des annotations et de permettre de rigidifier leur cadre, est de pouvoir les échanger en vue d'un travail collaboratif ou d'un partage d'opinion. Il s'agira donc d'étudier comment partager de tels schémas, comment adapter ou mettre en correspondance ces schémas, ou encore s'il est pertinent de mettre en place des serveurs de schémas d'annotation.

Pour accompagner ce travail théorique, des collaborations sur quelques projets sont envisagées. Tout d'abord, au sein de l'équipe d'enseignants-chercheurs « Interactions : Formes, Pratiques, Situations » du laboratoire ICAR (UMR 5191 du CNRS) étudiant les interactions entre personnes à l'aide de documents audiovisuels et de transcriptions de ceux-ci. Cette équipe effectue des transcriptions des gestes et paroles dans les vidéos pour étudier différents comportements interactionnels entre des personnes. Ils sont donc constamment en activité de lecture active de documents audiovisuels. Un

deuxième terrain de collaboration est envisagé avec cette équipe sur le thème de la gestion de corpus de documents audiovisuels, rejoignant par là notre notion de SIAV personnels. Enfin, deux projets pourront être abordés concernant l'élaboration d'un environnement auteur collaboratif. Le premier en vue d'analyses cinématographiques, le second dans le cadre de l'enseignement à distance via la plateforme « e-lycée ».

# Bibliographie

- [Aubert 03] O. Aubert, P-A. Champin & Y. Prié. *Instrumentation de documents audiovisuels : temporalisations et détemporalisation dans le projet Advène*. In Workshop "Temps et documents numériques", 2003.
- [Aubert 04a] O. Aubert, P-A. Champin & Y. Prié. *Advène : une plate-forme pour l'édition et la visualisation d'hypervidéos*. In Journée Modèles Documentaires de l'Audiovisuel (Semaine du Document Numérique 2004), 2004.
- [Aubert 04b] O. Aubert, P-A. Champin & Y. Prié. *The advène model for hypervideo document engineering*. In LIRIS UCBL Research Report RR-2004022, 2004.
- [Aubert 04c] O. Aubert & Y. Prié. *Documents audiovisuels instrumentés - Temporalités et détemporalisations dans les hypervidéos*. Document numérique, vol. numéro spécial "Temps et Document", pages 143–168, 2004.
- [Aubert 05a] O. Aubert & Y. Prié. *Advène : active reading through hypervideo*. In Proceedings of the sixteenth ACM conference on Hypertext and hypermedia, 2005.
- [Aubert 05b] O. Aubert & Y. Prié. *Des hypervidéos pour créer et échanger des analyses de documents audiovisuels*. In H2PTM'05, 2005.
- [Auffret 99] G. Auffret & Y. Prié. *Managing Full-indexed Audiovisual Documents : a new perspective for the Humanities*. Computer and Humanities, 1999.
- [autonomy Inc. 04] autonomy Inc. *Virage video System*, 2004.
- [Bachimont 99] B. Bachimont, G. Auffret, V. Brunie & T. Dechilly. *Représentation et structuration des connaissances pour les bibliothèques audiovisuelles*. Document numérique, 1999.
- [Bachimont 00] B. Bachimont & T. Dechilly. *Une ontologie pour éditer des schémas de description audiovisuels, extension pour l'inférence sur les descriptions*. In Pierre Tchounikine, editeur, Actes de IC'00, Toulouse, mai 2000.
- [Berners-Lee 89] T. Berners-Lee. *Information Managment : A proposal*, 1989.
- [Bertini 05] M. Bertini, A. DelBimbo & C. Torniai. *Automatic Video Annotation using Ontologies Extended with Visual Information*. In Proceedings of the 13th annual ACM international conference on Multimedia, 2005.

