

THÈSE

présentée devant

L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE LYON

pour obtenir

LE GRADE DE DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

par

Yannick PRIÉ

Modélisation de documents audiovisuels en Strates Interconnectées par les Annotations pour l'exploitation contextuelle

Soutenue le 15 décembre 1999 devant la commission d'examen

Jury :

Rapporteurs : Marie-France Bruandet
Marc Nanard

Examineurs : Bruno Bachimont
Françoise Chassaing
Alain Mille
Jean-Marie Pinon

Thèse préparée au Laboratoire d'Ingénierie des Systèmes d'Information, INSA de Lyon, et au Laboratoire Image Signal et Acoustique, CPE-Lyon

Mis en page avec la classe thloria.

THÈSE

présentée devant

L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE LYON

pour obtenir

LE GRADE DE DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

par

Yannick PRIÉ

Modélisation de documents audiovisuels en Strates Interconnectées par les Annotations pour l'exploitation contextuelle

Soutenue le 15 décembre 1999 devant la commission d'examen

Jury :

Rapporteurs : Marie-France Bruandet
Marc Nanard

Examineurs : Bruno Bachimont
Françoise Chassaing
Alain Mille
Jean-Marie Pinon

Thèse préparée au Laboratoire d'Ingénierie des Systèmes d'Information, INSA de Lyon, et au Laboratoire Image Signal et Acoustique, CPE-Lyon

Mis en page avec la classe thloria.

ECOLES DOCTORALES

➤ **MATERIAUX DE LYON**

INSAL – ECL -UCB. Lyon1 – Univ. De Chambéry – ENS

Responsable : Professeur A. HOAREAU, UCBL (Tél. : 04.72.44.85.66)

Formations doctorales associées :

- Génie des Matériaux (Pr. R. FOUGERES, Tél : 04. 72. 43. 81 .49)
- Matière condensée surfaces et interfaces (Pr. G. GUILLOT, Tél : 04.72.43.81.61)
- Matériaux polymères et composites (Pr. H. SAUTEREAU, Tél : 04.72.43.81.78)

➤ **MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE (MEGA)°**

Responsable : Professeur J. BATAILLE, ECL (Tél : 04.72.43.8079)

Formations doctorales associées :

- Acoustique (Pr. J.L. GUYADER, Tél : 04.72.43.80.80)
- Génie Civil : Sols, matériaux, structures, physique du bâtiment (Pr. P. LAREAL, Tél : 04.72.43.82.16)
- Mécanique (Pr. G. DALMAZ, Tél : 04.72.43.83.03)
- Thermique et Energétique (Pr. M. LALLEMAND, Tél : 04.72.43.81.54)

➤ **ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE (EEA)**

INSAL - ECL – UCB. Lyon1 – Univ. de Saint-Etienne

Responsable : Professeur G. GIMENEZ, INSAL (Tél : 04.72.43.83.32)

Formations doctorales associées :

- Acoustique (Pr. J.L. GUYADER, Tél : 04.72.43.80.80)
- Automatique Industrielle (Pr. SCAVARDA, Tél : 04.72.43.83.41)
- Dispositifs de l'électronique intégrée (Pr. P. PINARD, Tél : 04.72.43.80.79)
- Génie biologique et médical (Pr. I MAGNIN, Tél : 04.72.43.85.63)
- Génie électrique (Pr. J.P. CHANTE, Tél : 04.72.43.87.26)
- Signal, Image, Parole (Pr. G. GIMENEZ, Tél : 04.72.43.83.32)

➤ **ECOLE DOCTORALE INTERDISCIPLINAIRE SCIENCES-SANTE (EDISS)**

INSAL – UCB Lyon1 – Univ. de Saint-Etienne – Univ. Aix-Marseille2

Responsable : Professeur A. COZZONE, CNRS-Lyon (Tél 04.72.72.26.75)

Formations doctorales associées :

- Biochimie (Pr. M. LAGARDE, Tél : 04.72.43.82.40)
- Génie biologique et médical (Pr. I. MAGNIN, Tél : 04.72.43.85.63)

AUTRES FORMATIONS DOCTORALES

➤ **ANALYSE ET MODELISATION DES SYSTEMES BIOLOGIQUE**

Responsable : Professeur S. GRENIER, INSAL
Tél : 04.72.43.83.56

➤ **CHIMIE INORGANIQUE**

Responsable : Professeur P. GONNARD, INSAL
Tél : 04.72.43.81.58

➤ **CONCEPTION EN BATIMENT ET TECHNIQUE URBAINES**

Responsable : Professeur M. MIRAMOND, INSAL
Tél : 04.72.43.82.09

➤ **DEA INFORMATIQUE DE LYON**

Responsable : Professeur J.M. JOLION, INSAL
Tél : 04.72.43.87.59

➤ **PRODUCTIQUE : ORGANISATION ECONOMIQUE ET GENIE INFORMATIQUE POUR L'ENTREPRISE**

Responsable : Professeur J. FAVREL, INSAL
Tél : 04.72.43.83.63

➤ **SCIENCES ET TECHNIQUES DU DECHET**

Responsable : Professeur P. MOSZKOWICZ, INSAL
Tél : 04.72.43.83.45

Janvier 1998

Institut National des Sciences Appliquées de Lyon

Directeur : J. Rochat

Professeurs

S.	AUDISIO	PHYSICOCHIMIE INDUSTRIELLE
J.C.	BABOUX	GEMPPM*
B.	BALLAND	PHYSIQUE DE LA MATIERE
D.	BARBIER	PHYSIQUE DE LA MATIERE
G.	BAYADA	MODELISATION MATHEMATIQUE ET CALCUL SCIENTIFIQU
C.	BERGER (Mlle)	PHYSIQUE DE LA MATIERE
M.	BETEMPS	AUTOMATIQUE INDUSTRIELLE
J.M.	BLANCHARD	LAEPSI**
C.	BOISSON	VIBRATIONS ACOUSTIQUES
M.	BOIVIN	MECANIQUE DES SOLIDES
H.	BOTTA	EQUIPE DEVELOPPEMENT URBAIN
G.	BOULAYE	INFORMATIQUE
J.	BRAU	CENTRE DE THERMIQUE
M.	BRISSAUD	GENIE ELECTRIQUE ET FERROELECTRICITE
M.	BRUNET	MECANIQUE DES SOLIDES
J.C.	BUREAU	THERMODYNAMIQUE APPLIQUEE
J.Y.	CAVAILLE	GEMPPM*
J.P.	CHANTE	COMPOSANTS DE PUISSANCE ET APPLICATIONS
B.	CHOCAT	UNITE DE RECHERCHE EN GENIE CIVIL
B.	CLAUDEL	LAEPSI**
M.	COUSIN	UNITE DE RECHERCHE EN GENIE CIVIL
M.	DIOT	THERMODYNAMIQUE APPLIQUEE
A.	DOUTHEAU	CHIMIE ORGANIQUE
R.	DUFOUR	MECANIQUE DES STRUCTURES
J.C.	DUPUY	PHYSIQUE DE LA MATIERE
H.	EMPTOZ	RECONNAISSANCE DES FORMES ET VISION
C.	ESNOUF	GEMPPM*
L.	EYRAUD (Prof. Emérite)	GENIE ELECTRIQUE ET FERROELECTRICITE
G.	FANTOZZI	GEMPPM*
M.	FAYET	MECANIQUE DES SOLIDES
J.	FAVREL	GRUPE DE RECHERCHE EN PRODUCTIQUE ET INFORM/ DES SYSTEMES MANUFACTURIERS
G.	FERRARIS-BESSO	MECANIQUE DES STRUCTURES
Y.	FETIVEAU	GENIE ELECTRIQUE ET FERROELECTRICITE
L.	FLAMAND	MECANIQUE DES CONTACTS
P.	FLEISCHMANN	GEMPPM*
A.	FLORY	INGENIERIE DES SYSTEMES D'INFORMATION
R.	FOUGERES	GEMPPM*
F.	FOUQUET	GEMPPM*
L.	FRECON	INFORMATIQUE
R.	GAUTHIER	PHYSIQUE DE LA MATIERE
M.	GERY	CENTRE DE THERMIQUE
G.	GIMENEZ	CREATIS***
P.	GOBIN (Prof. émérite)	GEMPPM*
P.	GONNARD	GENIE ELECTRIQUE ET FERROELECTRICITE
M.	GONTRAND	COMPOSANTS DE PUISSANCE ET APPLICATIONS
R.	GOUTTE (Prof. Emérite)	CREATIS***
G.	GRANGE	GENIE ELECTRIQUE ET FERROELECTRICITE
G.	GUENIN	GEMPPM*
M.	GUICHARDANT	BIOCHIMIE ET PHARMACOLOGIE
G.	GUILLOT	PHYSIQUE DE LA MATIERE
A.	GUINET	GRUPE DE RECHERCHE EN PRODUCTIQUE ET INFORM/ DES SYSTEMES MANUFACTURIERS
J.L.	GUYADER	VIBRATIONS ACOUSTIQUES
J.P.	GUYOMAR	GENIE ELECTRIQUE ET FERROELECTRICITE
J.M.	JOLION	RECONNAISSANCE DES FORMES ET VISION
J.F.	JULLIEN	UNITE DE RECHERCHE EN GENIE CIVIL
A.	JUTARD	AUTOMATIQUE INDUSTRIELLE
R.	KASTNER	UNITE DE RECHERCHE EN GENIE CIVIL
H.	KLEIMANN	GENIE ELECTRIQUE ET FERROELECTRICITE
J.	KOULOUMDJIAN	INGENIERIE DES SYSTEMES D'INFORMATION
M.	LAGARDE	BIOCHIMIE ET PHARMACOLOGIE
M.	LALANNE	MECANIQUE DES STRUCTURES
A.	LALLEMAND	CENTRE DE THERMIQUE
M.	LALLEMAND (Mme)	CENTRE DE THERMIQUE
P.	LAREAL	UNITE DE RECHERCHE EN GENIE CIVIL
A.	LAUGIER	PHYSIQUE DE LA MATIERE
Ch.	LAUGIER	BIOCHIMIE ET PHARMACOLOGIE
P.	LEJEUNE	GENETIQUE MOLECULAIRE DES MICROORGANISMES

