

# L'indexation critique de documents audiovisuels

Martin Jean-Pascal\*

\*LIMSI/CNRS, Université Paris Sud, jpmartin@limsi.fr

## *Résumé*

*Les approches classiques d'indexation Documents AudioVisuels (DAV) n'intégrant pas de modèle de lecture dans le modèle de document, nous présentons l'indexation critique qui repose sur l'analyse critique de document. Trois modèles d'analyse critique sont présentés et illustrés. Ils sont ensuite formalisés afin de pouvoir être intégrés comme extension de la norme MPEG-7.*

***Mots-clés :** indexation critique, indexation par le contenu, systèmes d'indexation audiovisuel, analyse critique.*

## *Abstract*

*The traditional approaches for audiovisual content indexing don't integrate a model of reading in the document's model. We present the critical indexing which rests on the critical analysis of documents. Three models of analysis criticizes are presented and illustrated. They are finally formalized in order to be able to be integrated as extensions of standard MPEG-7.*

***Keywords :** critical indexing, audiovisual content indexing, critical analysis.*

## 1. INTRODUCTION

### 1.1. La modélisation de documents audiovisuels

L'indexation de Documents AudioVisuels (DAV) est un enjeu majeur pour de nombreux acteurs et utilisateurs de documents audiovisuels, tels que les journalistes, les documentaristes, les analystes de films, les historiens du cinéma, les enseignants en cinématographie... La mise en

œuvre de cette indexation est aujourd'hui bien souvent considérée comme insatisfaisante d'un point de vue qualitatif. Nombre d'acteurs du domaine de la télédiffusion emploient des documentalistes pour l'indexation manuelle et textuelle de leurs fonds, ce qui est coûteux en temps et trop informel pour pouvoir être totalement exploité par un système d'information.

Plusieurs voies sont explorées aujourd'hui afin de palier à cette double contrainte : méthodes numériques d'analyse du signal visant à extraire des descripteurs de haut niveau, modèles du document formalisant au mieux des aspects sémantiques, utilisation de connaissances externes et de raisonnements, développement de méthodologies de formalisation de domaine pour leur modélisation *ad hoc*, etc.

Nous reprenons dans cet article le constat que l'homme ne peut pas être évacué du processus d'indexation du DAV car il est le seul à même de comprendre, d'interpréter, ou de ressentir le message latent présenté par le document [14]. Ainsi, au-delà de l'identification des formes présentées à l'image, l'indexation doit décrire le contenu, qu'il soit discursif, narratif ou énonciatif, qu'il fasse référence à des éléments externes au film ou qu'il exploite les connaissances du lecteur. C'est pourquoi l'indexation objective du DAV n'existe pas, car l'indexation dépend nécessairement d'un discours porté sur le document. Mais loin de considérer que cette difficulté est rédhibitoire, nous proposons de porter ce discours au premier plan de l'indexation : il faut envisager l'indexation du contenu par sa représentation, représentation produite par un discours riche de sens. Nous en déduisons qu'il faut élaborer des modèles de DAV incluant une modélisation de la *lecture* du DAV.

Plusieurs disciplines étudient l'analyse documentaire et fournissent des grilles, des méthodes qui, si elles sont supportées par un système d'information, permettent la mise en place d'une *indexation critique* des documents audiovisuels. Étant donné que ces analyses peuvent servir d'index dans un système documentaire, nous présentons par la suite plusieurs formes d'analyses (l'analyse symbolique, l'analyse sémiotique et l'analyse componentielle) et nous réfléchissons à leurs pertinences sur le plan de l'usage en documentation et sur celui de leur faisabilité technique. L'enjeu de l'article est d'envisager l'exploitation de ces approches critiques. Tout en sachant qu'elles fournissent à la fois une proposition de modèle statique et une méthodologie dynamique de construction, nous n'abordons que la modélisation statique. La question de l'opérationnalisation de l'indexation est laissée de côté, celle-ci restant à ce stade de la réflexion à la charge de l'opérateur d'indexation.