- [Bird 00] S. Bird, D. Day, J. Garofolo, J. Henderson, C. Laprun & M. Liberman. *ATLAS : A Flexible and Extensible Architecture for Linguistic Annotation*. In Proceedings of the Second International Language Resources and Evaluation Conference, 2000.
- [Bénel 03] A. Bénel. *Consultation assistée par ordinateur de la documentation en Sciences Humaines : Considérations épistémologiques, solutions opératoires et applications à l'archéologie*. PhD thesis, LIRIS, INSA de Lyon, Lyon, France, December 2003.
- [Boavida 04] M. Boavida, S. Cabaço & N. Correia. *Videozapper : A system for delivering personalized video content*. *Multimedia Tools and Applications Journal*, 2004.
- [Bocconi 05] S. Bocconi, F. Nack & L. Hardman. *VoxPopuli : a Tool for Automatically Generating Video Documentaries*. In Proceedings of the sixteenth ACM conference on Hypertext and hypermedia, 2005.
- [Bottoni 04] P. Bottoni, R. Civica, S. Leviardi, L. Orso, E. Panizzi & R. Trinchesi. *MADCOW : a Multimedia Digital Annotation System*. In Proceedings of the working conference on Advanced visual interfaces, 2004.
- [Bulterman. 03] D.C.A. Bulterman. *Using SMIL to Encode Interactive, Peer-Level Multimedia annotations*. In Proceedings of the 2003 ACM symposium on Document engineering, 2003.
- [Chambel 02] T. Chambel & N. Guimaraes. *Context perception in video-based hypermedia spaces*. In Proceedings of the thirteenth Conference on Hypertext and Hypermedia, 2002.
- [Christel 05] M.G. Christel & A.G. Hauptmann. *The Use and Utility of High-Level Semantic Features in Video Retrieval*. In Proceedings of the 4th International Conference on Image and Video Retrieval, 2005.
- [Costa 02] M. Costa, N. Correia & N. Guimarães. *Annotations as Multiple perspectives of video content*. In Proceedings of the tenth ACM international conference on Multimedia, 2002.
- [Edmondson 04] R. Edmondson. *Audiovisual archiving : philosophy and principles*, 2004.
- [Geurts 05] J. Geurts, J. van Ossenbruggen & L. Hardman. *Requirements for practical multimedia annotation*. In Workshop on Multimedia and the Semantic Web, 2005.
- [Graham 03] J. Graham, B. Erol, J. Hull & D-S. Lee. *The video paper multimedia playback system*. In Proceedings of the eleventh ACM international conference on Multimedia, 2003.
- [Grnbaek 96] K. Grnbaek & R. Trigg. *Toward a Dexter-based Model for Open Hypermedia : Unifying Embedded References and Link Objects*. In Seventh ACM Conference on Hypertext, pages 149–160, Washington DC., 1996.
- [Halasz 94] F. Halasz & M. Schwartz. *The Dexter hypertext reference model*. *Communications of the ACM*, vol. 37, no. 2, pages 30–39, 1994.

- [Hardman 94] L. Hardman, D. C. A. Bulterman & G. van Rossum. *The Amsterdam hypermedia model : adding time and context to the Dexter model*. Commun. ACM, vol. 37, no. 2, pages 50–62, 1994.
- [Hauptmann 02] A. Hauptmann, R. Baron, W. Lin, M. Chen, M. Derthick, M-C-R. Jin & R. Yan. *Video classification and retrieval with the informedia digital video library system*. In TREC, 2002.
- [Hochon 94] J.C. Hochon & F. Evrard. *Lecture professionnelle et gestion personnalisée de documents textuels*. ICO Québec, vol. 6, 1994.
- [internet Archive 04] internet Archive. *Internet Movies Archive*, 2004.
- [Isaac 04] A. Isaac & R. Troncy. *Designing and Using an Audio-Visual Description Core Ontology*. In Workshop on Core Ontologies in Ontology Engineering held in conjunction with the 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW'04), Whittlebury Hall, Northamptonshire, UK, October 2004.
- [Kim 04] S. Kim, M. Slater & E.J.Jr Whitehead. *WebDAV-based Hypertext Annotation and Trail System*. In Proceedings of the fifteenth ACM conference on Hypertext and hypermedia, 2004.
- [Kipp 01] M. Kipp. *Anvil - A Generic Annotation Tool for Multimodal Dialogue*, 2001.
- [Marshall 97] C. Marshall. *Annotation : from paper books to the digital library*. In Proceedings of the ACM Digital Libraries '97 Conference, Philadelphia, PA, 1997.
- [MIP+ 06] MIP+. *MIP+ - Association pour le développement du Multimédia Informatisé Pédagogique francophone*, 2006.
- [Nack 99a] F. Nack & A.T. Lindsay. *Everything You Wanted To Know About MPEG-7 : Part 1*. IEEE Multimedia, vol. July-September, pages 65–77, 1999.
- [Nack 99b] F. Nack & A.T. Lindsay. *Everything You Wanted To Know About MPEG-7 : Part 2*. IEEE Multimedia, vol. October-December, pages 64–73, 1999.
- [Nack 01] F. Nack & W. Putz. *Designing annotation before it's needed*. In In proceedings of the ninth ACM international conference on Multimedia, 2001.
- [Nanard 01] M. Nanard & J. Nanard. *Cumulating and sharing end users knowledge to improve video indexing in a video digital library*. In JCDL '01 : Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, pages 282–289, New York, NY, USA, 2001. ACM Press.
- [Nelson 99] T.H. Nelson. *Xanalogical structure, needed now more than ever : parallel documents, deep links to content, deep versioning, and deep re-use*. ACM Comput. Surv., vol. 31, no. 4es, page 33, 1999.
- [Obendorf 03] H. Obendorf. *Hypermedia creation : Simplifying annotation support for real-world-settings : a comparative study of active reading*. In Proceedings of the fourteenth ACM conference on Hypertext and hypermedia, 2003.