A.	LUBRECHT	MECANIQUE DES CONTACTS
Y.	MARTINEZ	INGENIERIE DES SYSTEMES D'INFORMATION
H.	MAZILLE	PHYSICOCHIMIE INDUSTRIELLE
P.	MERLE	GEMPPM*
J.	MERLIN	GEMPPM*
J.P.	MILLET	PHYSICOCHIMIE INDUSTRIELLE
M.	MIRAMOND	UNITE DE RECHERCHE EN GENIE CIVIL
N.	MONGEREAU (Prof. Emérite)	UNITE DE RECHERCHE EN GENIE CIVIL
R.	MOREL	MECANIQUE DES FLUIDES
P.	MOSZKOWICZ	LAEPSI**
P.	NARDON	BIOLOGIE APPLIQUEE
A.	NAVARRO	LAEPSI**
A.	NOURI (Mme)	MODELISATION MATHEMATIQUE ET CALCUL SCIENTIFIQUE
M.	OTTERBEIN	LAEPSI**
J.P.	PASCAULT	MATERIAUX MACROMOLECULAIRES
G.	PAVIC	VIBRATIONS ACOUSTIQUES
J.	PERA	UNITE DE RECHERCHE EN GENIE CIVIL
G.	PERRACHON	THERMODYNAMIQUE APPLIQUEE
J.	PEREZ (Prof. Emérite)	GEMPPM*
P.	PINARD	PHYSIQUE DE LA MATIERE
J.M.	PINON	INGENIERIE DES SYSTEMES D'INFORMATION
D.	PLAY	CONCEPTION ET ANALYSE DES SYSTEMES MECANQUES
J.	POUSIN	MODELISATION MATHEMATIQUE ET CALCUL SCIENTIFIQUE
P.	PREVOT	GROUPE DE RECHERCHE EN APPRENTISSAGE, COOPER. ET INTERFACES MULTIMODALES
R.	PROST	CREATIS***
M.	RAYNAUD	CENTRE DE THERMIQUE
J.M.	REYNOUARD	UNITE DE RECHERCHE EN GENIE CIVIL
E.	RIEUTORD (Prof. Emérite)	MECANIQUE DES FLUIDES
J.	ROBERT-BAUDOY (Mme)	GENETIQUE MOLECULAIRE DES MICROORGANISMES
D.	ROUBY	GEMPPM*
P.	RUBEL	INGENIERIE DES SYSTEMES D'INFORMATION
C.	RUMELHART	MECANIQUE DES SOLIDES
J.F.	SACADURA	CENTRE DE THERMIQUE
H.	SAUTEREAU	MATERIAUX MACROMOLECULAIRES
S.	SCARVARDA	AUTOMATIQUE INDUSTRIELLE
D.	THOMASSET	AUTOMATIQUE INDUSTRIELLE
M.	TROCCAZ	GENIE ELECTRIQUE ET FERROELECTRICITE
R.	UNTERREINER	CREATIS***
J.	VERON	LAEPSI**
G.	VIGIER	GEMPPM*
A.	VINCENT	GEMPPM*
P.	VUILLERMOZ	PHYSIQUE DE LA MATIERE
Directeurs de recherche C.N.R.S.		
Y.	BERTHIER	MECANIQUE DES CONTACTS
P.	CLAUDY	THERMODYNAMIQUE APPLIQUEE
N.	COTTE-PATTAT (Mme)	GENETIQUE MOLECULAIRE DES MICROORGANISMES
P.	FRANCIOSI	GEMPPM
J.F.	GERARD	MATERIAUX MACROMOLECULAIRES
M.A.	MANDRAND (Mme)	GENETIQUE MOLECULAIRE DES MICROORGANISMES
J.F.	QUINSON	GEMPPM
A.	ROCHE	MATERIAUX MACROMOLECULAIRES
Directeurs de recherche I.N.R.A.		
G.	BONNOT	BIOLOGIE APPLIQUEE
G.	FEBVAY	BIOLOGIE APPLIQUEE
S.	GRENIER	BIOLOGIE APPLIQUEE
Y.	MENEZO	BIOLOGIE APPLIQUEE
Directeurs de recherche I.N.S.E.R.M.		
A.F.	PRINGENT (Mme)	BIOCHIMIE ET PHARMACOLOGIE
I.	MAGNIN (Mme)	CREATIS***

GEMPPM* : Groupe d'étude métallurgie physique et physique des matériaux

LAEPSI** : Laboratoire d'analyse environnementale des procédés et systèmes industriels

CREATIS*** : Centre de recherche et d'applications en traitement de l'image et du signal

A Marie-Gabrielle

Remerciements

Mes remerciements vont tout d'abord aux membres du jury et en particulier aux rapporteurs : je remercie Marie-France Bruandet et Marc Nanard d'avoir bien voulu accepter de rapporter cette thèse dans des délais relativement courts, ainsi que des nombreuses marques d'intérêt qu'ils ont bien voulu porter à mes travaux.

J'ai eu le grand plaisir de compter dans mon jury Bruno Bachimont — dont j'avais téléchargé la thèse d'épistémologie il y a quelques années — et que j'ai eu la bonne surprise de retrouver ensuite à la direction de la recherche de l'INA. Ses connaissances et compétences dans de nombreux domaines en font un interlocuteur de choix, qui m'a toujours accueilli à l'INA avec grand intérêt.

Je remercie Françoise Chassaing — qui a piloté du côté du CNET le projet SESAME dans lequel la recherche s'inscrivait — d'avoir pu se libérer pour faire partie de mon jury ainsi que de ses efforts pour faire prendre en compte les résultats de la recherche au sein même de son équipe de développement au CNET.

Jean-Marie Pinon a piloté le projet SESAME du côté lyonnais et a dirigé ma recherche avec une grande efficacité tant administrative qu'humaine, et je l'en remercie beaucoup.

Alain Mille m'a encadré et supporté sans compter au long de ces trois années. De nos discussions houleuses — et loin des « consensus mous » — sont nés un grand nombre des concepts ici présentés. Qu'il trouve ici l'expression de ma gratitude pour ces quelques années de collaboration fructueuse et amicale.

Cyril Declair, du même projet, a partagé mon bureau pendant la plus grande partie de cette thèse, ce qui représente en soi une performance. L'équipe du laboratoire RFV a joué le jeu du projet et je tiens particulièrement à remercier Emmanuel Etiévent pour son enthousiasme, ainsi que Jean-Michel Jolion, qui a fait preuve d'une grande ouverture d'esprit et a accepté de relire la première partie de ce mémoire malgré un emploi du temps très chargé. Előd Egyed-Szigmond a rejoint notre équipe en octobre 1998 afin de travailler sur les Strates-IA. Je lui suis extrêmement redevable de sa grande efficacité dans la mise au point du prototype SESAME et de son enthousiasme, et je lui souhaite bonne chance pour la thèse récemment commencée. Je remercie enfin nos partenaires de France3 Rhône-Alpes-Auvergne et de l'INA Centre-Ouest qui nous ont toujours chaleureusement reçus.

Les mentions spéciales des remerciements concernent tout d'abord Tahar Limane, collègue enseignant de CPE-Lyon et spécialiste de la théorie des graphes, qui a accepté de consacrer de longs moments à discuter de l'algorithme d'instanciation. Je l'en remercie amicalement.

Gwendal Auffret, thésard de l'INA, rencontré à une journée consacrée à la sémantique linguistique est rapidement devenu un ami, et nous avons étroitement et joyeusement collaboré sur plusieurs articles. Je le remercie de son intérêt pour ma recherche et pour avoir accepté de relire, en tant que spécialiste MPEG-7, la partie de mémoire que je consacre à cette norme.

Igor Babou, de l'ENS Fontenay St-Cloud, docteur deux jours avant moi, m'a amicalement aiguillé vers les Ateliers Méthodologiques de l'Inathèque et je le remercie de nos longues discussions sur le tout et le rien sémiotique, en attendant de remettre celà.

Il convient également de remercier les divers professionnels de l'archivage audiovisuel qui ont bien voulu me recevoir ou discuter de leur pratique avec moi : Claire Corneloup, de l'INA, avec laquelle nous avons étudié quelques documents audiovisuels, mais aussi certains membres des Ateliers de l'Inathèque, telle Cécile Kattng de la BNF.

J'adresse enfin mes remerciements aux directeurs de laboratoires qui m'ont accueilli : M. Jourlin du LISA à CPE-Lyon et J. Kouloumdjian puis R. Laurini du LISI à l'INSA-Lyon, ainsi que mes amitiés à l'ensemble des membres de ces laboratoires qui m'ont entouré au cours de ces

quelques années. Citons plus spécialement l'ex-équipe du 404 ainsi que Mathieu Exbrayat dont j'ai le plaisir de partager le bureau.

Mention très spéciale à Ioannis Kanellos, de l'ENST Bretagne, avec lequel j'ai eu la grande chance d'« entrer en recherche » et qui m'a ouvert des horizons intellectuels insoupçonnés, lesquels réapparaissent dès qu'ils le peuvent, c'est à dire souvent.