## 1.2. La rhétorique audiovisuelle

Que ce soit l'historien fouillant les archives de l'INA, que ce soit le journaliste politique cherchant à illustrer un sujet engagé, ou que ce soit le critique de cinéma cherchant une référence, tous ont besoin de données audiovisuelles qu'ils considèrent à cet instant comme remarquables. Ils cherchent des séquences très courtes mais très informatives, des « images chocs », des images rares ou des images à forte valeur symbolique (un drapeau déchiré, un tank enlisé).

Nous notons qu'un document vidéo destiné à la diffusion reflète une part de subjectivité provenant en parti de l'auteur, l'auteur étant lui même pris à parti par les intérêts sociétaux ou économiques de la radiodiffusion<sup>1</sup>. Le documentaire et le film ont une forte valeur démonstrative et tous deux cherchent à produire un effet au-delà de l'information brute donnée au spectateur – même dans un divertissement – comme si le document audiovisuel avait une intention propre révélée par son dispositif argumentatif. Cette intention est la conséquence de l'intention de l'auteur, ou d'un groupe d'auteur (telle que l'école de documentaires de Riga fondée par Herz Frank<sup>2</sup>). Bien que de nombreuses vidéos de journaux télévisés semblent neutres et strictement informatives, ces vidéos contiennent presque toujours un message latent : une vidéo montrant des tirs de missiles filmés du haut d'un hôtel de Bagdad montre des images de guerre sans truquage. Mais le dispositif journalistique mettant en scène ces images souligne certainement à *quel point* la guerre fait rage.

## 1.3. Les modèles du document audiovisuel

Cependant les aspects que nous venons d'évoquer ne sont pas abordés de front par la communauté de recherche d'information audiovisuelle dont la démarche générale est majoritairement *bottom-up* (allant du signal au concept), et procède soit par une approche de décomposition du signal, soit par une approche de description orientée objet, soit par une approche de stratification. La décomposition du signal procède par segmentation

---

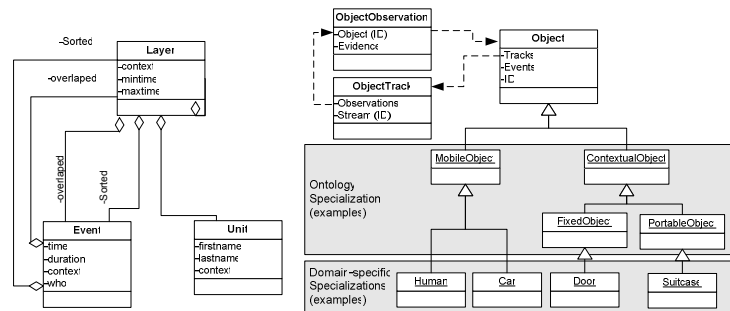
<sup>1</sup> Pierre Bourdieu a appliqué sa méthode sociologique à la télévision, dénonçant la « censure invisible » imposée par l'audimat et les contraintes du marché et ayant pour conséquence que « *ce qui aurait pu devenir un extraordinaire instrument de démocratie directe ne se convertisse pas en instrument d'oppression symbolique* » Pierre Bourdieu (1996). *Sur la télévision*. Transcription de l'enregistrement de deux émissions réalisées en 1996 dans le cadre d'une série de cours du Collège de France (CNRS Audio-Visuel).

<sup>2</sup> Dans son dernier documentaire, « Flashback », Herz Franz s'interroge sur les limites du documentaire et décide de montrer la fin de sa propre vie repoussée par des chirurgiens et cardiologues.

successive du flux, entre segments d'images et segments sonores, puis entre image, icône et texte pour les premiers et entre bruitage, musique et parole pour les segments sonores, etc, la décomposition continuant tant que des procédés d'analyse numérique peuvent opérer des distinctions significatives.

La démarche de modélisation orientée objet vise à identifier spatialement et temporellement tous les éléments présents à l'image tout en les séparant du fond. Elle est utilisée dans la norme MPEG-4 [11].