- [Pfeiffer 03] S. Pfeiffer, C. Parker & C. Schremmer. *Annodex : a simple architecture to enable hyperlinking, search and retrieval of time-continuous data on the web*. In Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, 2003.
- [Romero 03] L. Romero & N. Correia. *HyperReal : A Hypermedia Model for Mixed Reality*. In Proceedings of the fourteenth ACM Conference on Hypertext and Hypermedia, 2003.
- [Saarela 98] J. Saarela. *Video Contents Models based on RDF*. In W3C Workshop on Television and the Web, 1998.
- [Sawhney 96] N.N. Sawhney, D. Balcom & I.E. Smith. *Hypercafe : Narrative and aesthetic properties of Hypervideo*. In UK Conference on Hypertext, 1996.
- [Schilit 98] B.N. Schilit, G. Golovchinsky & M.N. Price. *Beyond paper : supporting active reading with free form digital ink annotations*. In Proceedings of the SIGCHI conference on Human factors in computing systems, 1998.
- [Shipman 03] F. Shipman, A. Grigensohn & L. Wilcox. *Image annotation and video summarization : Generation of interactive multi-level video summaries*. In Proceedings of the eleventh ACM international conference on Multimedia, 2003.
- [Shipman 05] F. Shipman, A. Girgensohn & L. Wilcox. *Hypervideo Expression : Experiences with Hyper-Hitchcock*. In Proceedings of the sixteenth ACM conference on Hypertext and hypermedia, 2005.
- [Shum 03] S.J. Buckingham Shum, P.A. Kirschner & C.S. Carr. *Visualizing Internetworked Argumentation*. In Visualizing Argumentation : Software Tools for Collaborative and Educational Sense-Making, 2003.
- [Slaughter 00] L. Slaughter, G. Marchionini & G. Geisler. *Openvideo : A framework for a test collection*. Journal of Network and Computer Applications, 2000.
- [Srikanth 05] M. Srikanth, J. Varner, M. Bowden & D. Moldovan. *Exploiting Ontologies for Automatic Image Annotation*. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 2005.
- [Stiegler 95] B. Stiegler. *Annotation, navigation, édition électronique : vers une géographie de la connaissance*. In Actes du séminaire "Hypermédiâs, Education et Formation", pages 27–37, 1995.
- [the M.I.P. 06] the M.I.P. *Elan : A linguistic multimedia annotator*, 2006.
- [Tran-Thuong 03] T. Tran-Thuong & C. Roisin. *Multimedia modeling using mpeg-7 for authoring multimedia integration*. In Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, 2003.
- [Uren 05] V. Uren, E. Motta, M. Dzbor & P. Cimiano. *Browsing for Information by Highlighting Automatically Generated Annotations : A User Study and Evaluation*. In Proceedings of the 3rd international conference on Knowledge capture., 2005.