Pour finir, mention de la famille à ma famille, que j'ai peu vue, un peu trop occupé par ce travail. J'espère que cela changera à l'avenir.

Table des matières

Introduction générale

1	Contexte de la recherche	17
1.1	Le projet SESAME	17
1.2	Problématique scientifique	17
1.3	Déroulement et principaux résultats	18
2	Plan du mémoire	19

Chapitre 1

Vers un système d'information documentaire idéal

1.1	Notion de document	23
1.1.1	Définition	23
1.1.2	Document numérique	25
1.1.3	Document multimédia	26
1.2	Système d'information documentaire	29
1.2.1	Qu'est ce qu'un système d'information documentaire?	29
1.2.2	Informatisation des SID	31
1.3	Vers des systèmes d'information documentaire idéaux	39
1.3.1	Indexation de documents multimédia	39
1.3.2	Utiliser l'information	41
1.3.3	Vers une indexation intelligente	44

Partie I Modélisation de documents audiovisuels numériques 47

Chapitre 2

Exploitation de DAV numériques dans un système d'information audiovisuelle

2.1	Document audiovisuel	50
2.1.1	L'image	50

2.1.2	Le flux d'images	51
2.1.3	Le son	52
2.1.4	Le montage audiovisuel : la véritable fabrique du document	52
2.1.5	Document audiovisuel	54
2.2	Document audiovisuel numérique	54
2.2.1	Formats de représentation de documents audiovisuels	54
2.2.2	Les apports du numérique	55
2.3	Systèmes d'information audiovisuelle	57
2.3.1	Utilisations de documents audiovisuels	57
2.3.2	Archives audiovisuelles	58
2.3.3	Fonctionnalités d'un SIAV	60
2.3.4	Annotation de documents audiovisuels numériques	61

<p>Chapitre 3</p> <p>Etat de l'art de la représentation de documents audiovisuels</p>

3.1	Des champs de recherche et des problématiques différentes	64
3.2	Caractéristiques d'annotation de documents audiovisuels	66
3.2.1	De la granularité de l'annotation	66
3.2.2	Primitives	69
3.2.3	Caractéristiques calculées de haut-niveau d'abstraction	70
3.2.4	Caractéristiques interprétées	73
3.3	Modèles de structuration	74
3.3.1	Structuration des annotations	75
3.3.2	Structuration implicite de documents audiovisuels	77
3.3.3	Structuration hiérarchique	78
3.3.4	Structuration par graphes	80
3.4	Fonctionnalités de systèmes d'information audiovisuelle	80
3.4.1	Indexation textuelle	80
3.4.2	Recherche par similarités et évolutions	81
3.4.3	Présentation de documents	82
3.4.4	Génération de documents	88
3.5	Conclusion	89

<p>Chapitre 4</p> <p>Réflexions sur la modélisation audiovisuelle</p>

4.1	Représenter pour exploiter	91
4.2	Sur les connaissances d'exploitation	93

4.2.1	L'image et le symbole	93
4.2.2	L'organisation des connaissances décrites: la structure	95
4.2.3	Organiser les connaissances de description	98
4.3	La notion de contexte dans l'audiovisuel et sa nécessaire prise en compte . . .	99
4.3.1	Contexte et systèmes d'information	99
4.3.2	Contexte et audiovisuel	101
4.4	Description et exploitation en contexte	103
4.4.1	Description de documents	103
4.4.2	Ecriture et lecture en contexte	104
4.5	Nécessités pour la modélisation audiovisuelle	105

Partie II Strates Interconnectées par les Annotations pour l'exploitation contextuelle de documents audiovisuels 109

**Chapitre 5
Modélisation de documents audiovisuels en Strates-Interconnectées par les annotations**

5.1	Annoter un flux audiovisuel	111
5.1.1	Appréhension de flux et objets d'intérêt	112
5.1.2	Dimensions d'analyse	112
5.1.3	Relations	113
5.1.4	Structuration des connaissances d'annotation	113
5.2	Présentation générale des Strates-IA	113
5.2.1	Graphes d'annotation	113
5.2.2	Connaissances d'annotation	120
5.2.3	Discussion	123
5.3	Une présentation plus formelle des Strates-IA	124
5.3.1	Objets des Strates-IA	125
5.3.2	Graphe orienté étiqueté d'objets	126
5.4	Conclusion de cette partie	128

**Chapitre 6
Manipulations de contextes dans les Strates-IA**

6.1	Contextes dans les Strates-IA	129
6.1.1	Un début de description	129
6.1.2	Une première approche du contexte	131

6.1.3	Définition du contexte	136
6.1.4	Graphes potentiels pour exprimer des contextes	137
6.2	Instanciation de graphes potentiels	140
6.2.1	Isomorphismes de sous-graphes partiels	141
6.2.2	Méthodes de recherche d'isomorphismes de sous-graphes	141
6.2.3	Notations et définitions	143
6.2.4	Algorithme de multi-propagation	148
6.2.5	Un exemple complet	150
6.2.6	Expérimentation et discussion	152
6.3	Outils de manipulation fondés sur les graphes potentiels	154
6.3.1	Graphes potentiels caractérisés	154
6.3.2	Dimensions d'analyse	159
6.3.3	Schémas de description	161
6.3.4	Graphes requêtes	166
6.3.5	Valences	169
6.4	Exploitation d'une base Strates-IA	170
6.4.1	Annoter — indexer	171
6.4.2	Rechercher	172
6.4.3	Naviguer	173
6.4.4	Analyser	175
6.4.5	Editer des documents, construire des documents pour la présentation	176
6.4.6	Extraire une documentation	176
6.5	Conclusion	177

Chapitre 7

Réalisations

7.1	Prototype de validation de l'algorithme de multi-propagation	179
7.2	Prototype SESAME	182
7.2.1	Principes généraux	182
7.2.2	Fonctionnalités	183
7.2.3	Evolutions	184

Chapitre 8

Vers une utilisation de l'expérience pour l'assistance à l'utilisateur

8.1	Modéliser la tâche d'un utilisateur pour aider à sa réalisation	188
8.2	Modéliser utilisation et tâches pour aider l'utilisateur	190
8.2.1	Définitions	191

8.3	Vers une généralisation de l'approche	195
8.4	Discussion : exploiter l'expérience	195
8.4.1	Les cas d'utilisation pour l'illustration et la validation	196
8.4.2	Apprentissage	196
8.4.3	Aide à l'utilisateur fondée sur l'expérience	197
8.5	Conclusion	198

Chapitre 9

Strates-IA : documents et connaissances

9.1	Structure, documents, connaissances et annotations	200
9.1.1	Evolution des structures de présentation	200
9.1.2	Structures sémantiques	200
9.1.3	Ajouter des connaissances par l'annotation	202
9.2	Descriptions documentaires	204
9.2.1	Langages de balises	205
9.2.2	Strates-IA et langage de balises	206
9.2.3	Arbres et graphes	208
9.2.4	MPEG-7	209
9.2.5	Approches orientées « connaissances »	217
9.2.6	Analyse des approches	223
9.3	Strates-IA et connaissances	226
9.3.1	Strates-IA et approches centrées sur les connaissances	226
9.3.2	Ecriture sur le flux	230
9.3.3	Gestion et évolution des connaissances Strates-IA	232
9.4	Conclusion	234

Conclusion générale

1	Bilan de la recherche	237
2	Discussion	239
3	Développements en cours	243

Glossaire	245
------------------	------------

Bibliographie	249
----------------------	------------

Introduction générale

1 Contexte de la recherche

1.1 Le projet SESAME

C'est sur la constatation que la puissance des ordinateurs, les capacités de stockage les normes de compression et les débits de réseaux permettraient dans un futur proche de diffuser et d'exploiter des documents audiovisuels numériques que le CNET-CCETT France Télécom (Centre National des Etudes en Télécommunication, Centre Commun d'Etudes en Télédiffusion et Télécommunications) a lancé en 1996 un appel d'offre dans la cadre d'une CTI (Concertation Thématique Informelle) sur le thème *Indexation et recherche par le contenu pour les services multimédias*.

Le projet SESAME (contrat 96-ME-17) est un des projets retenus et associe quatre laboratoires lyonnais¹, deux entreprises partenaires² ainsi qu'un partenaire technique³

SESAME (Système d'Exploration de Séquences Audiovisuelles et Multimédia enrichi par l'Expérience) se donnait pour objectif d'étudier les possibilités offertes par l'exploitation de documents audiovisuels numériques sous les angles suivants : *traitement d'image pour l'indexation* (RFV) ; *distribution et accès parallèles à des données audiovisuelles* (LIP) ; *bases de données audiovisuelles* (LISI) ; *aide à l'utilisateur fondée sur l'expérience des sessions d'exploitation d'un système de recherche d'information audiovisuelle* (LISA-LISI).

La recherche présentée dans ce mémoire s'est déroulée à partir du 20 novembre 1996 sous la direction conjointe de Alain Mille (LISA, CPE, équipe RàPC) et de Jean-Marie Pinon (LISI, INSA-Lyon) sur cette dernière problématique, avec un statut de chercheur contractuel de l'Insa de Lyon financé par le projet SESAME.

