

Connaissances et documents audiovisuels

Un modèle pour l'exploitation contextuelle des annotations

Yannick Prié^{*,} — Alain Mille^{**} — Jean-Marie Pinon^{*}**

** Laboratoire d'ingénierie des systèmes d'information - INSA-Lyon
20 av. Albert Einstein, F-69621 Villeurbanne cedex*

Yannick.Prie@insa-lyon.fr; pinon@lisi.insa-lyon.fr

*** LISA - CPE-Lyon, Equipe Raisonnement à Partir de Cas
43 Bd du 11 Novembre 1918 BP 2077 F-69616 Villeurbanne cedex*

Yannick.Prie@cpe.fr; am@cpe.fr

RÉSUMÉ. Cet article est consacré à l'étude de l'utilisation de connaissances contenues dans les documents audiovisuels¹. Nous montrons d'abord qu'il est nécessaire d'explicitier celles-ci sous la forme d'annotations rendant possible leur exploitation contextuelle. Nous présentons ensuite le modèle des Strates Interconnectées par les Annotations (Strates-IA) et les outils contextuels permettant une telle exploitation. Nous étudions enfin la nature des différentes connaissances utilisées dans un système Strates-IA.

ABSTRACT. This article is devoted to the study of the use of knowledge contained in audiovisual documents. We first show that this knowledge must be explicitated as annotations that allow its contextual exploitation. Then we present the Annotation Interconnected Strata model (AI-Strata) and the contextual tools that allow such an exploitation. We finally study the kind of knowledge used in a AI-Strata system.

MOTS-CLÉS : documents audiovisuels, annotations, connaissances, contextes.

KEYWORDS: Audiovisuel documents, Annotations, Knowledge, Context.

1. Ce travail s'inscrit dans le cadre du projet SESAME financé par le CNET, consultation thématique 96-ME-17. This work is part of the SESAME project granted by the French Center for Telecommunication Research under contract 96-ME-17.

1. Introduction

La notion de connaissance est comprise dans cet article comme implicitement liée au fait qu'elle permet à l'homme d'interpréter convenablement son environnement en situation d'action. Les techniques de l'intelligence artificielle ont étudié comment un système informatique pouvait représenter de la manière la moins approximative possible ces connaissances et les exploiter pour résoudre des problèmes ou assister les utilisateurs dans leurs tâches [SCH 93]. Rendre accessible à l'homme ce type de connaissances a été un des premiers objectifs des systèmes à base de connaissances [NEW 82]. La disponibilité d'une base de connaissances cohérente adaptée aux tâches de l'utilisateur constituerait en effet une assistance idéale pour la résolution de problème et agirait un peu comme une sorte d'encyclopédie cohérente opérationnelle. La mise en place et la maintenance d'une telle base de connaissances se révèle un problème extrêmement difficile et le plus souvent insurmontable pour un domaine d'expertise tel que celui d'une entreprise entière ¹.

A contrario, la mise en place d'un système documentaire est relativement aisée et pour faciliter l'exploitation des informations contenues dans les documents, un système documentaire propose en général des mécanismes d'indexation permettant *a minima* de retrouver un document selon quelques mots-clés ou mots simples appartenant à telle ou telle partie du document. Mais si un document quelconque véhicule bien des informations utiles et partageables, elles doivent être obligatoirement interprétées par l'homme pour accéder au statut de connaissances exploitables dans sa tâche, ce que la simple indexation par mots-clés ne garantit bien entendu pas.

C'est dans le cadre de la gestion de documents audiovisuels (DAV) que nous proposons une troisième voie – médiane – facilitant le passage au statut de connaissances des contenus documentaires, grâce à l'interconnexion de termes (d'annotations) issus d'une « base de connaissances » posés sur des parties de documents (des strates) au sein d'un graphe unique. L'exploitation de ce graphe de Strates Interconnectées par des Annotations (Strates-IA) est réalisée « en contexte », c'est-à-dire en se fondant d'abord sur la mise en évidence de contextes d'interprétation, sans *a priori* obligatoire sur la sémantique formelle des relations présentes dans la base de connaissances.

Pour présenter notre travail, nous analysons d'abord les rapports entretenus entre les connaissances potentiellement présentes dans les documents audiovisuels et les documents eux-mêmes. Les conclusions de cette analyse nous amènent alors à insister sur la nécessité de décrire explicitement les documents audiovisuels pour les exploiter, et la tâche de description est montrée comme générique aux tâches d'exploitation des systèmes d'information audiovisuelle (indexer, rechercher, naviguer, analyser...).

L'annotation est alors présentée comme l'opération atomique de toute tâche de description, et la contextualisation des annotations est proposée comme le moyen

1. Voir à ce sujet les difficultés rencontrées par D. Lenat [LEN 98].

privilegié de « donner du sens » aux termes mobilisés pour une tâche particulière. Nous présentons ensuite le modèle des Strates-IA comme support de description et d'exploitation des connaissances contenues dans les documents audiovisuels. Un journal télévisé sert d'illustration aux concepts présentés et plusieurs exemples permettent de décrire la richesse d'expression possible du modèle.

L'exploitation contextuelle de documents audiovisuels annotés selon le modèle des Strates-IA est alors décrite en montrant diverses utilisations possibles s'appuyant sur la notion de graphe potentiel caractérisé. Le contrôle des opérations de contextualisation est rendu possible par la définition des dimensions d'analyse, des valences et des schémas de description. Les tâches de l'utilisateur bénéficient alors de cadres précis pour leur mise en œuvre.

La validation du système est ensuite discutée au travers des prototypes réalisés, et la discussion est l'occasion de situer le travail et son ambition par rapport à des approches orientées « système à base de connaissances ».

2. Documents audiovisuels et connaissances

2.1. Diversité des utilisations de DAV et des connaissances documentaires qu'elles mobilisent

Les documents audiovisuels sont utilisés et manipulés dans un certain nombre d'organisations dans des objectifs très divers. Ainsi les bibliothèques ou le dépôt légal de la télévision² (Inathèque) mettent à disposition du public un très grand nombre de documents audiovisuels. Les entreprises comme les chaînes de télévision, le département Archives de l'INA (Institut national de l'audiovisuel) ou d'autres encore conservent les émissions produites, réalisées et diffusées.

Les documents audiovisuels conservés et exploités sont donc de types très divers, des films aux manuels techniques, des journaux télévisés aux bandes d'enregistrement d'un reporter ou aux clips vidéo. Les documents ou les fragments de documents peuvent être : soit simplement *visualisés* pour répondre à un besoin d'information particulier, par exemple apprendre tout sur la vie et l'œuvre d'une star de la chanson ou bien comment assembler les pièces d'un circuit électronique [LIE 94] ; soit *analysés*, ce qui correspond à une visualisation attentive faisant souvent appel à des lectures et relectures [AUF 99b], par exemple dans un cours d'université sur la télévision, ou bien dans le cadre d'une formation à la communication aidée par la vidéo ; soit enfin *réutilisés* dans de nouveaux documents, par exemple des images d'archive peuvent être « citées » dans une émission sur la télévision ou un film promotionnel.