L'approche dite de stratification part de la description sommaire du signal, par-dessus laquelle des strates de description sont déposées [1]. Ces strates s'empilent pour atteindre progressivement des descriptions de plus haut niveau sémantique<sup>3</sup>. Les strates peuvent être utilisées pour représenter des événements comme pour le format AEDI défini par l'INA (dont nous proposons une représentation UML sur la FIG. 1, à gauche). Les strates peuvent ne pas être contraintes par une sémantique *a priori*, celle-ci pouvant être spécifiée de façon *ad hoc* par annotation [17, 22]. Plus récemment, le standard VERL pour *Video Event Representation Language* a été proposé [3, 7]. Ce modèle permet l'annotation par événements auxquels des connaissances sont associées (FIG. 1, à droite).



**FIG. 1** - A gauche, diagramme de classes UML du format AEDI permettant la définition de strates événementielles. A droite, ontologies de domaine associées aux objets pour le format VEML/VERL

Le standard MPEG-7 fait la fusion entre ces approches en permettant la description objet, l'annotation et la stratification. Cependant, bien que cette norme permette la description du contenu conceptuel, il faut pour cela l'étendre en ajoutant de nouveaux schèmes de description sémantique

<sup>3</sup> Nous ne critiquons bien évidemment pas cette démarche qui n'empêche d'aucune façon une description commençant par la constitution d'une strate de très haut niveau sémantique.

(SDS). En fait, MPEG-7 ne prévoit pas directement l'utilisation de connaissances formelles, mais [10, 20] ont montré que la norme pouvait être étendue avec des langages de représentation de connaissance.

## 2. INTERPRÉTATION AUDIOVISUELLE

Aucune des approches présentées n'intègre explicitement un modèle de lecture du DAV. Pourtant, Christian Metz après avoir proposé une syntagmatique du cinéma conclut à son propos que ce « n'est pas une langue parce qu'il contrevient à trois caractères importants du fait linguistique : une langue est un *système* de *signes* destinés à l'*intercommunication*. Trois éléments de définition. Or le cinéma, comme les arts et parce qu'il en est un, est une 'communication' à sens unique : c'est en fait un *moyen d'expression* beaucoup plus que de communication » [4]. [15] ajoute que « Non seulement un film ne produit pas de sens en lui-même, mais tout ce qu'il peut faire, c'est bloquer un certain nombre d'investissements signifiants » [5]. Ainsi le document AV ne se comprend vraiment que par l'intermédiaire d'un lecteur procédant à une interprétation. Le document n'a aucun sens patent, car c'est la lecture qui lui donne sens.

Afin de rendre l'expression du contenu tangible pour un système d'information, nous proposons de réaliser une indexation que nous qualifions de critique puisqu'elle vise à indexer le discours critique porté sur le contenu signifiant d'un document d'après un lecteur. Ces discours critiques peuvent revêtir de nombreuses formes. Nous en illustrons trois centrées sur l'analyse du signe cinématographique : l'analyse symbolique, l'analyse sémiotico-narrative et l'analyse componentielle. Bien que toutes trois proviennent de l'analyse du texte, elles s'appuient sur le paradigme sémiologique dont le champ est plus large que celui de la linguistique. Comme le souligne Peirce, les signes sont « la matière première grâce à quoi tout être [...] communique avec d'autres êtres sur la base d'un quelconque système de communication » [6]. Nous souhaitons appliquer ces analyses au document audiovisuel. Pour chacune de ces méthodes, nous fournissons une représentation des concepts et de leurs relations sous la forme d'un diagramme de classe UML, aisé à lire, et fournissant un modèle utile en prévision d'une d'instrumentation.

### 2.1. Le symbole et son analyse

Un symbole est un signe particulier doté d'une signification abstraite : c'est un signe dyadique puisqu'il est constitué d'un signifiant, d'un signifié, et ne dispose pas de référent. Un exemple classique de symbole est la colombe qui représente la paix. Pour [21], le symbole est l'usage d'une unité signifiée pour faire penser à une autre unité signifiée. Selon les auteurs, la relation symbolique peut être analogique ou arbitraire. En corollaire, le champ du symbole ne peut pas être limité. Par exemple, dans les conduites culturelles, tout (la mode, la publicité, l'urbanisme, l'architecture, le discours politique...) est porteur de symbolisme [12], ce qui inclus naturellement l'audiovisuel.