1.2 Problématique scientifique

Les objectifs du projet initial résument bien les enjeux scientifiques de la recherche menée, et le passage suivant tiré du dossier soumis au CNET dans le cadre de l'appel d'offre en définit les contours : « *Le challenge est de réussir à «indexer» ces informations pour les retrouver efficacement en fonction de demandes qui peuvent être multimodales et varieront d'un contexte à l'autre, voire d'un individu à l'autre. Pour tracer les contours plus précis de la recherche impliquée par*

1. Le laboratoire Reconnaissances des Formes et Vision (RFV) de l'Insa de Lyon; le Laboratoire d'Informatique du Parallélisme (LIP) de l'ENS-Lyon; le Laboratoire Image Signal et Accoustique (LISA) de CPE-Lyon; le Laboratoire d'Ingénierie des Systèmes d'Information (LISI) de l'Insa de Lyon.

2. France 3 Rhône-Alpes-Auvergne et l'INA (Institut National de l'Audiovisuel) Centre-Est

3. le Centre Inter-universitaire de calcul Scientifique et Médical (CISM).

un tel objectif général, il convient d'analyser les différentes facettes du problème à résoudre pour réaliser un système correspondant à notre projet. [...] Certaines caractéristiques brutes (entités géométriques, qualité, etc.) peuvent être, en partie, extraites de l'examen des images, formant un jeu de descripteurs «résumant» les propriétés «objectives» de la séquence audiovisuelle. Une sémantique externe (par exemple un titre, des mots-clés, un résumé, etc.) peut être ajoutée par un opérateur. Ces deux niveaux sémantiques sont associés avec l'information brute dans la base de données [...] Il peut s'agir de points de vue que l'on pourrait qualifier de «canoniques», c'est à dire correspondant à une classe générique d'utilisateurs, ou de points de vue individuels. Les premiers peuvent faire appel à la modélisation, tandis que les seconds nécessitent de mémoriser la sémantique individuelle liée à l'information dans ce contexte, ce qui implique d'analyser la manière dont la recherche y est menée. [...] Il existe potentiellement autant de systèmes d'indexations secondaires qu'il existe de contextes pour les sémantiques de contenu (canoniques ou individuelles) [...] Un index peut être vu comme une extension de la base de séquences audiovisuelles devant permettre de retrouver «efficacement» les séquences elles-mêmes [...] L'utilisateur exploite les possibilités des index grâce à un moteur de recherche qui l'assiste dans sa tâche. La recherche est interactive et le moteur de restitution permet de rendre compte de l'avancement de la recherche (aperçus par exemple) et bien entendu de fournir l'accès à l'information quand elle est déterminée sans ambiguïté. [...] Le moteur de recherche doit élaborer ces informations à partir des indications fournies par l'utilisateur et des informations mises en correspondances dans les index. ».

La problématique scientifique du travail de recherche s'articule donc autour de ces quelques problèmes :

- la représentation de documents audiovisuels de façon à prendre en compte leurs caractéristiques essentielles, telle que la temporalité, la composante fortement visuelle, la multiplicité des analyses possibles (traitement du signal pour le calcul de primitives de description de bas-niveau d'abstraction, analyse de haut-niveau conceptuel) ;
- l'indexation pour la recherche d'information et l'utilisation de documents dans le cadre de tâches multiples (recherche simple pour la visualisation, analyse, réutilisation, *etc.*), lesquelles ne sont pas toutes prévues à l'indexation, par des utilisateurs qui tous diffèrent : bref le partage de descriptions documentaires ;
- l'assistance aux diverses tâches d'exploitation des documents audiovisuels basées sur les descriptions des documents, et de façon plus générale l'exploitation de document fondée sur les connaissances documentaires, les rapports entre documents et connaissances.

1.3 Déroutement et principaux résultats

Il s'agissait donc de se placer au carrefour de plusieurs domaines de recherche (intelligence artificielle, recherche d'information documentaire, description documentaire, systèmes à base de connaissances), afin d'essayer d'appréhender de façon globale le problème.

La thèse s'est tout d'abord déroulée autour des problématiques de la recherche d'information, de la prise en compte de l'expérience dans le raisonnement à partir de cas, et de l'indexation audiovisuelle. Sur la constatation qu'il n'existait pas de modèle de représentation suffisamment adapté à notre volonté de gérer connaissances documentaires et expérience d'utilisation, nous avons mis au point le modèle des Strates Interconnectées par les Annotations (Strates-IA). Ce

modèle permet de résoudre le conflit entre les approches de segmentation *a priori* et de stratification. Il permet de plus de prendre en compte les contextes audiovisuels, en considérant que toute annotation participe d'une structure, laquelle est support d'annotation contextuelle. La suite de la thèse a consisté en la mise au point des outils d'exploitation contextuelle des Strates-IA et de l'algorithme fondamental d'instanciation, l'étude de la possibilité de stocker l'expérience d'utilisation de façon expliquée, ainsi que la comparaison avec d'autres systèmes émergents dans le cadre du couplage entre documents et connaissances. Un premier prototype validant l'approche par graphes et un deuxième permettant une exploitation des graphes de connaissances Strates-IA ont été mis au point. Un travail sur des aspects plus liés à l'utilisation de l'audiovisuel en sciences humaines a également été mené et a conduit à quelques publications.

Un résultat important de la recherche est le modèle des Strates-IA pour la représentation de documents audiovisuels. Ce modèle autorise en effet une description libre (« écrite ») des documents, sans accorder le moindre primat à une structure documentaire à partir de laquelle toute autre description devrait s'organiser. L'approche documentaire classique suppose en effet un type d'utilisation partagé par tous, et une manière de décrire qui ne l'est pas moins, ce qui d'une certaine manière contraint voire fige les protocoles d'indexation et de recherche. Dans notre cas, au contraire, l'objectif de réutilisation des descriptions dans le cadre de tâches variées nécessite un *cadre* de description homogène des contenus documentaires et une réutilisation *contextuelle* des descriptions. Le cadre fourni permet également d'envisager de façon raisonnée (c'est à dire selon un point de vue explicite) l'association des connaissances de haut et de bas-niveau d'abstraction.

Un second résultat important est la proposition d'outils et de mécanismes génériques d'exploitation contextuelle des Strates-IA. Les graphes potentiels qui s'instancient dans le graphe des Strates-IA représentent un outil de base à partir duquel tous les autres sont construits. L'algorithme de multi-propagation pour l'instanciation de graphes potentiels que nous proposons possède quelques caractéristiques intéressantes, telle que celle d'être *anytime* et d'être piloté par une heuristique simple qu'il est possible de redéfinir. Les schémas de description nous permettent de fournir des guides sur la manière de décrire, et s'expriment de manière très semblable aux descriptions elles-mêmes. Il est ainsi aisé de passer des descriptions (réellement utilisées) aux schémas de descriptions abstraits qui pourront alors être organisés et réutilisés. L'exploitation contextuelle des Strates-IA permet de plus de faire le lien entre la tâche et la volonté de l'utilisateur (non atteignables directement) et l'expression des celles-ci sous la forme de graphes potentiels c'est à dire de visées de contextualisation qui sont eux manipulables.

La proposition de compléter la description d'un système d'information documentaire à l'aide d'un modèle d'utilisation (par exemple celui des Strates-IA) et de modèles simplifiés mais explicites de tâches permet le stockage d'une expérience d'utilisation expliquée par des connaissances⁴ (et non brute) ouvre des voies prometteuses pour l'aide à l'utilisateur fondée sur l'expérience.

2 Plan du mémoire

Le chapitre introductif de ce mémoire *1- Vers un système d'information documentaire idéal* est consacré à une étude générale des systèmes d'information documentaire, qui nous permet d'aborder tour à tour les notions de documents numériques et multimédia, de recherche d'information documentaire et d'indexation. Nous proposons alors ce que seraient les caractéristiques d'un système d'information documentaire multimédia idéal, les tâches qu'il permettrait d'accom-

4. Par opposition à une trace « brute » sans référence à une base de connaissances.

plir, et définissons une indexation « intelligente » comme permettant de définir nativement des index manipulables comme des connaissances.

La première partie de ce document *I- Modélisation de documents audiovisuels numériques* est consacrée à une étude des documents audiovisuels et de leur représentation dans les systèmes d'information documentaire. Cette partie se compose des chapitres 2 à 4.

Le chapitre 2- *Exploitation de documents audiovisuels numériques dans un système d'information audiovisuelle* nous permet de faire connaissance avec le médium audiovisuel et ses principales caractéristiques, notamment le fait qu'un document audiovisuel soit monté. Nous étudions ensuite les apports du numérique à l'audiovisuel en terme de formats, avant de nous intéresser à l'utilisation des documents audiovisuels, aux archives audiovisuelles et aux fonctionnalités attendues d'un système d'information audiovisuelle (SIAV). Nous présentons enfin la notion d'annotation qui sera centrale à notre recherche.

Le chapitre 3- *Etat de l'art de la représentation de documents audiovisuels*, a un caractère essentiellement bibliographique. Après une introduction permettant de situer les différents champs de recherche intéressés par la modélisation audiovisuelle, nous organisons les caractéristiques d'annotation en trois niveaux dépendant de la manière dont elles ont été mises en place. Nous étudions ensuite un certain nombre d'approches de structuration de caractéristiques (implicites, hiérarchiques, en graphes), avant de présenter les fonctionnalités principales des SIAV que nous avons pu rencontrer.

Le chapitre 4- *Réflexions sur la modélisation audiovisuelle* est une mise en perspective de l'ensemble des notions vues dans la première partie. Nous discutons ainsi à nouveau le statut des descriptions et des connaissances d'exploitation d'un SIAV, d'une part au niveau des caractéristiques de représentation (où nous montrons qu'il est nécessaire de s'appuyer sur des descriptions symboliques), d'autre part, en ce qui concerne la structure d'organisation de ces caractéristiques (qui doivent prendre la forme la plus générale possible). Nous évoquons ensuite la notion de contexte, dans les documents audiovisuels, mais aussi comme contexte de tâche d'exploitation, et proposons de considérer toute tâche d'exploitation d'un SIAV comme tâche de description contextuelle de documents audiovisuels. Nous définissons alors la notion de description comme écriture en contexte d'annotations, et concluons ce chapitre en énumérant quelques unes des nécessités à notre sens fondamentales pour la modélisation audiovisuelle.