2. Citons également la volonté récente du gouvernement des Etats-Unis de financer la restauration et la préservation de quelques films anciens « témoignages du génie américain ».

Les connaissances documentaires mobilisées au cours de ces tâches peuvent s'exprimer à différents niveaux et dépendent du type d'utilisation du document. Ce peuvent par exemple être des connaissances factuelles présentées comme telles dans des documents dont c'est le genre (un manuel technique ou un reportage), ce qui correspond au contenu du discours. Une analyse de la forme du discours peut également être menée, par exemple sur les types de mouvements de caméra utilisés ou bien la rhétorique du reportage de guerre. D'autres niveaux de connaissances existent encore, par exemple le nom et l'âge d'un acteur jouant dans un film, la date de la prise de son d'une interview, ou bien le statut juridique d'un document par rapport à sa réutilisation.

2.2. Interpréter les documents pour en expliciter les connaissances

Il apparaît en fait que toute connaissance mobilisée dans une lecture d'un document audiovisuel n'est utilisée *en tant que telle* que parce qu'elle est *explicitée* à un moment ou à un autre. Elle résulte alors d'une *interprétation* du document ou d'un de ses fragments dans le cadre d'une volonté d'action sur un matériau contenant *implicitement* toute la connaissance qu'on voudra y trouver.

Les connaissances de montage par exemple ³, les marques de structuration utilisées lors de la construction d'un flux n'existent plus dans le document final (hormis les ruptures de plan et de son). Atteindre à ces connaissances de structuration signifie les *(re-)construire* lors d'une tâche d'analyse du flux. Il en va de même pour toute autre connaissance liée à l'utilisation d'un document audiovisuel, et la multiplicité des interprétations et des connaissances véhiculées par un document audiovisuel est donc liée à la multiplicité des utilisations qu'il est possible d'en faire. La fonction sociale de la pratique documentaire est alors de « permettre un accès à des corpus de connaissances *via* l'organisation de repères qui utilisent des modes de représentation spécifique » [COT 99].

Il s'agit donc de s'interroger sur la question de ce que les technologies numériques peuvent apporter à la gestion de corpus de connaissances implicites, notamment en termes d'explicitation et d'exploitation des connaissances véhiculées.

3. Un document audiovisuel est presque toujours *monté*, c'est-à-dire que son créateur organise et construit les flux vidéo et audio à partir d'autres morceaux de flux, de manière à induire contextuellement des inférences narratives, factuelles ou autres chez le spectateur. Par exemple il associera un plan représentant un personnage devant une fenêtre à un autre plan représentant l'extérieur d'une maison avec fenêtre, tandis qu'une même bande-son permettra de faire le lien contextuel entre la situation du personnage dans la maison et celle de cette maison dans le paysage.

3. Nécessités pour la représentation de DAV dans des systèmes d'information audiovisuelle

Les *systèmes d'information audiovisuelle* (SIAV) sont rendus possibles par les progrès technologiques récents tant en termes de puissance des ordinateurs (stockage et visualisation) que de débits des réseaux (transmission). Il devient alors possible d'envisager l'exploitation directe des documents audiovisuels sous forme numérique.

3.1. Décrire des DAV pour les exploiter, exploiter des DAV en les décrivant

Remarquons tout d'abord que l'appréhension standard d'un document audiovisuel se fait de manière temporalisée et que le mode de lecture fondamental en est la simple *visualisation*, qui peut s'agrémenter de fonctions telles que celles que l'on trouve sur un magnétoscope, avance et retour rapides.

Avec la numérisation, le document audiovisuel devient *manipulable* dans une machine, c'est-à-dire qu'il est désormais techniquement possible d'accéder quasi instantanément à l'une quelconque de ses parties (accès direct), de naviguer d'une partie de document à une autre, mais aussi de découper, de monter, en un mot de construire de nouveaux documents numériques⁴. Il est alors nécessaire de décrire informatiquement les documents afin d'en opérationnaliser les diverses exploitations possibles, au travers de la manipulation des descriptions.

Nous considérons [PRI 98] que les principales tâches d'utilisation d'un système d'information audiovisuelle correspondent toutes à des tâches de *description* d'un document par un utilisateur. Ainsi, l'*indexation* consiste à décrire un document, la *recherche* demande une description du document tel qu'il est recherché, l'*analyse* est une sur-description détaillée, et l'*édition* ou la *génération* [AUF 99a] correspondent à la description du document tel qu'il va être produit. La *navigation* enfin est une description partielle de ce que l'on cherche à atteindre.

3.2. Annoter des documents audiovisuels pour les décrire

La description d'un document audiovisuel correspond à son *indexation*, c'est-à-dire sa paraphrase sous une forme sémiotique interprétable permettant son exploitation, ici numérique. Or, à la différence du texte, l'audiovisuel ne repose pas sur un système fonctionnel connu et donc manipulable comme la langue. Par exemple, les primitives (textures, formes, mouvements) que l'on peut extraire des images n'ont

4. Certains documents sont d'ores et déjà créés directement sous forme numérique, du filmage par une caméra numérique au montage et à la diffusion.

pas de sens, et n'existent qu'au niveau du calcul et non à celui de l'interprétation [BAC 99]. En d'autres termes, le contenu audiovisuel n'existe pas tant qu'il n'a pas été explicité suite à une interprétation, un document audiovisuel ne peut être son propre index (alors qu'un texte peut l'être) et les unités de manipulation de documents audiovisuels doivent être créées explicitement.

Ceci implique – jusqu'à ce qu'un hypothétique langage audiovisuel soit mis en place – qu'il n'existe pas de système d'indexation primaire de l'audiovisuel autre que celui existant par défaut et lié aux repères temporels (si $t_{DebutDoc} \leq t \leq t_{FinDoc}$ alors t peut être considéré comme index du document Doc). Ces index sont alors fortement décorrélés du contenu sémantique du document, et donc peu utiles.

Un deuxième niveau d'index devra alors permettre de décrire ce contenu, qui sera interprété et explicité (cf. 2.2), et les index pourront prendre le statut de connaissances dans le cadre de tâches d'exploitation des documents⁵. L'enjeu premier induit par le numérique consiste donc en description du document audiovisuel à l'aide de descripteurs symboliques explicites qui en autoriseront la manipulation.

L'*annotation* est l'opération qui consiste à décrire une partie d'un document, en lui attachant un descripteur.

L'approche générale d'annotation est l'approche dite de *stratification* [DAV 93]: la mise en place des descripteurs y guide la création des parties de document (ou strates), en leur adjoignant un contenu sémantique. C'est, par exemple, parce que telle action a été repérée dans le document entre deux bornes temporelles que l'on décide de créer une *strate* annotée par le codage de cette action. On notera qu'alors le découpage du document en parties n'est pas réalisé *a priori* (suivant un éventuel modèle hiérarchique de document), mais quand le besoin s'en fait sentir, de façon directement liée au contenu annoté.