L'analyse symbolique vise à expliciter les symboles, en établissant des liens entre des signaux (qu'ils soient textuels, réels ou audiovisuels) et des significations abstraites et normées. Ces liens peuvent être vus comme les segments d'un *parcours interprétatif* défini par [18] à propos du texte comme « une suite d'opérations permettant d'assigner un ou plusieurs sens à un passage ou à un texte ».

Louis Herbert propose d'explicitier ces parcours interprétatifs par des opérations de réécriture « par laquelle on réécrit un ou plusieurs signes, signifiants, signifiés en un ou plusieurs signes, signifiants ou signifiés différents » [9]. Il cite pour exemple la phrase « L'aigle a terrassé l'Ours » offrant la possibilité de réécrire « L'aigle » en /Etats-Unis/ et « l'Ours » en /URSS/, grâce à la connexion symbolique qui apparaît lors de la lecture<sup>4</sup>.

/colombe/ → /paix/

« L'Aigle » → /Etats-Unis/ ou « l'Ours » → /URSS/

Appliquée à l'audiovisuel, la réécriture permet de proposer des lectures différentes ou plus abouties à partir du signal. Par exemple, un plan montrant un drapeau en berne peut être interprété entre autres choses comme une défaite, comme un temps calme, ou un climat apaisé.



→ "Drapeau en berne" → /défaite/

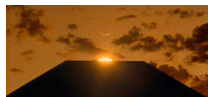
Dans cet exemple, la réécriture du signe audiovisuel ("drapeau en berne") en un signifié ("défaite") est faite par lexicalisation lors de la lecture de l'image.

---

<sup>4</sup> Nous utilisons ici une notation qui distingue "signe", signifiant, 'signifié', /sème/, //classe sémantique//, → [réécriture].

## 2.2. De la compréhension à l'interprétation

Du point de vue de la psychologie cognitive, « comprendre, c'est construire des interprétations » [19]. On pourrait dire de l'interprétation qu'il s'agit d'une compréhension limitée, faisant moins consensus, plus singulière, telle que celle que nous proposons à partir d'un plan du film *2001: A Space Odyssey* de Stanley Kubrick tourné en 1968 :



→ /monolithe/ → /Vie/

D'autres interprétations peuvent être faites à partir de cette image :



→ /monolithe/ → /Évolution/

→ /monolithe/ → /Technologie/

Ces interprétations peuvent être considérées comme complémentaires. En ce cas, le monolithe est à la fois signe d'évolution/ de Technologie/ et de /vie/. De plus, ces trois signes observés conjointement peuvent mener à l'interprétation suivante :

/Monolithe/ → /Vie/ + /Évolution/ + /Technologie/ → /Humanité/

Nous observons qu'il est possible de construire des signes parallèlement les uns aux autres, deux chaînes de signes ayant le même signifiant initial et le même référent terminal. De l'un à l'autre, des parcours d'interprétations différents peuvent être établis. Ainsi, différentes granularités pour l'interprétation peuvent coexister sans se contredire, et des strates de réécriture symbolique se forment.

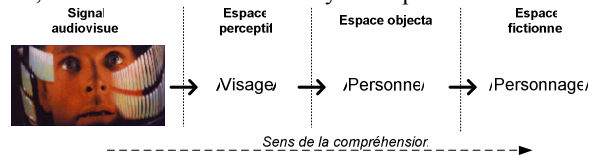


FIG. 2 - Exemple d'interprétation à trois niveaux d'imbrication de signes à partir de l'image de l'acteur principal

La chaîne de signes représentée sur la Fig. 2 permet l'explicitation de la relation entre le signifiant d'un personnage (l'acteur, et plus précisément la représentation audiovisuelle de l'acteur), le concept de personnage fictif et le personnage de Dave Bowman dans « *2001: A Space Odyssey* ».