La seconde partie de ce mémoire *II- Strates-Interconnectées par les Annotations pour l'exploitation contextuelle de documents audiovisuels* présente notre modèle de description de documents audiovisuels ainsi que la manière dont il est possible d'exploiter contextuellement un système Strates-IA. Nous proposons également un cadre enrichi à la modélisation des systèmes d'information permettant de penser le stockage de l'expérience d'utilisation, avant de nous intéresser aux rapports entre documents et connaissances et de situer les Strates-IA dans ce contexte. Cette partie se compose des chapitres 5 à 9.

Le chapitre 5- *Modélisation de documents audiovisuels en Strates-Interconnectées par les Annotations* présente les bases de notre approche de description, qui consiste à repérer des *objets d'intérêt* dans le flux audiovisuel. Nous proposons alors de décrire les objets d'intérêt par des *éléments d'annotation* annotant des *unités audiovisuelles*. Les éléments d'annotation sont structurés

entre eux à l'aide d'un seul type de relation, la *relation élémentaire*. Une base de connaissances, ensemble organisé d'*éléments d'annotation abstraits*, permet de structurer les connaissances d'annotation. Au final, les unités audiovisuelles, éléments d'annotation, éléments d'annotation abstraits et leurs relations forment un unique graphe. Nous présentons alors de manière plus formelle ce graphe comme un graphe orienté étiqueté par des objets.

Le chapitre 6- *Manipulation de contextes dans les Strates-IA* présente un exemple de description en Strates-IA d'un document audiovisuel, ainsi que la notion de *contexte* dans le graphe Strates-IA, qui est défini comme extrémité de chemin⁵. Nous illustrons ce concept et définissons les *graphes potentiels* comme graphes génériques permettant de décrire des contextes. Les graphes potentiels *s'instancient* dans le graphe général, et cette opération correspond à une recherche d'isomorphisme de sous-graphes partiels. Nous présentons en conséquence un *algorithme de multi-propagation* pour la recherche d'instances, lequel est piloté par une heuristique simple et donne de bons résultats. La section suivante est consacrée à divers outils d'exploitation contextuelle des Strates-IA : *dimensions d'analyse*, *graphes potentiels caractérisés*, *schémas de description* pour pouvoir contraindre l'annotation, *valence* comme possibilités de relations pour les éléments d'annotation en contexte. La suite du chapitre présente de quelle manière il est possible d'utiliser les outils mis en place dans les diverses tâches d'exploitation d'un système d'information audiovisuelle : annotation, recherche, navigation, *etc.*

Le chapitre 7- *Réalisations* présente rapidement les deux prototypes construits autour des concepts des Strates-IA. Le premier nous permet de valider les algorithmes d'instanciations, mais souffre de possibilités d'exploitation limitées. Le second prototype implante graphiquement l'ensemble des notions que nous avons présentées et fournit une base solide d'expérimentation.

Le chapitre 8- *Vers une utilisation de l'expérience pour l'assistance à l'utilisateur* est pour nous l'occasion de proposer un cadre enrichi pour la modélisation de systèmes d'information à l'aide d'un *modèle d'utilisation* unique et de *modèles de tâches* décrivant les éléments du modèle d'utilisation mobilisés dans le cadre de tâches et sous-tâches. Il devient alors possible de mettre en place des *cas d'utilisation* rationalisés par les modèles de tâche et d'utilisation, mais aussi par des explications internes, non prévues à la base, sous la forme de relations. Nous présentons quelques pistes d'exploitation de l'expérience expliquée stockée dans les cas d'utilisation pour l'apprentissage et l'aide à l'utilisateur.

Le chapitre 9- *Strates-IA : documents et connaissances* est consacré au thème général des documents et des connaissances, et nous permet de compléter la bibliographie du chapitre 3. Nous étudions dans un premier temps l'évolution des structures documentaires en lien avec les structures de présentation, et proposons de considérer finalement l'ensemble des structures documentaires comme des structures de connaissances, utilisées dans des tâches quelconques, pouvant entre autres concerner la présentation. Nous étudions alors les descriptions documentaires en considérant deux tentatives de décrire des documents audiovisuels à l'aide de langages de balises. Nous situons les Strates-IA par rapport à ces approches et décrivons l'état actuel de la future norme de description de documents multimédia MPEG-7, avant de la comparer aux Strates-IA. Nous présentons ensuite cinq approches orientées « connaissances » pour la description de documents, et analysons ces approches sous l'angle de l'inférence comme opération sur des

5. Le noeud extrémité d'un chemin est déclaré comme étant en contexte avec le noeud de départ selon le chemin qui les réunit.

connaissances de description. Nous étudions enfin les Strates-IA sous l'angle des connaissances, présentons la notion d'inférence contextuelle et discutons la notion d'écriture sur le flux audiovisuel. Quelques propositions sur la gestion et l'évolution des connaissances Strates-IA viennent terminer ce chapitre.

La conclusion est l'occasion de présenter les perspectives d'application et de poursuite de la recherche entreprise.

Chapitre 1

Vers un système d'information documentaire idéal

Sommaire

1.1	Notion de document	23
1.1.1	Définition	23
1.1.2	Document numérique	25
1.1.3	Document multimédia	26
1.2	Système d'information documentaire	29
1.2.1	Qu'est ce qu'un système d'information documentaire?	29
1.2.2	Informatisation des SID	31
1.3	Vers des systèmes d'information documentaire idéaux	39
1.3.1	Indexation de documents multimédia	39
1.3.2	Utiliser l'information	41
1.3.3	Vers une indexation intelligente	44

Ce premier chapitre poursuit un double objectif.

Il s'agit d'une part de présenter un ensemble de concepts liés aux documents, aux documents numériques, aux documents multimédias, aux systèmes d'information documentaires, et à la recherche d'informations. Au cours du mémoire, nous pourrons ainsi référer à l'un ou l'autre de ces concepts pour les préciser ou les remettre en cause sous le couvert d'une définition minimale. Nous ne prétendons bien entendu pas épuiser les sujets que nous évoquerons dans ces quelques pages — plusieurs thèses n'y suffiraient pas.

Il s'agit d'autre part de proposer ce que serait pour nous un *système d'information documentaire idéal*, dans lequel êtres humains et machines coopéreraient à la résolution des problèmes des premiers. Il conviendra alors de garder en mémoire ce schéma générique idéal comme un sorte de « Graal » motivant la modeste quête présentée dans ce mémoire.

1.1 Notion de document

1.1.1 Définition

Dans sa définition la plus générale, un *document* est une trace de l'activité humaine. Si l'on élimine de cette définition un certain nombre d'objets (par exemple les documents archéologiques),

et qu'on se limite aux traces intellectuelles, alors on considérera qu'un document a été conçu dans l'objectif d'être interprété (lu, vu, écouté, visionné) par un certain nombre de personnes différentes de la ou les personnes qui l'ont mis en place. Un document est donc à un moment ou à un autre *mis à disposition* de son public potentiel. Par exemple une lettre sera envoyée à son destinataire, un livre sera édité et publié, un reportage d'émission télévisuelle sera diffusé.

[Auffret et Bachimont, 1999], étendant une définition de [Furuta, 1997], considèrent ainsi qu'un document « est une unité auto-suffisante représentant une contribution intellectuelle, et publié sur un média pour des raisons spécifiques. Un document exhibe, dans une certaine limite, une structure intentionnelle qui définit comment les éléments de son contenu sont organisés selon des axes dans l'objectif d'être interprétés par un lecteur comme témoignage de cet objectif original de publication ».

Ces définitions permettent de préciser un certain nombre de points et de définitions importants quand il s'agit de documents.

Par exemple un document existe en soi du fait de son *unité documentaire*, c'est à dire de ce qui fait qu'il est interprétable comme tel de façon auto-suffisante. Cette notion peut être rapprochée de celle de *textualité* qui est ce qui rend un texte irréductible à une suite de phrases [Rastier, 1995a].

D'autre part, un document est selon Bruno Bachimont « un objet matériel exprimant un *contenu*. Il est indissociable d'un *support matériel* (un écran, une feuille de papier, etc.), support d'inscription où un contenu est exprimé. Ce contenu s'exprime en une forme interprétable (lettres, formes iconiques, etc.) pour un lecteur. Cette forme peut être dite sémiotique dans la mesure où elle fait signe pour un lecteur d'un sens qui lui est adressé. Ce que l'on appelle contenu est l'abstraction que l'on peut faire des différents documents exprimant la même chose » [Bachimont, 1999].

Ensuite, un certain nombre de contraintes découlent du support d'inscription et de la forme sémiotique choisie, contraignant « l'expression du contenu et ses conditions de réception et d'interprétation. En particulier, deux types de contraintes émergent :

- « – une structure *logique* qui impose un ordonnancement (une « mise ensemble ») d'éléments de contenus exprimés selon un mode d'expression donné. C'est le choix par exemple d'énoncer une argumentation selon un ordre plutôt qu'un autre. Les structures logiques se formalisent partiellement par les définitions des types de documents que l'on peut rencontrer dans des formalismes comme SGML.