3.3. Contextualisation des annotations et exploitation d'un SIAV

Nous avons montré dans [PRI 99a] l'importance du contexte d'une annotation pour en guider l'interprétation. Nous distinguons plusieurs types de contextes. Le contexte *temporel* est le contexte d'appréhension le plus simple, lié à la temporalité du médium. Par exemple, dans le cadre de la stratification, les annotations de deux strates qui se recouvrent sont dans le même contexte temporel. Le contexte *sémantique* concerne toutes les autres relations contextuelles. Le contexte *structurel* par exemple provient d'une structuration du document: les annotations d'un plan peuvent annoter contextuellement la séquence qui le contient, tandis que le plan peut hériter des annotations du document (par exemple le nom du réalisateur). Ce type de contexte peut être rapproché de la notion d'attribut « propagé » au long de relations

5. Tout d'abord dans la tâche primaire d'indexation.

structurelles [CHI 97]. Les autres contextes sémantiques font appel à divers liens de coréférence entre objets temporels annotés d'un ou de plusieurs documents : par exemple, le fait que le composant électronique présenté à la fin du document est bien le même que celui présenté au début, ou bien qu'un lien entre une musique et un personnage permettra plus tard de n'utiliser que la musique pour symboliser le personnage.

Si le contexte temporel est partagé dans toute utilisation d'un document audiovisuel (simple visualisation), le contexte sémantique est lié à une utilisation particulière du document suivant une visée de contextualisation réalisée dans un tâche particulière. Nous avons vu que nous considérons toute tâche d'utilisation d'un SIAV comme une tâche de description. Comme cette description correspond en fait à l'explicitation d'annotations symboliques et leur mise en place dans des contextes dans lequel ils vont prendre un sens particulier, la « mise en contexte », la contextualisation sera l'opération de base de toute description dans une situation d'action spécifique. Ainsi, dans l'indexation, chaque annotation mise en place pour décrire un objet d'intérêt du flux prendra sens dans son contexte. Une recherche consistera quant à elle en la description des annotations cherchées et de leurs contextes significatifs. La navigation consistera à suivre des contextes signifians, etc.

Il apparaît en conclusion de cette partie 1) que toute connaissance contenue dans un document audiovisuel doit être explicitée pour être utilisable 2) que toute exploitation de document audiovisuel numérique passe par l'annotation (l'indexation) de celui-ci 3) que le contexte audiovisuel doit être pris en compte quel que soit le type de tâche de l'utilisateur.

Nous proposons donc d'écrire (au sens large du terme) sur le document, avec le document. Son utilisation autre que la visualisation résultera alors d'une lecture et/ou d'une écriture contextuelles des annotations. L'ordinateur et la technologie accompagneront l'utilisateur aux niveaux de la représentation de l'annotation (partie 4) et de l'utilisation contextuelle de ces annotations (partie 5).

4. Modélisation des annotations de documents audiovisuels

Nous utiliserons au long de cet article un exemple audiovisuel extrait du journal de France 2 du 13 juillet 1996⁶. La figure 1 présente une vue globale du document : le journal est composé d'un indicatif, d'une ouverture, et de quelques reportages à chaque fois introduits par le présentateur. L'un de ces reportages a pour sujet la visite de Nelson Mandela à Paris à l'occasion du 14 juillet, et peut se décomposer (avec au moins une image par plan) en un lancement du sujet ; une introduction présentant la rencontre de Mandela et Chirac au château de Rambouillet ; un rappel de

6. Nous remercions le Département Innovation de l'Institut National de l'Audiovisuel qui a fourni les matériaux de tests reproduits dans cet article

la rencontre de Mandela et Mitterrand quelques années plus tôt ; une rétrospective sur la vie de Mandela en Afrique du Sud avec sa condamnation, ses années de prison et sa libération ; et enfin une conclusion dans laquelle Mandela et Chirac marchent dans les allées du château de Rambouillet.

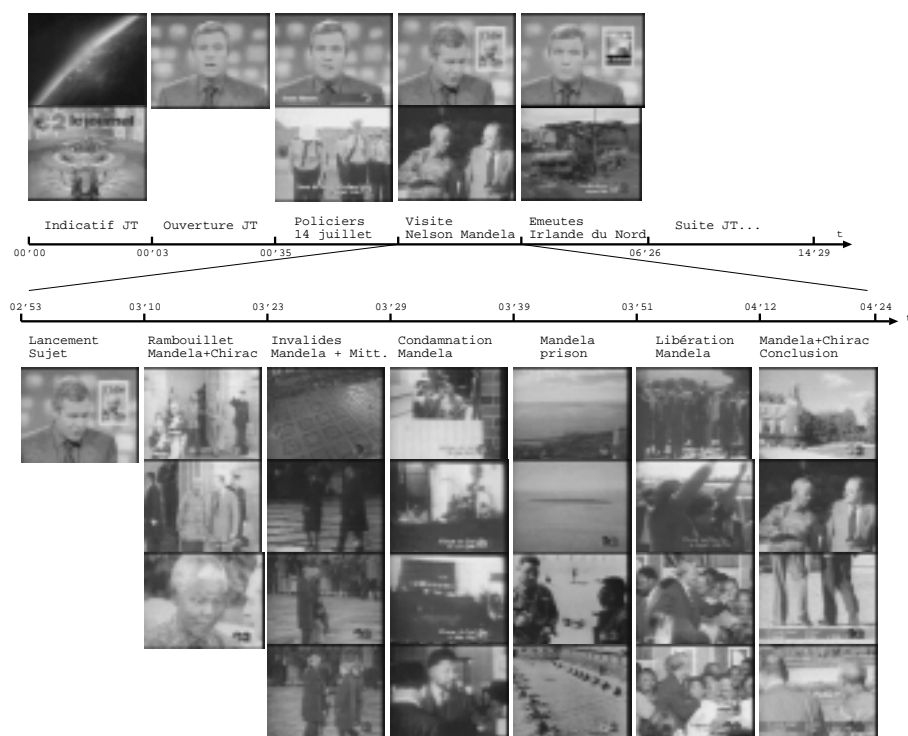


Figure 1. *Vue d'ensemble du journal télévisé*

Les annotations de documents audiovisuels peuvent correspondre à plusieurs types de descripteurs. Les descripteurs de bas-niveau sont liés à des résultats issus de traitement du signal (image ou son) sur lesquels on pourra faire une recherche de similarité, par exemple des histogrammes de couleur ou des spectres sonores. Les descripteurs dits de haut-niveau peuvent permettre d'identifier des unités structurales (par exemple un plan ou un document), ou bien des personnages (Mandela, Mitterrand) ou encore des actions (poignée de main) ou des vues (gros plan, plan d'ensemble), etc.