L'analyse de la Fig. 2 est construite séquentiellement, tout d'abord par l'identification des percepts (espace perceptif), suivie de leur reconnaissance dans l'espace objectal, et suivie d'une mise en signification dans l'espace fictionnel.

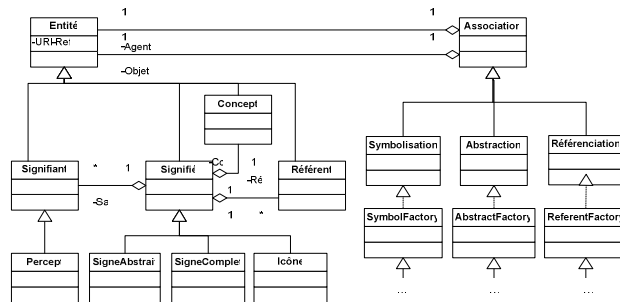


FIG. 3 - Diagramme des classes UML permettant l'instanciation de signes et de relations de réécritures de signes. Les relations d'abstraction et de référencement sont possible en plus de la relation de symbolisation.

Nous avons proposé dans [14] une extension de ce modèle d'interprétation symbolique, prenant la forme de réseau de signes formels. Ceux-ci permettent la réécriture sémiotique du signal audiovisuel. Les différents types de signes et les différentes associations possibles entre signes sont présentés dans la FIG. 3. Pour chaque signe, les références intensionnelles (Concept) et extensionnelles (Référent) peuvent être définies.

### 2.3. Analyse componentielle

Une autre forme d'analyse est couramment appliquée au texte. Il s'agit de l'analyse componentielle (ou sémique [8, 16]), qui tente de rendre compte de l'organisation structurale des contenus lexicaux, en identifiant et en décomposant les signifiés. L'identification des signifiés se fait par comparaison des signifiés entre eux afin de trouver les éléments communs et les éléments distinctifs. Ces comparaisons permettent la décomposition des signifiés (sémèmes) et leur organisation structurale selon qu'ils partagent ou non des valeurs (ou sèmes). [2] rapportent que l'extraction des relations taxinomiques entre signifiés n'est pas toujours explicite dans un corpus. Leur manifestation doit être extraite par un travail particulier.

Bien que l'analyse sémique soit appliquée au langage, elle peut se transposer à l'analyse de tout ce qui fait sens et donc au flux audiovisuel.



Les discussions portant sur les relations entre lexèmes ou phonèmes et morphèmes, ainsi que celles entre morphèmes et sémèmes doivent être reconduites pour l'image et le son.

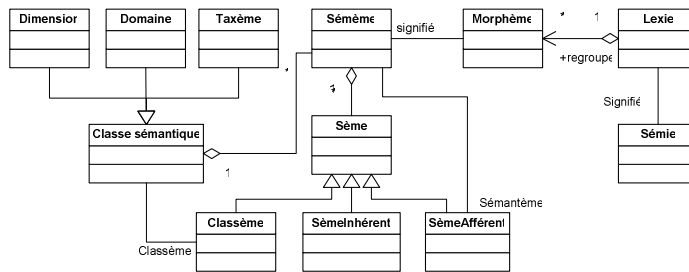


FIG. 4 - Unités manipulées en sémantique componentielle (Diagramme UML)

Nous avons représenté les relations entre les classes conceptuelles primitives de la sémantique componentielle sur la FIG. 4.

## 2.4. L'analyse narratologique

L'analyse narratologique (ou sémiotique narrative) étudie le document selon trois niveaux.

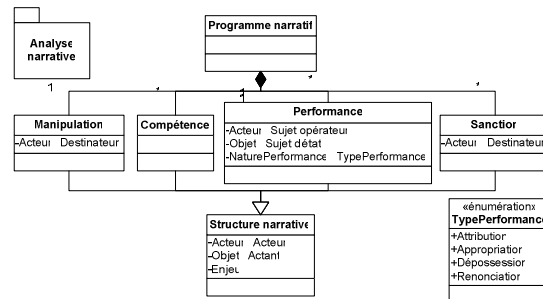


FIG. 5 - Classes identifiant les phases la narration et leurs propriétés (diagramme UML).