- [W3C 06a] W3C. *XML Schema 1.1 Part 2 : Datatypes*, W3C Working Draft 17 February 2006, 2006.
- [W3C 06b] W3C. *XML Schema 1.1 Part 2 : Structures*, W3C Working Draft 17 February 2006, 2006.
- [Waller 03] R. Waller. *Functionality in Digital Annotation : Imitating and supporting real-world annotation*. Ariadne <http://www.ariadne.ac.uk>, vol. 35, 2003.
- [Whitehead. 02] J. Whitehead. *Uniform comparison of data models using containment modeling*. In Proceedings of the 13th ACM conference on Hypertext and hypermedia, 2002.
- [Zhou 05] T.T. Zhou, T. Gedeon & J.S. Jin. *Automatic Generating Detail-on-Demand Hypervideo Using MPEG-7 and SMIL*. In Proceedings of the 13th annual ACM international conference on Multimedia, 2005.

## Annexe A

# Outils d'annotation

Dans cette annexe, nous présenterons les outils d'annotation que nous avons évoqué dans le chapitre 2. Nous commencerons par présenter le logiciel CLAN puis ELAN [the M.I.P. 06] du Max Plank Institute et enfin ANVIL [Kipp 01] du Centre de Recherche Allemand pour l'Intelligence Artificielle (DFKI), en termes de fonctionnalités offertes et d'interface.

### A.1 CLAN

CLAN (Computerized language analysis) est un logiciel développé au sein de l'université de Carnegie Mellon. Cet outil permet d'effectuer l'analyse de données transcrites dans le format CHILDES, ainsi que d'autres fonctions telles que de la recherche de termes, de l'analyse d'interactions, ou du calcul de fréquence d'apparition d'un mot. Ce logiciel facilite la transcription et l'analyse d'interactions verbales et non verbales sous forme de liste de segments textuels (chaque ligne/paragraphe correspond par exemple à une prise de tour) qui peuvent être alignés avec un signal audio ou vidéo. CLAN facilite l'insertion de la représentation de catégories syntaxiques, morphologiques et phonétiques sous forme d'annotations linéaires (segmentation des prises de tours et annotation dans des lignes parallèles). CLAN permet de lier une transcription à une vidéo et il est alors possible de relire uniquement certains passages, en se basant sur les lignes de la transcription. L'application offre une fenêtre d'édition permettant de saisir sa transcription en respectant le format nécessaire. Le logiciel offre aussi toute sorte de calculs et d'analyses exécutables sur la transcription saisie. Il est cependant uniquement possible d'imprimer le texte sans aucun alignement temporel sur les données audio ou vidéo. Aucun paramétrage ni aucun export dans des formats tels que pdf ne sont possibles actuellement.

La figure A.1 montre un exemple de transcription du discours entre une mère et son enfant.

```
@Begin
@Participants: CHI Sarah Target_child, MOT Carol Mother
*MOT: sure go ahead [c].
%cod: $A
%spa: $nia:gi
*CHI: can I [c] can I really [c].
%cod: $A $D. $B.
%spa: $nia:fp $npp:yq.
%sit: $ext $why. $mor
*MOT: you do [c] or you don't [c].
%cod: $B $C.
%spa: $npp:pa
*MOT: that's it [c].
%cod: $C
%spa: $nia:pa
@End
```

FIG. A.1 – Exemple de transcription d'une conversation dans CLAN

## A.2 ELAN

ELAN (EUDICO Linguistic Annotator) est un logiciel d'annotation permettant de créer, éditer et visualiser des annotations pour des documents audio ou vidéo. Il a été développé au sein de l'Institut Max Plank. Elan est particulièrement adapté à l'analyse du langage, des signes et des gestes. Il supporte l'affichage simultané des documents audiovisuels et de leurs annotations ou la lecture d'un document audio et de ses annotations, les liens entre les annotations et les documents, les liens entre plusieurs annotations, la possibilité de définir des « tiers » pour structurer les annotations, l'export sous forme de fichiers textuels, et la recherche dans les annotations.