- « – une structure *matérielle* imposant une présentation (une « mise en forme » de ces éléments conformément à l'ordonnancement prescrit. C'est le choix par exemple d'une typographie, d'une mise en pages donnée, etc. Cette structure matérielle se formalise partiellement par la notion de feuille de style que l'on rencontre en traitement électronique des documents. »

[Bachimont, 1999].

Les différentes structures internes au document sont mises en place dans un *contexte de production*, dans le cadre d'une *pratique* donnée (par exemple la lettre médicale à une confrère, ou bien le texte de loi). C'est le *genre* du document qui détermine les contraintes, par exemple les structures logiques utilisées. C'est le *contexte de réception* du document qui détermine quant à lui la manière dont les structures du document seront utilisées et interprétées.

Au final, « un document n'est pas une simple juxtaposition des éléments qu'il contient. Conçu dans un contexte de production donné, consulté dans un contexte de réception, réalisé conformément à des structures logique et matérielle, un document en tant que tel conditionne l'interpréta-

tion de ses éléments. Un document cristallise dans ses structures et contextes des connaissances et des prescriptions interprétatives pesant sur la lecture à effectuer » [Bachimont, 1999].

Le document pouvant être interprété en dehors d'un contexte de réception canonique (correspondant à celui prévu lors de la production), la reconnaissance d'éléments de structures — jouant le rôle d'autant d'indices — permet de déterminer le genre du document, lequel à son tour impose une lecture canonique. Inversement, l'analyse du document en vue d'en repérer les structures peut permettre de reconstituer un contexte de production *a posteriori*. Plus simplement, il est possible de constituer un *genre*, pour peu que certaines structures se répètent entre divers documents. Les beaux jours de l'analyse littéraire reposent entre autres sur ces constatations. Plus pragmatiquement, la Text Encoding Initiative (TEI [Burnard et Sperberg-McQueen, 1994]) a eu pour objet de déterminer quelles étaient les principales définitions de structure logique permettant de formaliser tous les textes en SGML.

Nous reviendrons plus loin sur ces problèmes, fondamentaux dès qu'il s'agit de représenter des documents structurés en tant que tels dans les systèmes d'information, ou bien encore quand le problème est justement de structurer des documents audiovisuels.

1.1.2 Document numérique

Avec l'avènement du numérique, le document quitte son support matériel natif — par exemple le papier pour l'écrit ou les bandes analogiques pour les documents audiovisuels — pour une représentation unique et binaire dans les mémoires des ordinateurs.

La mise en place d'un document numérique peut se faire de deux manières différentes : le document peut être numérisé à partir de son support premier — par exemple un document papier est scanné ; ou bien le document est directement pensé et écrit sous forme numérique par la médiation par exemple d'un logiciel de traitement de texte.

Quelle que soit la manière dont est obtenu le document numérique, sa forme matérielle lui transmet désormais les propriétés propres au numérique : le document peut et même doit être l'objet de *calculs* afin d'être manipulé et appréhendé.

B. Bachimont, dans la tradition du travail mené à l'UTC sous l'impulsion de B. Stiegler, analyse dans [Bachimont, 1999] le document numérique. Il considère ainsi que celui-ci « met en jeu les distinctions suivantes :

- « – le *support d'enregistrement du document* : le document numérique est codé par une structure discrète [...] ;
- « – la *forme d'enregistrement* : l'enregistrement numérique se fait selon un format de codage qui permet d'établir la correspondance entre le document tel qu'il doit être consulté et lu, et sa structure interne ;
- « – le *support d'appropriation du document* : [...] l'enregistrement numérique est en mémoire et ne peut être accédé directement. Il faut qu'un calcul produise à partir de l'enregistrement une représentation lisible ou intelligible pour un utilisateur [...] sur un support où l'utilisateur accède *directement* au document [...] sans devoir faire appel à une médiation calculatoire, c'est à dire à l'exécution d'un programme. Ce support est un support d'appropriation car l'utilisateur s'approprie le contenu du document. Les supports d'appropriation sont par exemple l'écran, des hauts parleurs, le papier ;
- « – la *forme d'appropriation du document* : la représentation affichée sur le support d'appropriation respecte une structure ou une forme telle qu'elle est directement accessible par

l'utilisateur. Cette forme est donc la forme d'appropriation permettant à un utilisateur de s'approprier le contenu. Ce dernier accède au contenu [par le recours] à la médiation d'un apprentissage de la forme d'appropriation. Par exemple, il a appris à lire, et la forme alphabétique d'affichage du document est une forme d'appropriation intelligible pour lui. Quand il y a plusieurs formes d'appropriation, le document est multimédia. Par exemple l'audiovisuel est multimédia car il mobilise l'image, la musique, le bruit et la parole. L'image peut aussi être multimédia si elle comporte des textes et des structures iconiques ;

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

Il s'ensuit qu'un document, en tant que mise ensemble d'éléments de contenu dans le cadre d'une structure matérielle, est « nécessairement ce qui est consulté dans le cadre d'une forme d'appropriation sur un support d'appropriation. Il ne correspond pas à l'enregistrement interne, contrairement à ce que la locution de « document numérique » pourrait laisser entendre [...] le document numérique, considéré sur le support d'enregistrement, n'est pas un document, mais une ressource à partir de laquelle peuvent être calculés autant de documents, c'est à dire de formes d'appropriation sur un support d'appropriation. » [Bachimont, 1999] L'enregistrement interne d'un document word est par exemple inintelligible, et peut faire appel à plusieurs fichiers (images externes) tandis que le document affiché sous une forme ou une autre dans l'interface du traitement de texte, ou bien imprimé, l'est.

Le document numérique est donc ce qui est présenté et reconnu sous forme intelligible en tant que tel par un utilisateur, après un calcul effectué sur une forme numérique. Quatre opérations entre ces deux différentes formes sont dès lors envisageables [Bachimont, 1999]. La *projection* est l'opération du calculable à l'intelligible visant à passer d'une forme interne à une forme d'appropriation. L'*abstraction* est l'opération inverse (par exemple une numérisation). La *transformation* est une opération du calculable au calculable ne conduisant pas à une forme d'appropriation, par exemple une reconnaissance de caractères sur un document numérisé, ou bien le calcul d'un fichier PostScript à partir d'un enregistrement Word. La *navigation* enfin, opération de l'intelligible à l'intelligible permet de passer d'une forme d'appropriation à une autre, qui ne dépend pas obligatoirement du même enregistrement.

L'analyse que nous venons de largement présenter permet de faire la part des choses entre les inévitables fichiers d'un système d'exploitation informatique et la manière dont le contenu de ces fichiers est présenté à l'utilisateur. Elle permet surtout de remettre le document à sa vraie place, comme *ensemble organisé de contenus prescrivant leur propre interprétation par un être humain*.

1.1.3 Document multimédia

Après avoir évoqué la notion de document numérique dans sa généralité, il convient d'avoir un mot sur celle de *document multimédia*.

De la façon la plus générale, un média permet de faire médiation entre une source et une destination. Nigay et Coutaz [Nigay et Coutaz, 1996] montrent dans le cadre d'une étude consacrée aux interactions homme-machine que plusieurs définitions plus particulières de *média* se rencontrent dans la littérature :

– certaines se concentrent sur le point de vue purement technologique : un média est toute

forme logique, purement technique, capable de véhiculer de l'information (un haut-parleur, un message électronique).

- d'autres définitions occultent ces dimensions matérielles et techniques, et considèrent les liens prédominants du média avec les capacités perceptuelles et cognitives humaine :
 - un média peut ainsi être un ensemble de qualités perceptives couplé au dispositif sensoriel humain nécessaire à la perception de ces qualités. Par exemple le couplage des qualités du son et de l'audition forment un média, qui peut s'assimiler à la notion de modalité sensorielle telle que comprise en psychologie.
 - un média peut être un système représentationnel [Alty, 1991]. Par exemple un texte, un graphe sont des systèmes représentationnels qui font appel à l'interprétation et à la compréhension.

Cette définition se rapproche de la définition vue plus haut, pour laquelle est multimédia tout ce qui mêle des formes d'appropriation différentes, c'est à dire différents systèmes représentationnels appris. Cette approche est celle que l'on retiendra dans ce document.

- la troisième définition de la notion de média, plus obscure, concilie perception et support matériel. Un média est alors un dispositif qui a la capacité matérielle de véhiculer des informations d'une certaine substance organisée en une forme [Bourguet, 1992]. Le média associé a substance et forme, mais n'a pas de capacité d'interprétation. Le message conceptuel d'une œuvre lui échappe ainsi.

Le *multimédia* est alors à chaque fois défini comme faisant appel à au moins deux médias différents, et les définitions de média variant, on conçoit que varient celles de multimédia, surtout dans le contexte numérique.