Nous nous plaçons dans le cadre général de la stratification, et considérons que toute annotation, quel que soit son niveau, est une explicitation par la personne qui décrit le flux de la détection d'un *objet d'intérêt* en *élément d'annotation* (EA). Par exemple, repérer dans le flux la présence de « Jacques Chirac » conduira à mettre en place un élément d'annotation (*Chirac*). Les éléments d'annotation sont caractérisés par un *terme* et un certain nombre d'attributs-valeurs. Par exemple $\langle \text{Journal Télévisé} ;$

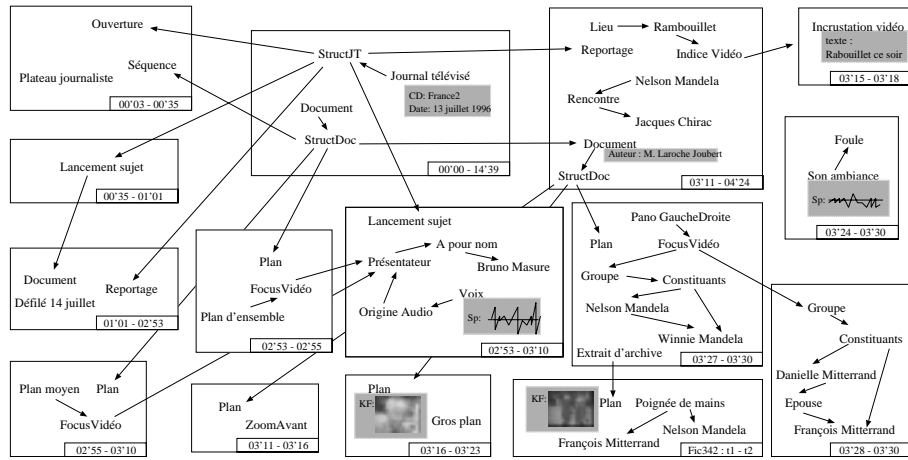


Figure 2. Une vue « emboîtée » de l'annotation d'un journal télévisé

$\langle Date : 13.07.96 \rangle$ est un élément d'annotation de terme *Journal Télévisé* ayant un attribut *Date*, et $\langle Plan ; KeyFrame : KF326.jpeg \rangle$ un EA décrivant un plan, avec un attribut contenant une image représentative du plan.

La mise en place d'un élément d'annotation *primitif* correspond à la création d'une strate ou *unité audiovisuelle* (UAV) qui représente un morceau du document audiovisuel, et est annotée par cet EA. Une unité audiovisuelle peut être annotée par autant d'éléments d'annotation que nécessaire, mais par un unique EA primitif.

Afin de structurer l'annotation, les EA peuvent être connectés par des *relations élémentaires* R_e dont la sémantique est limitée au simple fait de mettre en relation. Une mise en relation peut ne pas être expliquée (par exemple entre $\langle Nelson Mandela \rangle$ et $\langle Winnie Mandela \rangle$), mais si on désire en exprimer la sémantique, on utilise un EA supplémentaire, qui sera en relation élémentaire avec l'EA de départ et l'EA d'arrivée. Par exemple $\langle Présentateur \rangle$ et $\langle Bruno Masure \rangle$ peuvent être mis en relation par l'intermédiaire de l'EA $\langle A \text{ pour nom} \rangle$ sous la forme $\langle Présentateur \rangle R_e \langle A \text{ pour nom} \rangle R_e \langle Bruno Masure \rangle$. La mise en place de relations élémentaires entre EA annotant des UAV différentes permet de mettre celles-ci en relation, d'où le nom du modèle : *Strates Interconnectées par les Annotations* (Strates-IA).

La figure 2 présente une vue d'un début d'annotation (résultant en partie d'un travail avec une documentaliste de l'INA Centre-Est) du journal télévisé de la figure 1. Les UAV sont représentées par un encadrement de leurs bornes temporelles (faisant référence sauf indication contraire au même fichier). Les EA annotant les UAV sont représentés par leur nom et le cas échéant leurs attributs et les flèches représentent les relations élémentaires. L'EA $\langle FocusVidéo \rangle$ permet de désigner l'objet principal visuel d'un plan. On notera la double structuration du journal télévisé en ses compo-

santes, et du document en séquences, plans, et autres documents, et que les relations élémentaires sont mises en place entre EA annotant une même UAV, ou bien deux UAV différentes, n'appartenant pas forcément au même document (par exemple l'EA *Extrait d'archive* est en relation avec l'EA *Plan* annotant une UAV représentant un morceau d'un autre fichier).

Afin d'avoir un vocabulaire d'annotation contrôlé, les éléments d'annotation sont des *inscriptions dans le flux d'éléments d'annotation abstraits* (EAA) organisés dans ce que nous appelons « base de connaissances », qui ne doit pas être considérée comme une ontologie définitoire. Elle correspond en effet *a minima* à un thésaurus portant sur les termes des EA, organisé en relation de spécialisation/abstraction, mais d'autres relations de thésaurus sont possibles (par exemple *voir aussi*), ou même des relations exprimant d'autres niveaux de connaissances (par exemple la relation *pays* exprimant la provenance d'un personnage ou d'un événement). Remarquons que dans la mesure où un EA intermédiaire de relation se révèle général et partagé (ce qui pourrait être le cas de *Epoque*), il pourra être pris en compte directement comme relation conceptuelle dans la base de connaissances, comme par exemple la relation *Pays*.

La figure 3 présente une base de connaissances liée à l'annotation décrite figure 2. Par commodité, nous avons « factorisé » certains attributs suivant la relation de spécialisation : l'EAA *Genre d'émission* possède un attribut *CD* (chaîne de diffusion) de type texte, ainsi que *EAA:Document*, *EAA:Séquence* et *EAA:Plan*.

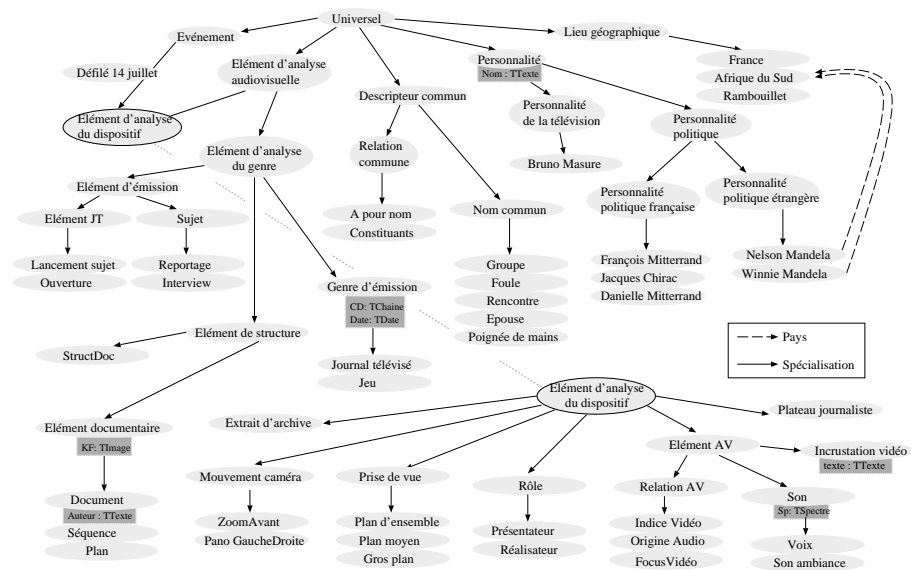


Figure 3. Un exemple de « base de connaissances » (pour des raisons de place, la branche Élément d'analyse du dispositif est détaillée séparément)

5. Exploitation contextuelle des documents annotés

5.1. Contextes et graphe potentiels

L'ensemble du système (flux annotés et base de connaissances) est un graphe orienté étiqueté dont les sommets sont les UAV, les EA et les EAA. Nous définissons alors le *contexte* d'un sommet x quelconque du graphe comme l'ensemble des sommets y_i qui peuvent être mis en relation avec lui, c'est-à-dire tels qu'il existe des chemins $x \dots y_i$ (indépendamment du sens des relations). Comme le graphe est connexe, le contexte complet d'un sommet du graphe est l'ensemble du graphe lui-même, ce qui exprime le fait que tout descripteur – comme connaissance totalement explicitée – est défini par l'ensemble des relations qu'il entretient avec les autres éléments du système. Cette notion, pour juste qu'elle soit, n'est cependant pas opératoire, et il s'agit d'être à même de contrôler l'expression des chemins contextuels dans le graphe.