Elan permet aussi d'importer ou d'exporter des transcriptions au format CHAT évoqué dans la partie précédente. Il est aussi possible d'exporter des clips au format SMIL.

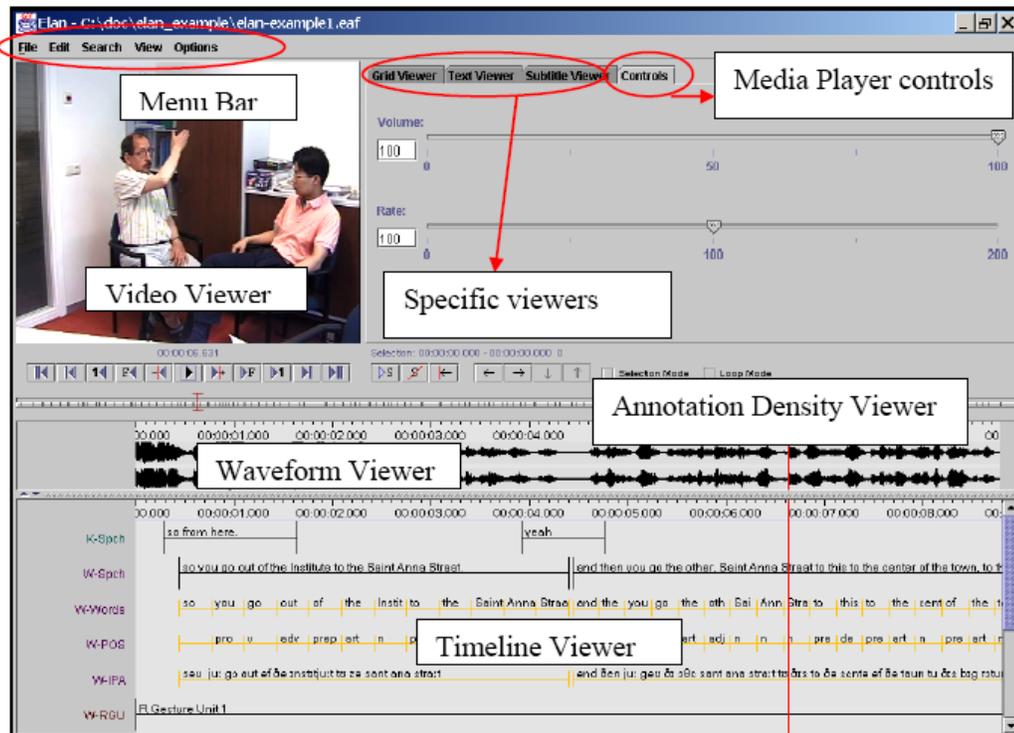


FIG. A.2 – Fenêtre principale du logiciel Elan

La fenêtre principale d'Elan est présentée dans la figure A.2. Elle se compose de différentes parties, toutes synchronisées entre elles.

La première partie est la fenêtre de visualisation de la vidéo. Celle-ci permet comme tout lecteur multimédia de lire la vidéo et de contrôler celle-ci. Une barre de temps permet aussi de naviguer directement dans la vidéo.

Un deuxième espace est utilisé par une représentation de l'enveloppe sonore du document. Celle-ci permet au lecteur de se repérer dans la vidéo et de sauter directement aux parties potentiellement intéressantes, dans le cas d'une analyse du langage par exemple.

L'espace supérieur droit de la fenêtre peut être utilisé pour différentes fonctions, sélectionnables via un système d'onglets. Ainsi on pourra y trouver les contrôles du lecteur multimédia, ou un affichage des sous-titres de la vidéo, sélectionnés suivant leur « tier », ou une grille de calcul type grille Microsoft Excel regroupant les annotations, ou encore une transcription bout-à-bout du texte de la vidéo.

Enfin, l'espace inférieur de la fenêtre est occupé par la représentation des annotations suivant une ligne de temps et organisées par « Tier ».

Le logiciel est très complet et offre beaucoup de fonctionnalités dans chacun des modes de représentation et d'analyse qu'il propose.