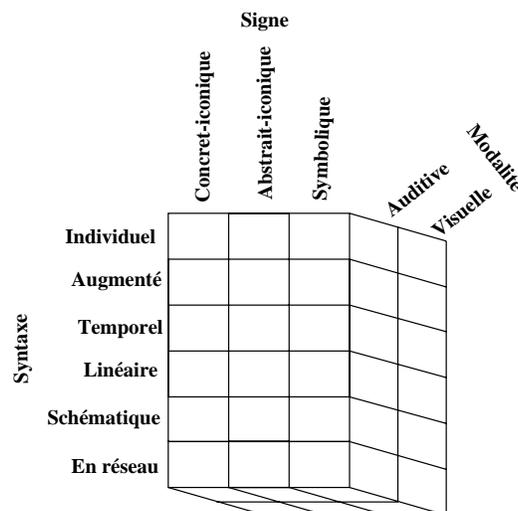


FIG. 1.1 – Les trois axes du modèle de Purchase : signe, syntaxe et modalité

Négligeant l'approche technologique, Purchase [1998] définit un texte¹ comme composé de signes dans un système de représentation également appelé média. Trois axes sont alors mis en

1. L'auteur prend ici le terme texte est dans son sens le plus large, une image peut alors être considérée comme un texte qu'on lit.

place (voir figure 1.1). La *nature* des signes utilisés tout d'abord, qui peuvent être concrets-ictoniques (une photographie), abstrait-ictoniques (un panneau dessiné indiquant des chutes de pierres), ou bien symbolique (quand il n'existe plus de lien perceptuel entre le symbole et ce qu'il représente, par exemple un mot ou un feu rouge). L'*arrangement* de ces signes ensuite : individuels (simple objet), augmentés (par exemple une photographie découpée), temporels (son ou vidéo), linéaires (un enchaînement d'images), ou bien schématiques (un diagramme arborescent d'organisation) ; catégories auxquelles elle ajoute la mise en réseau possible grâce au numérique. La *modalité* enfin, auditive et visuelle principalement².

Considérant les différentes possibilités de combinaison suivant ces trois axes, il est alors possible de décrire les documents existants ; par exemple la bande dessinée fait partie de l'abstrait ictonique linéaire dans la modalité visuelle.

Il est surtout possible de proposer un certain nombre de définitions de la communication multimédia. La première concerne « la production, la transmission et l'interprétation d'un texte-composite quand au moins deux des textes-composants font appel à des systèmes de représentation différents ». On retrouve là la définition la plus générale de multimédia, déjà entrevue. La deuxième définition fait appel à au moins deux modalités perceptives, cette définition s'accorde bien avec la conception répandue qu'est multimédia « ce qui fait appel à de l'image et du son ». Les troisième et quatrième définitions insistent sur la mise en réseau (la délinéarité, ou plutôt la multilinéarité déjà évoquée dans [Holm Nelson, 1965]) autorisée par l'informatique, et sur l'interactivité possible avec l'utilisateur. On retrouve dans cette définition tout ce qui est « *hyper-* » que l'on trouve dans les *hypertexte* et les *hypermédia*, ainsi que la notion de *navigation*. La dernière définition s'attache à ce qu'un au moins de textes-composants soit linéaire et temporalisé : ceci correspond à la considération qu'est multimédia ce qui comporte entre autres du son et/ou de l'image animée.

Les diverses définitions du multimédia et de la communication multimédia données par Purchase dans le cadre d'un système intégré permettent de couvrir les différentes définitions généralement proposées.

Il nous semble cependant utile d'insister sur le fait que les documents multimédias numériques sont redevables à l'informatique sur au moins deux points. Le premier concerne la possibilité de mêler de façon aisée plusieurs documents ou morceaux de documents de formes d'appropriation différentes. Par exemple un texte est présenté sur l'écran accompagné d'une image ou d'un document vidéo. L'important est ici que l'ordinateur est chargé de gérer la présentation des documents, lesquels sont éventuellement temporalisés : un texte défilera par exemple de droite à gauche avant d'être remplacé par une image au bout de dix secondes, tandis qu'une musique sera jouée d'un bout à l'autre de la présentation. L'assemblage, le calcul de la présentation échoit à la machine à chaque occurrence du document devant un lecteur. Cet assemblage se réalise d'après une description des opérations (structure de plan de présentation) qu'il y a lieu de réaliser à partir d'enregistrements de documents ou de morceaux de documents différents, ayant des formes d'enregistrement et d'appropriation différents.

Le deuxième point fondamental est de rendre opératoire la possibilité de *liens* entre parties d'un même document ou de documents différents, autorisant une nouvelle interactivité de l'utilisateur par rapport au document. La vieille notion d'hypertexte pressentie il y a bien longtemps par V. Bush [1945] est remise au goût du jour et devient effective, ouvrant des voies insoupçonnées dans les rapports de l'humanité et de l'information. Les réseaux *hypertextes* permettent

2. Il convient d'y ajouter le toucher, qui est appelé à connaître un avenir intéressant. Ainsi les aveugles peuvent lire du braille généré par un ordinateur, ce qui sera également bientôt possible sur le WEB, tandis que des vêtements équipés de vibreurs permettront dans le futur de ressentir les documents musicaux (mais ne seront-ils pas alors autre chose?) directement au niveau de la peau.

de naviguer d'une partie de document textuels à l'autre, tandis que les *hypermédi*as étendent la notion à tout type de document. Il devient possible de cliquer sur un texte et de jouer un enregistrement musical, ou même de naviguer d'une image à l'autre. Le World Wide Web illustre l'importance toujours croissante qu'est amené à prendre dans nos vies ce type de réseaux.

Ce petit tour des définitions du multimédia achevé, nous pouvons tenter une définition synthétique d'un document multimédia numérique, qui correspond plus ou moins alors à la quatrième définition de Purchase : *est document multimédia un document gérant de façon interactive à l'aide de liens une présentation de multiples morceaux de documents ou documents entiers, faisant appel à des formes d'appropriation et d'enregistrement multiples, le statut effectif de document lui étant donné par sa cohérence d'ensemble, laquelle découle de l'origine humaine du document*³.

1.2 Système d'information documentaire

Après avoir rapidement présenté dans la partie précédente ce que sont les documents en général et étudié les conséquences de l'avènement du numérique sur ceux-ci, notamment les documents multimédias, il convient de s'intéresser à la notion de *système d'information documentaire* (SID). Nous en décrivons d'abord rapidement le cadre le plus général hors informatique, avant d'étudier les conséquences de son intégration dans un SID numérique en terme de manipulation et d'utilisation de documents numériques.

1.2.1 Qu'est ce qu'un système d'information documentaire?

Définition

Par définition, un système d'information documentaire est un système permettant d'exploiter une collection de documents. Par exploitation, nous entendons principalement toutes les opérations liées à l'archivage de documents et à leur recherche par un utilisateur en vue d'une utilisation. Deux types de systèmes d'information documentaires peuvent se distinguer : ceux destinés à la conservation et au patrimoine, par exemple les bibliothèques et les dépôts légaux, et ceux plus liés à une utilisation plus immédiate et balisée, comme les services de documentation d'entreprises. La politique de choix des documents à conserver ou à écarter dépend bien entendu des objectifs de la collection.

Historiquement, la fonction de gestion d'un ensemble de documents a échu à des spécialistes — documentalistes ou conservateurs — qui doivent d'une part stocker les documents et en assurer la pérennité, d'autre part en rendre l'accès possible. Le développement de *systèmes d'index* a été rendu nécessaire dès que le nombre des documents a dépassé les faibles capacités de mémorisation humaine.

Indexation de documents

Nous définissons ici de façon très générale — mais non historique — un index comme *quelque chose permettant d'accéder à autre chose*. Le processus d'indexation consiste alors en la mise en place d'objets permettant d'accéder à d'autres objets. Appliqué au contexte documentaire, les objets cibles de l'indexation peuvent être :

- les documents eux-mêmes ;

3. Cette définition pourrait aussi convenir au terme *hypermédi*a, mais ce terme nous semble voué à disparaître, nous lui préférons donc *multimédia*.

- les parties de documents (le chapitre 3, la page 27, l'illustration de la page 34) ;
- des ensembles de documents (tous les journaux télévisés de France 2).

Afin de pouvoir accéder à l'objet d'indexation, il y a lieu d'identifier celui-ci de façon non ambiguë, donc de lui associer un *identifiant*. L'identifiant d'un document peut être par exemple son numéro d'arrivée dans la collection. L'identifiant d'un ensemble de documents peut être le numéro associé à une étagère, ou bien un ensemble d'identifiants de documents (ou au moins un moyen d'y accéder). L'identifiant d'une partie de documents est le plus souvent défini à partir de l'identifiant du document et d'un identifiant interne. Cet identifiant interne peut être explicite (le chapitre 3) ou bien créé au besoin (la page 27, l'image correspondant à la 4ème minute du reportage).

Nous posons qu'un *identifiant représente un index en soi à partir du moment où il est interprétable dans une pratique d'exploitation du document donnée*. Par exemple même un simple numéro d'ordre d'arrivée peut être utilisé pour rechercher des documents récents. Un numéro de page d'un ouvrage donné est un index de cette page dans le contexte de la manipulation de cet ouvrage : il y signifie un emplacement donné (le milieu par exemple). En fait, un identifiant représente un index à partir du moment où il est exprimé dans une forme sémiotique interprétable par un utilisateur, pour peu que celui-ci en ait la clé, c'est à dire les règles et le contexte de production.

A l'identifiant — index de base — est en général associé le véritable index, c'est à dire la description de l'objet de l'indexation. Ainsi, une émission hebdomadaire (*i.e.* l'ensemble de ses occurrences) peut être indexée à l'INA par une fiche-collection décrivant son contenu général et son dispositif (type de scènes, progression de l'émission). Un document sera simplement décrit dans une bibliothèque en terme d'auteur, de titre et d'un certain nombre de mots-clé ; une partie de document pourra avoir le même type de description.

Bachimont [1999] définit l'indexation dans une bibliothèque comme *la paraphrase d'un contenu en une forme sémiotique interprétable permettant l'exploitation du contenu indexé dans la cadre d'une pratique donnée*.