Le niveau discursif vise à identifier les figures, à les classer et à étudier leur agencement (trois pôles : acteurs, temps et lieux à organiser thématiquement). Le niveau narratif identifie des états (introduit des notions de sujet et d'objet, où de rôle) et leurs transformations selon quatre phases : manipulation, compétence, performance et sanction (que nous traduisons en diagramme de classe UML sur la FIG. 5).

L'analyse narratologique est particulièrement adaptée pour l'analyse de la diégèse (la diégèse est l'univers fictionnel porté par le récit), de sa mise en discours et de la forme prise par son énonciation.

### 3. MODÈLE D'INDEXATION CRITIQUE

Les trois modèles d'analyse présentés peuvent soit servir directement de langage d'annotation structurée du flux audiovisuel, soit venir en complément d'une modélisation MPEG-7. Dans le premier cas, une traduction des modèles d'analyse critique en schémas XML permet l'annotation structurée du flux audiovisuel par l'intermédiaire de logiciels d'annotation existants.

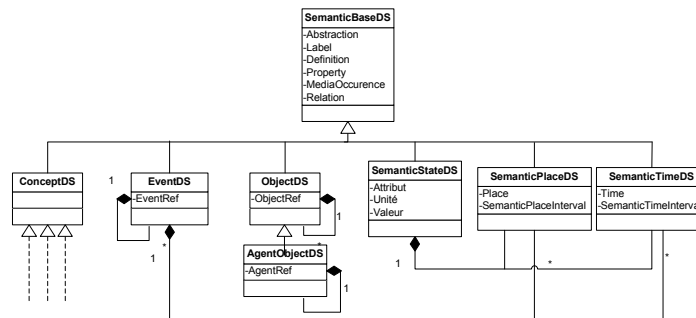


FIG. 6 - Possibilité d'extension de la norme MPEG-7 par héritage des classes du schéma de description SemanticBaseDS (Diagramme UML).

Les possibilités d'extension de la sémantique de la norme MPEG-7 sont représentées sur la FIG. 6 (par les flèches en pointillés). Ces possibilités peuvent être utilisées en faisant hériter de nouvelles classes à la classe ConceptDS. Ces classes peuvent être celles représentées sur les diagrammes d'analyse des figures FIG. 3, FIG. 4 et FIG. 5. Cela permet d'intégrer des éléments d'analyse critique (symbolique, componentielle ou narratologique) au modèle standard MPEG-7.

Bien que nous ne présentons pas ici de solutions techniques, nous soulignons la possibilité d'assister l'utilisateur dans la construction de son analyse, notamment en définissant des schémas de réécriture valides dans un contexte identifié. Cette approche a été présentée dans [14].

## 4. CONCLUSION

Pour les analystes, le document audiovisuel « résiste » à son analyse automatique, et de nombreuses voies doivent être explorées pour faire « parler » le document. Ce dernier sert de « vérité terrain » pour formuler et tester des hypothèses sur son message latent.

Dans cet article, nous avons envisagé d'utiliser l'analyse critique comme élément final d'une indexation poussée à son terme : l'« indexation critique ». Cette forme d'indexation s'appuie sur des approches sémiologiques offrant des possibilités quand à la représentation des contenus par l'ajout d'informations nées de la lecture du document. Nous avons en conséquence fourni des modèles sémiotiques primitifs pouvant servir de support à l'analyse critique. Ces modèles permettent de prendre en compte le caractère individuel de l'analyse, ce qui est une manière de réintroduire des aspects pragmatiques lors de la modélisation. En effet, Morris définit la pragmatique comme l'« étude des relations que les signes entretiennent avec leurs utilisateurs » [13].