Dans cet objectif, nous définissons un *graphe potentiel caractérisé* (GPC) comme un graphe construit sous les contraintes des Strates-IA (types de sommets, relations étendues aux relations temporelles⁷), avec la liberté supplémentaire qu'il est possible de banaliser tout ou partie des termes et attributs d'un certain nombre de sommets. Ces sommets sont alors appelés sommets génériques, par exemple l'EA (*). Certains sommets du graphe potentiels sont caractérisés, c'est-à-dire qu'on leur associe un terme qui sera utile pour leur exploitation (par exemple la réunion de deux GPC suivant tel et tel sommet). Un graphe potentiel caractérisé est *instancié* dans le graphe général, c'est-à-dire que l'on cherche tous les sous-graphes partiels du graphe général qui sont isomorphes (aux contraintes de généralité près) avec le graphe potentiel.

Nous avons mis au point à cet effet un algorithme de *multi-propagation* qui permet de mener une recherche d'isomorphisme de sous-graphe partiel dans le cas particulier où on connaît des associations de départ. Ces associations sont liées à des nœuds du graphe potentiel qui ont des correspondants non ambigus dans le graphe global, par exemple des EAA (uniques par définition) ou bien des UAV ou EA explicitement spécifiés, ce qui est toujours le cas dans notre application. La recherche d'isomorphisme est alors *propagée* à partir des associations connues en suivant une heuristique permettant d'aller rapidement vers la solution. Les solutions sont rendues par l'algorithme au fur et à mesure de leur découverte, ce qui est une caractéristique intéressante en recherche d'information [PRI 00].

Les graphes potentiels caractérisés permettent d'exprimer des relations contextuelles multiples entre sommets du graphe. Ils peuvent être manipulés, joints, découpés et représentent donc l'outil de base permettant d'exploiter un système Strates-IA. Toute requête au graphe doit être exprimée par des graphes potentiels, et leur

7. Ces relations se trouvent entre deux éléments d'annotation, et expriment les relation temporelles des unités audiovisuelles que ceux-ci annotent respectivement.

utilisation est représentative de la visée de contextualisation, donc d'action sur des connaissances d'un utilisateur. Nous en présentons quelques exemples dans la suite.

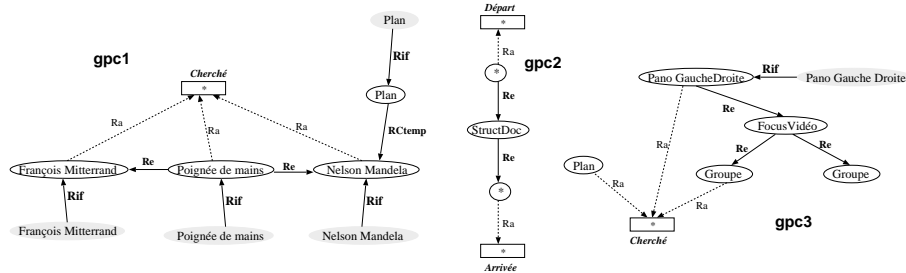


Figure 5. Exemples de graphes potentiels, les termes de caractérisation de certains sommets sont en italiques

5.2. Exploiter le graphe global à l'aide de graphes potentiels

La figure 5 présente trois graphes potentiels caractérisés, liés à diverses utilisations.

Le premier, *gpc1* illustre une utilisation simple d'un graphe potentiel pour une recherche d'une unité audiovisuelle telle qu'elle soit contenue dans un $\langle Plan \rangle$ (relation temporelle $\langle CTemp \rangle$), et annotée par $\langle Nelson Mandela \rangle$, $\langle François Mitterrand \rangle$ et $\langle Poignée de main \rangle$ en relations. L'instanciation de *gpc1* est une recherche à partir des quatre associations connues liées aux quatre EAA du graphe potentiel⁸, et le sommet *Cherché* de son unique instance désignera l'UAV [*Fic342* : $t1 - t2$].

L'exemple de recherche utilisant *gpc1* se révèle intéressant à deux titres. D'une part, la possibilité de requête est nettement plus riche que par exemple si on cherchait une strate annotée par $\langle Nelson Mandela \rangle$, $\langle François Mitterrand \rangle$ et $\langle Poignée de main \rangle$, comme simples mots-clés, sans que leurs relations contextuelles soient explicitées. D'autre part, nous ne nous trouvons pas non plus ici dans un système à base de connaissances dans lequel un concept (par exemple $\langle Poignée de main \rangle$) serait strictement et uniquement défini dans toutes ses utilisations possibles et les inférences réalisables à partir de lui (comme ce serait le cas avec des graphes conceptuels).

Le deuxième graphe potentiel caractérisé *gpc2* décrit un chemin contextuel très simple et permet de naviguer dans le graphe global suivant des relations de structuration documentaire explicitées à l'aide de l'EA $\langle CStruct \rangle$. Il convient pour l'utiliser d'identifier le sommet UAV caractérisé par *Départ* avec l'UAV à partir de laquelle on

8. L'heuristique de l'algorithme spécifie alors que la première étape de la propagation se fera à partir de l'EAA qui est le moins utilisé pour l'annotation (figure 2), soit sur notre exemple $\langle EAA : Poignée de main \rangle$, utilisée une seule fois.

souhaite naviguer, et d’instancier le graphe *gpc2'* ainsi défini. Par exemple, utiliser *gpc2* à partir de l’UAV [00'00 – 14'29] permet de proposer d’atteindre les UAV [00'03 – 00'35], [02'55 – 03'10], [02'53 – 02'55] et [03'11 – 04'24] (sommets des instances caractérisés par *Arrivée*).

Le troisième graphe potentiel caractérisé *gpc3* permet de rechercher des unités audiovisuelles illustrant un panoramique de la gauche vers la droite passant d’un groupe à un autre. Instancié dans notre exemple, il permet de retrouver l’UAV [03'27 – 03'30].

5.3. Des moyens de contrôler l’annotation

Le graphe *gpc1* peut être mis en place sans connaître exactement l’utilisation de l’EA *⟨Poignée de main⟩* (éventuellement en cherchant par essais/erreurs la bonne syntaxe). Ceci n’est cependant pas possible pour *gpc3*, où il est très difficile de deviner la manière dont un utilisateur a utilisé les EAA à sa disposition pour exprimer sa visée de description d’un panoramique gauche-droite. Il apparaît alors qu’il est nécessaire d’être à même de spécifier comment décrire des flux à l’aide des éléments d’annotation et de leurs mises en relations, afin de pouvoir ensuite utiliser ces connaissances pour la mise en place de requêtes. Nous présentons dans la suite de cette partie quelques outils d’exploitation des Strates-IA basés sur les graphes potentiels caractérisés et permettant, à divers degrés, le contrôle de l’annotation.