## A.3 ANVIL

Anvil est un outil développé au Centre de Recherche Allemand pour l'Intelligence Artificielle (DFKI) pour l'annotation de contenu audiovisuel. Les annotations typées (regroupées par pistes) s'insèrent dans le document par référencement temporel. Chaque élément d'annotation possède une référence temporelle de début, une de fin et un contenu, et appartient à une « piste », type d'annotations. Les valeurs des annotations peuvent être des chaînes de caractère, des booléens, des ensembles de valeurs, ou des liens. Chaque « piste » a son propre schéma d'annotation, et lors de la création d'annotations, la fenêtre d'édition est spécialisée en fonction de la piste de l'annotation. Chaque annotation peut aussi être accompagnée d'un commentaire libre, qui n'est pas pris en compte dans le schéma d'annotation. Les schémas sont prédéfinis dans Anvil, pour certains types d'annotations telles que les gestes, les positions, les comportements. La transcription de langage n'est pas intégrée, cependant, il est possible d'importer des transcriptions de certains logiciels spécialisés.

Anvil permet l'export sous forme textuelle de chaque piste d'annotations. Il permet aussi l'export de la même piste de différents fichiers d'un projet sous forme de table d'éléments écrits les uns sous les autres. Enfin, Anvil peut générer la documentation relative à un projet.

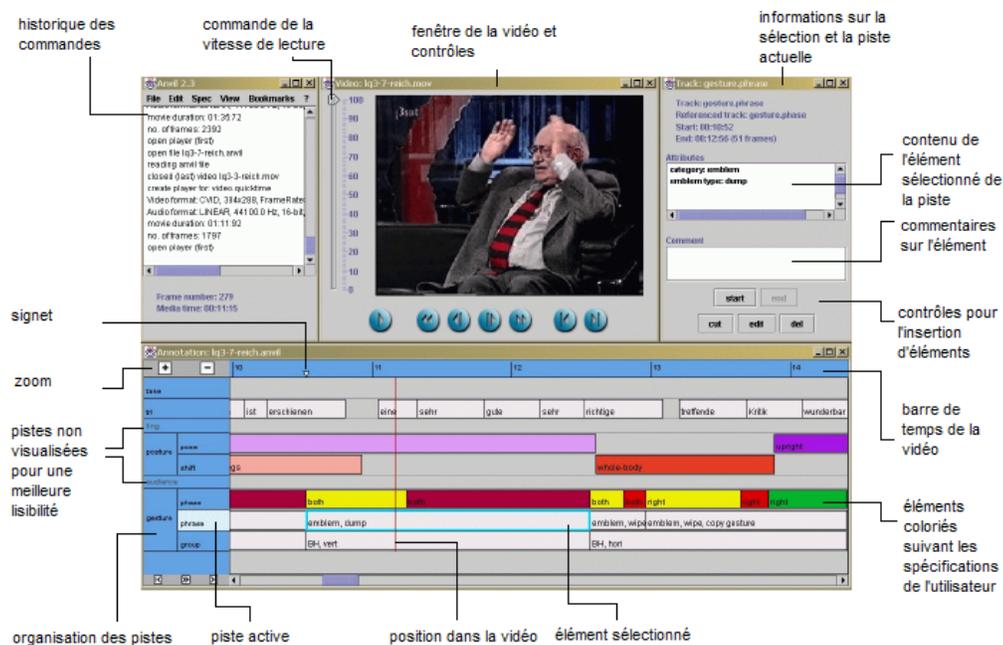


FIG. A.3 – Fenêtre principale du logiciel Anvil

La figure A.3 montre la fenêtre principale du logiciel Anvil. On peut y voir, au centre, le lecteur vidéo avec la barre de contrôle en dessous. Il est possible de contrôler la vitesse de lecture de la vidéo de façon intuitive avec la barre de vitesse à gauche de la vidéo. L'espace tout à gauche présente un historique des dernières actions effectuées. Sur la droite sont disponibles les informations relatives à l'élément d'annotation sélectionné et sur la piste à laquelle il appartient.