Comme ces informations sont autant de points d'accès pour des systèmes d'indexation-recherche, elles potentialisent de nouveaux usages. En premier lieu, l'analyse sémiotique peut servir de méthode d'évaluation de la compréhensibilité, de l'intelligibilité, de la clarté du discours, et favoriser la construction de documents pédagogiques adressés aux étudiants. Elle est aussi utile pour vérifier la qualité illustrative de vidéos de savoir-faire d'entreprise. Enfin, la construction d'un fond documentaire de films « classiques » auxquels des analyses critiques formalisées doit naturellement être envisagée.

## 5. RÉFÉRENCES

- [1] Aguiere-Smith, T.G. and G. Davenport. *The stratification system: A design environment for random access video*. in *Proc. 3rd International Workshop on Network and Operating System Support for Digital Audio and Video*. La Jolla, CA, USA, 1992.
- [2] Bachimont, B., V. Malaisé, and P. Zweigenbaum. *Vers une combinaison de méthodologies pour la structuration de termes en corpus: premier pas vers des ontologies dédiées à l'indexation de documents audiovisuels*. in *Actes de la conférence "ISKO-France: L'organisation des connaissances: approches conceptuelles"*. Grenoble, 2003.
- [3] Bolles, B. and R. Nevatia, *A Hierarchical Video Event Ontology in OWL*. ARDA Challenge Workshop Report, 2004.

- [4] Chamming's, L. and P. Escande, *Pour une description sémiotique des objets audiovisuels*, Inter-Ateliers, INA, 2004.
- [5] Claude, R., *Analyse et critique de film*, 2006.
- [6] Eco, U., *Le signe*. Le Livre de poche, ed. Labor: Payot, 1980.
- [7] François, A.R.J., et al., *VERL: An Ontology Framework for Representing and Annotating Video Events*. IEEE Computer. **12**(14): p. 76-86, 2005.
- [8] Greimas, A.J., *Sémantique structurale: recherche de méthode*, Paris: Larousse, réédition Presses Universitaires de France, 1986, 1966.
- [9] Hébert, L., *La sémantique interprétative en résumé*. revue-texto, 2002.
- [10] Hunter, J. and L. Armstrong. *A Comparison of Schema for Video Metadata Representation*. in *The Eighth International World Wide Web Conference*. Toronto: Toronto Convention Centre, 1999.
- [11] ISO/IEC, *14496-11. Information technology - Coding of audio-visual objects - Part 11: Scene description and application engine*, 2003.
- [12] Jameux, D., *Symbole*, in *Encyclopédia Universalis*, 2004.
- [13] Kerbrat-Orecchioni, C., *Sémantique*, in *Encyclopædia Universalis*, 2004.
- [14] Martin, J.-P., *Description sémiotique de contenus audiovisuels*, Université Paris Sud: Orsay, p. 275 pages, 2005.
- [15] Odin, R., *Pour une sémio-pragmatique du cinéma*. Iris. **1**(1): p. 67-82, 1983.
- [16] Pottier, B., *Théorie et analyse en linguistique*, Paris: Hachette, 1987.
- [17] Prié, Y., *Modélisation de documents audiovisuels en Strates Interconnectées par les Annotations pour l'exploitation contextuelle*, in *Informatique*, Institut national des sciences appliquées de Lyon: Lyon. p. 276, 1999.
- [18] Rastier, F., *Arts et sciences du texte*, Paris: Presses Universitaires de France, 2001.
- [19] Richard, J.-F., *Introduction*, in *Cours de psychologie*. R. Guilglione J.-F. Richard: Paris. p. 7, 1999.
- [20] Tran-Thuong, T., *Modélisation et traitement du contenu des médias pour l'édition et la présentation de documents multimédias*, in *Département d'informatique*, Institut National de Recherche en Informatique et en Automatique: Grenoble, 2003.
- [21] Vaillant, P., *Glossaire de sémiotique*, in *revue-texto*, 1999.
- [22] Weiss, R., A. Duda, and D. Gifford, *Composition and Search with Video Algebra*. IEEE Multimedia Tools and Applications. **2**(n°1): p. 12-25, 1995.