Dimensions d’analyse (DA). Une dimension d’analyse permet de découper la base de connaissances en regroupant divers éléments d’annotation abstraits qui seront utilisés au cours d’une analyse (d’une annotation) d’un document audiovisuel. Par exemple, une dimension d’analyse *⟨DA: Hommes politiques⟩* regroupera tous les EAA désignant des hommes politiques, une autre sera consacrée aux actions ou aux mouvements de caméra. Une dimension d’analyse peut également ne désigner qu’un seul élément d’annotation, par exemple *⟨DA: Plan⟩*, ou être construite à partir d’autres dimensions d’analyse. Par exemple, on peut avoir une DA *⟨DA:Prise de vue⟩* composée des DA *⟨DA:Vue⟩* (désignant *⟨EAA: Plan d’ensemble⟩*, *⟨EAA: Plan moyen⟩*, *⟨EAA: Gros plan⟩*) et *⟨DA:Mouvement Caméra⟩* (désignant *⟨EAA: Zoom Avant⟩* et *⟨EAA: Pano GaucheDroite⟩*).

Les dimensions d’analyse représentent la visée de description de l’utilisateur, et on en utilisera plusieurs pour décrire un flux, par exemple celles des hommes politiques et des actions, des mouvements de caméra et des incrustations visuelles. On remarquera également que si certaines d’entre elles sont « naturelles » à l’organisation de la base de connaissances, car basées sur la relation de spécialisation comme *⟨DA:Mouvement Caméra⟩*, d’autres peuvent simplement exprimer une visée de description locale à une tâche, ce qui est par exemple le cas de *⟨DA:Vue⟩* telle que nous l’avons présentée.

Valences. Une valence⁹ est un attribut d'un EAA $\langle EAA:x \rangle$ exprimant des possibilités de relations élémentaires des EA $\langle x \rangle$ qui en sont issus. Une valence est donc composée d'un graphe potentiel désignant une relation contextuelle entre un point de départ EA $\langle x \rangle$ et des points d'arrivée $\langle y_i \rangle$, avec lesquels il est possible de mettre $\langle x \rangle$ en relation élémentaire. Par exemple $\langle EAA:Poignée\ de\ main \rangle$ pourra avoir une valence indiquant qu'à la mise en place de l'EA $\langle Poignée\ de\ main \rangle$, il y aura lieu de chercher dans un certain contexte (par exemple le contexte de cooccurrence temporelle) si des EA issus d'une dimension d'analyse $\langle DA: Personnalité\ politique \rangle$ existent et peuvent être mis en relation avec lui [PRI 99b]. Les valences peuvent également être considérées comme des indications décrivant des schémas locaux d'annotation possibles. Nous étendons cette notion de schéma local de description afin de pouvoir contraindre fortement une mise en place d'annotations.

Schémas de description (SD). Les schémas de descriptions sont des graphes orientés dont les sommets sont étiquetés par des dimensions d'analyse spécifiant des EA à utiliser et dont les relations spécifient :

- des relations entre ces EA, du type $R_e[n, m]$ pour les relations élémentaires,
- des relations d'annotation R_a entre EA source de la relation et unités audiovisuelles qu'annotent les EA cibles,
- des relations temporelles (par exemple *après*).

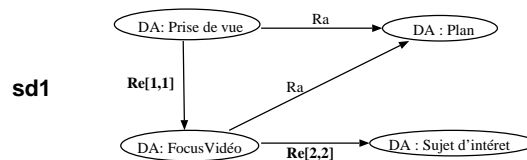


Figure 6. Un schéma de description exprimant la nécessité de décrire un plan par un type de prise de vue et un focus vidéo vers un sujet d'intérêt

Un schéma de description correspond donc à la description d'un réseau d'éléments d'annotation qu'il convient de mettre en place. Ainsi, le schéma de description *sd1* de la figure 6 spécifie-t-il qu'une unité audiovisuelle annotée par $\langle Plan \rangle$ devra également être annotée par un EA extrait de $\langle DA: Prise\ de\ vue \rangle$ (par exemple $\langle Pano\ GaucheDroite \rangle$) en relation élémentaire avec un EA $\langle Focus\ Vidéo \rangle$ lui-même en relation élémentaire avec deux EA extraits de la dimension d'analyse $\langle DA: Sujet\ d'intérêt \rangle$ ¹⁰. Le schéma de description *sd1* correspond à un schéma utilisé dans l'annotation de la figure 2, et il est possible de s'en inspirer pour mettre en place des graphes potentiels caractérisés correspondant à des requêtes valables dans le graphe d'annotation. Ainsi, décrire l'annotation d'une unité audiovisuelle telle qu'on la re-

9. Au sens d'un liaison potentielle, par analogie avec celles des molécules en chimie.

10. Par exemple définie à l'aide des dimensions d'analyse $\langle DA: Personnalités \rangle$ et $\langle DA: Noms\ communs \rangle$.

cherche (cf. 3.1) en utilisant *sd1* permet de mettre en place le graphe potentiel caractérisé *gpc3* de la figure 5.

Les schémas de descriptions permettent de spécifier les annotations à mettre en place de façon locale, en décrivant des relations possibles entre ensembles d'éléments¹¹. A la possibilité de les utiliser pour construire les graphes potentiels requêtes – c'est-à-dire de retrouver les contextes ayant présidé à la mise en place de l'annotation par l'annotateur – s'ajoute celle de générer aisément des schémas de description à partir de sous-graphes d'annotation. Il devient alors possible d'apprendre et de réutiliser les manières de décrire des documents audiovisuels par les utilisateurs.

6. Résultats et discussion

6.1. Prototypes

Un premier prototype a été implanté afin de valider l'approche de description par graphes et surtout de vérifier les performances de l'algorithme de multipropagation. Celles-ci se révèlent de loin supérieures à celles des algorithmes de recherche d'isomorphisme de sous-graphes partiels connus (par exemple [COR 98]) et satisfont les contraintes que nous nous sommes fixées [PRI 00].

Un second prototype SESAME¹² offre une interface graphique conviviale permettant la mise en place de la base de connaissances, de dimensions d'analyse, ainsi que l'annotation de flux audiovisuels (figure 7). L'exploitation de la base fondée sur les requêtes à l'aide de graphes potentiels et sur l'exploration sont en cours d'implantation. Une expérimentation sur un corpus de quelques heures de documents audiovisuels pourra alors être menée et permettre de valider notre approche (travail poursuivi dans le cadre du projet RNRT RECIS¹³).