Enfin, la partie inférieure est occupée par une barre de temps, selon laquelle sont disposées les annotations, organisées suivant des lignes représentant chacune une piste. L'organisation est la même que celle de la ligne de temps d'Elan ou que celle d'Advene, avec en plus la possibilité de cacher les pistes inintéressantes.

## Annexe B

# MPEG-7

Les travaux sur ce format de description de contenus multimédias ont démarré en 1996 au sein du Moving Picture Expert Group (MPEG). Les standards MPEG précédents (MPEG-1, MPEG-2, MPEG-4) étant clairement orientés sur la façon de coder les contenus audiovisuels, le but de ce nouveau format est de standardiser un ensemble de descripteurs de documents audiovisuels et de structures de ces descripteurs et de leur relations (schémas de description). Il définit aussi un langage de définition de ces schémas de description, le « Description Definition Language » (DDL). Il est donc possible d'indexer et de rechercher du contenu audiovisuel auquel sont associées des descriptions au format MPEG-7 [Nack 99a, Nack 99b], l'idée de ce format étant de pouvoir rendre le contenu audiovisuel autant accessible et aussi facilement indexable que du texte.

Le principal but de ce format est donc de fournir des méthodes et des outils permettant de décrire des contenus multimédias de manière précise. Pour cela, MPEG-7 fournit quatre grandes classes de descripteurs, représentant quatre aspects fondamentaux des documents audiovisuels :

- description du média
- description physique
- description perceptuelle
- description par transcription

Le premier type de description à fournir est celui concernant le média concerné. Cette classe permet de décrire le type de source pour la vidéo, analogique ou numérique, les limites des plans de la vidéo, la distance focale de la caméra, *etc.* Ce sont des descripteurs de bas niveau de la vidéo.

La deuxième classe de descripteur regroupe ce qui concerne la structure physique de la vidéo. Ce sont encore des descripteurs de bas niveau, qui ne correspondent pas à des perceptions humaines. Il s'agit de descripteurs non ambigus, généralement calculables à partir du document source brut, comme par exemple la fréquence.

La classe de description perceptuelle définit des caractéristiques telles que la couleur, la texture de l'image, le son présent, *etc.*

Enfin, la classe transcription permet de présenter une reconstruction du dialogue de la vidéo, ou des notes de la musique.

Au delà de ces classes de description, une classe décrivant l'architecture existe dans ce format. Elle permet de décrire les éléments des autres classes de description et de les structurer.

Pour terminer, une classe de description annotative permet de prendre en compte

les analyses et annotations humaines concernant le document.

MPEG-7 offre donc un ensemble de schémas permettant de décrire et d'annoter les documents audiovisuels. De plus, ce format offre des possibilités de créer ses propres schémas de descriptions grâce à un langage de définition de description inclus dans le format.

# Glossaire

Annotation : inscription sur un document ou liée à celui-ci, décrivant, mettant en évidence ou ajoutant de l'information à une partie ou à la totalité d'un document.

Document audiovisuel : 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.

Document audiovisuel annoté : document audiovisuel auquel on a ajouté une structure d'annotations liée spatio-temporellement au flux de ce document.

Hyperlien : lien navigable entre deux composants d'un hyperdocument.

Hypermédia : ensemble de ressources numériques intégrées sur un même support, constituées de données de nature différente (texte, son, image et fichier) avec des liaisons multiples qui donnent, à la consultation de ces ressources, une dimension réticulaire.

Hypervidéo : vue d'un document audiovisuel annoté, permettant d'une part, d'enrichir le document audiovisuel par ses annotations, et d'autre part de contrôler directement le flux audiovisuel à partir de celles-ci.

Lecture active : activité composée d'une lecture critique et d'une production par rapport à un document.

Orientation de la lecture : Raison pour laquelle le lecteur lit.

Plan : prise de vue sans interruption.

Scène : composition de plans.

SIAB : Système d'information audiovisuels, système fournissant des moyens d'indexer, de rechercher et d'utiliser des documents audiovisuels.

Système hypervidéo : système permettant de construire, modifier, visualiser des hypervidéos à des degrés variables.

Vue d'un document : manière de visualiser un document.