6.2. Exploiter un espace d'information

L'exploitation de documents audiovisuels, décrits suivant le modèle des Strates-IA, se fait en utilisant systématiquement les outils d'exploitation de contextes relationnels présentés. Ainsi, nous avons vu que la recherche de documents ou de morceaux de

11. On peut donc dire qu'ils décrivent des arrangements syntagmatiques entre éléments issus d'ensembles paradigmatiques et pourraient permettre au système de se rapprocher d'un système fonctionnel tel que celui de la langue.

12. SESAME : Système d'Exploration de Séquences Audiovisuelles et Multimédia enrichi par l'Expérience.

13. RECIS : Recherche et Exploration de Contenus Image et Son, collaboration INSA-Lyon, CNET, INRIA, Philips et Nouvelles Frontières.

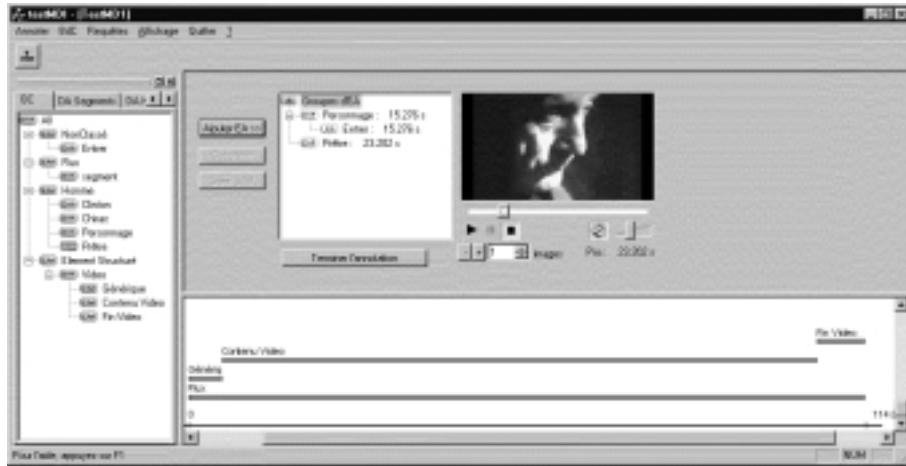


Figure 7. Interface d'annotation du prototype SESAME : les dimensions d'analyse utilisées sont décrites à gauche, le flux est représenté dans un lecteur (haut-droite), mais aussi par les unités audiovisuelles qui le décrivent (en bas). Deux unités audiovisuelles sont en cours de création (haut-milieu)

documents audiovisuels consiste à mettre en place les graphes potentiels caractérisés exprimant les descriptions recherchées sous la forme de mises en contexte de descripteurs. Ces graphes potentiels peuvent être construits entièrement mais l'utilisateur s'inspirera le plus souvent de graphes-requêtes canoniques, d'exemples d'annotation ou de parties de schémas de description. La navigation est bien entendu naturelle aux Strates-IA, puisqu'il est possible de suivre toutes les relations mises en place, mais aussi d'atteindre des nœuds décrits par des contextes.

Il apparaît donc que la recherche par requêtes et celle par navigation convergent vers des épisodes intégrés d'*utilisation d'information* [CHI 97] dans un espace d'information. Dans cet espace, les connaissances permettant de décrire (éléments d'annotations abstraits), les connaissances de description (éléments d'annotations et relation élémentaire) et les documents (ici représentés par des unités audiovisuelles) sont exploitables suivant des mises en contexte mettant en jeu ces trois niveaux. Cette exploitation est rendue possible par la représentation comme éléments d'un graphe unique de tous les éléments du système.

6.3. Situation des Strates-IA par rapport aux principales approches de représentation documentaire fondées sur les connaissances

Dans le domaine des hypermédias, l'approche fondée sur les connaissances est utilisée et concerne principalement les connaissances externes aux documents, per-

mettant d'y accéder. Un exemple se trouve dans [NAN 95], qui utilise un réseau sémantique avec une hiérarchie de types et des nœuds instances liés à des morceaux de documents. L'ensemble du système est alors une correspondance entre un réseau de connaissances et un ensemble de documents. La navigation et l'inférence sont possibles sur les instances de concepts, qui représentent des connaissances factuelles explicites valides pour des documents techniques. Des expressions de chemins liées aux concepts permettent de supporter la lecture du document suivant des visées définies à l'avance. L'utilisation des concepts du réseau est relativement figée (puisque décrits *hors* des documents) et nécessite la mise en place des notions de *qualificateurs* de relations afin d'en gérer les différentes utilisations.

Au contraire, dans les Strates-IA, chaque élément d'annotation est unique, une connaissance *en contexte* (instance dans le document), qui peut s'écarter du sens de l'élément d'annotation abstrait (concepts hors du document) dont il est extrait, notamment du fait de ses relations élémentaires. Toute requête est alors naturellement contextuelle puisqu'elle peut spécifier tout à la fois des prescriptions sur les concepts et leurs instances.

Dans le cadre de l'extraction de connaissances à partir de documents, [MAR 96] propose de décrire des morceaux m de documents (éléments SGML) par des graphes conceptuels en associant à des morceaux contenus dans m des liens vers des concepts, des relations entre concepts permettant de mettre en place un graphe conceptuel, lequel est alors considéré comme annotation de m . La navigation dans les graphes conceptuels vers les parties de documents est alors possible, et les requêtes consistent à rechercher des parties annotées par des graphes conceptuels permettant d'en déduire des graphes requêtes.

Dans le domaine de la recherche d'information, [CHI 97] choisit de décrire des documents directement à l'aide de graphes conceptuels. Ainsi, une structure est définie (document, chapitre, section...) avec des relations structurelles entre éléments de structure, et chacun de ces éléments peut ensuite être décrit à l'aide de graphes conceptuels en exprimant le contenu. La recherche de parties de document s'effectue en utilisant la notion de transmission d'attributs : on considère qu'un élément de structure est décrit par les descriptions des éléments qu'il contient, ce qui correspond à une opération de joint maximal sur celles-ci. Une requête « orientée précision » est un graphe conceptuel dont il s'agit de trouver un élément de structure qui permette de le déduire¹⁴.

Les relations entre éléments d'annotation des Strates-IA permettent de lier des parties de documents, à la manière de [MAR 96], mais nous ne nous limitons pas à des graphes de description liant des parties contenues dans d'autres. Nous choisissons, de la même manière que [CHI 97], de décrire les connaissances de structure et les connaissances « de contenu » dans un réseau unique, considérant que les connais-

14. On est alors dans le cadre du *modèle logique* de recherche d'information.

sances de structure d'un document audiovisuel font partie de son contenu, du fait de l'absence de structure documentaire dans leur représentations. A la différence des deux approches présentées, nous ne considérons cependant pas que les contraintes de description doivent être exprimées dans les concepts (sous la forme de mises en relation strictes), mais plutôt dans la manière dont ceux-ci sont utilisés (à la base en tant que mots-clés). Nous laissons ainsi *a priori* toute latitude à l'utilisateur pour l'organisation de ses graphes d'annotation, les schémas de description contenant localement les règles de description. Nous considérons en effet que les connaissances d'annotation sont par nature semi-structurées et ne se laissent pas obligatoirement ramener à des schémas globaux pré-construits.

Le statut des connaissances varie donc. Des connaissances et des relations toujours vraies de la base de connaissances, on en arrive à des structurations locales et non prévues d'instances de ces connaissances en contexte, tandis que l'utilisation à l'aide de graphes potentiels décrivant des contextualisations ajoute un degré de liberté supplémentaire à la fois aux requêtes et à la navigation. Ceci ouvre la voie à la prise en compte continue de nouvelles connaissances sur le document, liées à son utilisation concrète. En effet, nous comptons utiliser la notion de contexte pour décrire des modèles explicites de tâches. Ces modèles de tâche associés au modèle d'utilisation général constitutif des Strates-IA, permettront de rationaliser les traces d'utilisation et en conséquence de fournir les mécanismes de base à la réutilisation de l'expérience concrète pour assister les tâches et participer à l'enrichissement des connaissances de la base [PRI 99b].

7. Conclusion

Nous avons dans cet article montré que toute utilisation informatique de documents audiovisuels autre que la simple visualisation nécessitait une explicitation des connaissances qu'ils contiennent, et qu'il y avait lieu d'exploiter contextuellement celles-ci. Nous avons donc présenté un modèle de description de documents audiovisuels en Strates-Interconnectées par les Annotations à même de prendre en compte cette complexité. Après avoir défini strictement le contexte d'un élément comme un chemin de relations dans le graphe d'annotation, nous avons présenté les graphes potentiels comme moyen de description de contextes et de requêtes au système. Ainsi, le contexte de la tâche de l'utilisateur (inaccessible par définition) est matérialisé sous la forme de visées de contextualisation. Sur le constat de la nécessité de contrôler l'annotation pour pouvoir l'utiliser autrement que de façon exploratoire, nous présentons divers outils permettant ce contrôle. Nous présentons alors rapidement les prototypes réalisés et discutons le statut des connaissances manipulées.

Quelques directions de travail sont ouvertes autour des Strates-IA. Le modèle, issu des travaux du projet SESAME sera utilisé dans le projet RECIS comme support global des annotations. L'étude et la formalisation des mécanismes de réutilisation

de l'expérience concrète non présentées ici est un travail qui sera poursuivi dans le cadre de ce dernier projet. Une autre étude concerne les possibilités d'utilisation du modèle comme opérationnalisation des principes de MPEG7, qu'il est à même de prendre en compte. Enfin, un travail s'est attaché à étudier les possibilités d'utilisation de XML pour la description de graphes Strates-IA, et sa prolongation sera concernera l'adaptation de l'algorithme d'instanciation de graphes potentiels sur des graphes non plus en mémoire, mais dans un ensemble de fichiers-documents structurés.

Remerciements

Nous remercions les relecteurs pour leurs pertinentes remarques sur la première version de cet article.

8. Bibliographie

- [AUF 99a] AUFFRET G., BACHIMONT B., « Audiovisual Cultural Heritage: from TV and radio archiving to hypermedia publishing », *Third European Conference on Research and Advanced Technology for Digital Libraries*, Paris, septembre 1999, LNCS 1696, Springer-Verlag, p. 58–75.
- [AUF 99b] AUFFRET G., PRIÉ Y., « Managing Full-indexed Audiovisual Documents: a New Perspective for the Humanities », *Computer and the Humanities, special issue on Digital Images*, 1999, Kluwer Academic Publishers, A paraître. 26 p.
- [BAC 99] BACHIMONT B., « Bibliothèques numériques audiovisuelles : des enjeux scientifiques et techniques », *Document numérique, Numéro spécial Les Bibliothèques Numériques*, vol. 2, n° 3–4, 1999, p. 219–242.
- [CHI 97] CHIARAMELLA Y., « Browsing and Querying: two Complementary Approaches for Multimedia Information Retrieval », *Hypertext – Information Retrieval – Multimedia HIM'97*, Dortmund, septembre 1997, Conférence invitée. Disponible sur l'Internet : <<http://lrb.cs.uni-dortmund.de/HIM97/Konferenzband/Chiarabella/Paper0.htm>>.
- [COR 98] CORDELLA L., FOGGIA P., SANSONE C., VENTO M., « Subgraph Transformations for inexact Matching of Attributed Relational Graphs », *Computing*, 1998, p. 43–52.
- [COT 99] COTTE D., « Le texte numérique et l'intériorisation des dispositifs documentaires », *Document numérique, Numéro spécial Les Bibliothèques Numériques*, vol. 2, n° 3–4, 1999, p. 259–279.
- [DAV 93] DAVIS M., « Media Streams: An Iconic Visual Language for Video Annotation », *Proceedings of the 1993 IEEE Symposium on Visual Languages*, Bergen, Norway, août 1993, IEEE Computer Society Press, p. 196–203.
- [LEN 98] LENAT D., *HAL's Legacy: 2001's Computer as Dream and Reality*, Chapitre « From 2001 to 2001: Common Sense and the Mind of HAL », MIT Press, 1998.
- [LIE 94] LIEBERMAN H., « A User Interface for Knowledge Acquisition from Video », *Proc. of AAAI-94 Conference*, Seattle, WA, août 1994, p. 527–534.
- [MAR 96] MARTIN P., ALPAY L., « Conceptual Structures and Structured Documents », *Proceedings of ICCS'96*, Australie, 1996, LNAI 1114, Springer-Verlag, p. 145–159.
- [NAN 95] NANARD J., NANARD M., « Adding macroscopic semantics to anchors in knowledge-based hypertext », *Int. J. Human-Computer Studies*, vol. 43, 1995, p. 363–382.
- [NEW 82] NEWELL A., « The knowledge level », *Artificial Intelligence*, n° 18, 1982, p. 87–127.
- [PRI 98] PRIÉ Y., MILLE A., PINON J.-M., « AI-STRATA: A User-centered Model for Content-based description and Retrieval of Audiovisual Sequences », *Int.*

Advanced Multimedia Content Processing Conf., Osaka, novembre 1998, LNCS 1554, Springer-Verlag, p. 328-343.

- [PRI 99a] PRIÉ Y., MILLE A., PINON J.-M., « A Context-Based Audiovisual Representation Model for Audiovisual Information Systems », *International and Interdisciplinary Conference on Modeling and using Context*, Trento, septembre 1999, LNAI 1688, Springer-Verlag p. 296–309.
- [PRI 99b] PRIÉ Y., MILLE A., PINON J.-M., « Modèle d'utilisation et modèles de tâches pour l'assistance à l'utilisateur basée sur l'expérience : le cas d'un système d'information audiovisuelle », *Ingénierie des Connaissances 1999*, Palaiseau, juin 1999, Afia, p. 21–30.
- [PRI 00] PRIÉ Y., LIMANE T., MILLE A., « Isomorphisme de sous-graphe pour la recherche d'information audiovisuelle contextuelle », *Reconnaissance de Formes et Intelligence Artificielle, RFIA'2000*, Paris, février 2000, A paraître, 10 p.
- [SCH 93] SCHREIBER A.-T., WIELINGA B.-J., BREUKER J.-A., Eds, « KADS: a Principled Approach to Knowledge-Based System Development », vol. 11 de *Knowledge-Based Systems*, Academic Press, 1993, 457 p.