Il s'agit donc de mettre en place des *documents de description* (puisqu'ayant contenu et forme matérielle, l'index a un statut documentaire *per se*) décrivant le contenu des documents, d'ensemble ou de parties de documents et leur servant d'index. Les documents de description peuvent bien entendu être eux-mêmes à leur tour l'objet d'une indexation. Par exemple un article de journal pourra servir d'index à un livre. Les index sont valables dans une pratique donnée, c'est à dire que leur création se fait dans un contexte précis, en vue d'*au moins une* utilisation prévue explicitement. Ainsi, indexer un document par une fiche regroupant nom d'auteur, titre, et un ensemble de mots-clé implique une utilisation de l'index suivant ces catégories.

L'interprétation d'un contenu et sa réécriture dans une pratique donnée sont en fait un moyen d'exploiter celui-ci. La génération elle-même du document produit une structure logique de mise ensemble de contenus (chapitres, listes, paragraphes), prescrivant l'interprétation mais pouvant également servir d'index. A partir du moment où un document possède une structure, y compris implicite, par exemple découlant de sa linéarité, il est possible d'utiliser cette structure pour y accéder, et d'y adjoindre un contenu sémantique dans le cadre d'une pratique.

A *maxima*, il est donc possible de dire qu'est index de contenu documentaire tout ce qui est utilisé comme tel, que cela ait été mis en place avec cette volonté ou non. Cette extension de la notion d'index nous sera bien entendu utile quand nous évoquerons le cas des documents numériques.

Accès aux documents

L'accès aux documents peut se faire pour l'utilisateur du système d'information documentaire des façons suivantes (ici décrites sous forme non électronique) :

- soit par l'intermédiaire du documentaliste, auquel il pose une requête ou décrit un besoin, à charge pour ce dernier de l'interpréter et de consulter les index afin de proposer un certain nombre de documents ;
- soit directement en se promenant à sa guise dans la collection de documents (cas des salles en libre-service d'une bibliothèque) ;
- soit enfin en accédant aux index mis à disposition (sous forme de classeurs de fiches triées par auteurs ou par titres par exemple), la recherche effective du document choisi se résumant à utiliser l'identifiant permettant de l'atteindre.

Il va de soi que toute recherche peut se dérouler suivant une ou plusieurs des modalités présentées. Par exemple il est possible, après avoir sélectionné un document, de chercher aux alentours si d'autres documents ne seraient pas intéressants (dans le cadre d'une recherche sur un thème particulier).

Les trois accès que nous venons de présenter, pour classiques qu'ils soient n'en servent pas moins de référence à nombre de métaphores utilisées en recherche d'information, particulièrement sur le Web. Il convient donc de les avoir en mémoire lorsque l'on évoque le passage au numérique d'un système d'information documentaire. Nous nous intéresserons désormais aux systèmes « tout numériques ».

1.2.2 Informatisation des SID

L'intégration de l'informatique dans les SID se déroule d'abord à deux niveaux :

- au niveau des index : historiquement, les fiches de description ont par exemple été numérisées dans des bases de données permettant leur interrogation sur plusieurs champs qui n'étaient auparavant pas accessibles directement⁴ ;
- au niveau des documents : les documents peuvent être numérisés ou produits sous forme électronique, et sont dès lors accessibles directement sous cette forme, les identifiants devenant expressions électroniques d'accès aux documents ou parties de documents (par exemple une URL).

Il apparaît en fait que c'est à partir du moment où les documents sont sous forme électronique que la pratique change : il devient possible de réaliser des *calculs* sur leur contenu. Cela est ainsi déjà le cas avec la possibilité de recherche (*i.e.* de calcul) dans les documents fortement structurés que sont les index, de correspondances entre les requêtes et les représentants de documents.

Mais le numérique autorise désormais une exploitation directe du document au niveau de son support d'enregistrement (décrit par sa forme d'enregistrement), y compris au niveau de l'indexation : celle-ci peut être automatisée, voire réalisée au besoin. Une recherche plein-texte est ainsi possible : si un mot-clé en soi ne représente pas un index, un mot-clé associé à un système de recherche (comme Altavista) est un index de l'ensemble des documents fournis par ce système en réponse à la requête utilisant avec le mot-clé. De la même manière, une balise HTML

4. Quoique certains anciens évoquent le système *keydex* qui permettait de réaliser des *ET* sur des fiches-index perforées .

peut désormais représenter un index. En d'autres termes, n'importe quelle partie du support d'enregistrement d'un document électronique est d'une part susceptible d'être accédée comme contenu documentaire, d'autre part de servir d'index pour ce contenu.

L'indexation est en effet partout dans le numérique, et lui est consubstantielle. Tout calcul est manipulation des données et d'index permettant d'y accéder, les index étant bien évidemment eux-mêmes de données⁵. Cela permet à [Bachimont, 1999] de conclure que « tout système numérique est un système documentaire et mobilise une indexation ». En fait, un système numérique est un système documentaire dès qu'il en existe une indexation. Par exemple, un code binaire généré par un compilateur est un document indexé par son source si celui-ci est disponible, et de toutes façons par une table des symboles.

Différentes approches de SID numériques L'approche précédemment évoquée nous amène à considérer toute *organisation de données* comme un document, il est alors possible de considérer tous les types de systèmes d'information comme des SID.

Smaïl [1994] considère que les systèmes d'information pouvaient se considérer sous les trois aspects suivants :

- un *système de gestion de bases de données* est un ensemble d'outils logiciels offrant un ensemble de fonctions : la définition des données dans un format explicite, la manipulation de ces données (création, recherche, modification et suppression), la gestion de l'intégrité, de la sécurité et de la confidentialité des données. Par exemple, un SGBD relationnel stocke des données en tables. Chaque table correspond à un type d'objet et a autant de colonnes qu'on a choisi d'attributs pour définir l'objet. Chaque ligne ou enregistrement représente une occurrence du type d'objet. Poser une requête d'interrogation consiste à fixer certaines valeurs d'attribut pour certaines relations et le résultat d'une recherche dans une base de données est alors l'ensemble des enregistrements s'appariant *exactement* avec la requête posée. Il va alors de soi que décrire un document index comme un ensemble d'attributs en relation autorise son stockage et sa recherche dans une base de données.
- un *système de recherche d'information*, comme son nom l'indique, met l'accent sur la fonction de *recherche d'information* (nous reviendrons sur cette notion plus loin). Aucune restriction n'est faite, en théorie, sur le type d'élément d'information considéré. En réalité, on parle de SRI lorsqu'il s'agit de traiter des documents ayant une existence propre et à caractère stable (mises à jour exceptionnelles) comme des journaux, des articles bibliographiques, des partitions musicales, des images... L'information réellement manipulée par le système se réduit le plus souvent à un représentant de chaque document, affecté pendant l'opération d'*indexation*. Le résultat de la requête consiste en un ensemble de références vers les documents susceptibles d'intéresser l'utilisateur.
- un *système à base de connaissances* (SBC) a pour objectif de répondre à des questions de l'utilisateur. Il utilise en général principalement une base de connaissances structurée, ainsi qu'un ensemble de règles d'inférence permettant de déduire des faits correspondant à des réponses. Les faits, les règles, les traces de raisonnement documentent la façon dont un problème a été résolu, c'est à dire les connaissances qui ont été mobilisées. On peut donc à juste titre considérer un SBC comme manipulant et agissant sur des documents.

5. On notera qu'à ce niveau il peut y avoir désémantisation presque totale des identifiants de données, c'est à dire qu'un être humain devra avoir une connaissance totale du système et de l'application pour pouvoir les utiliser en tant que tels : c'est par exemple le cas dans des séances de débogage les plus ardues.

Ajoutons à ces systèmes d'information les systèmes d'exploitation qui permettent la manipulation et la gestion de documents dans des structures organisées. Ranger par exemple un ensemble d'images dans un répertoire en leur donnant des noms explicites revient à mettre en place un système d'information documentaire appliqué aux images, dont les index sont les noms de fichiers.

L'évolution des systèmes d'information nous semble tendre vers une plus grande intégration et une bien moins grande séparation entre domaines. Ainsi, certains systèmes de gestion de bases de données sont dits déductifs, c'est à dire qu'il permettent de faire des inférences comme les SBC. D'autres prennent d'entrée en compte les liens et la navigation [Amann *et al.*, 1994]. Les bases de données semi-structurées ont vu le jour afin de s'adapter aux documents et aux structures documentaires (par exemple les sites Web) : la notion de schéma de base de données disparaît au profit d'une organisation des données en arbre ou en graphes suivant celle des documents ou des ensembles de documents. Ajoutons à celà la notion d'interactivité, d'abord propre aux SRI, qui est désormais prise en compte dans des SBC beaucoup moins rigides que par le passé. Les documents multimédias, interactifs par eux-mêmes et objets des SID ajoutent aux possibilités offertes de recherche et d'exploitation interactives. Il nous semble donc que les cloisonnements entre domaines de recherche devraient, si ce n'est disparaître, du moins changer de position et favoriser l'émergence de nouveaux systèmes de gestion de documents. Nous reviendrons plus largement sur quelques-uns de ces aspects plus loin dans le rapport, il nous semble cependant utile de mettre d'ores et déjà en exergue les changements en cours. Cela s'accorde également avec notre volonté d'essayer de définir un système d'information documentaire idéal.

Fonctions des SID numériques : Les fonctionnalités d'un système d'information documentaire sont les suivantes :

- *stockage, gestion de l'intégrité et des accès* aux documents : toutes fonctions basiques nécessaires à la bonne marche d'un système utilisé en vraie grandeur ;
- *transmission* des documents et des requêtes si le système fonctionne en réseau : gestion du transfert des différentes informations entre clients et serveurs, éventuellement répartis, gestion de la qualité de service ;
- *indexation des documents* : intégration des documents dans le système et description dans son *format de représentation*, en fonction du type d'unité documentaire choisi, traitement automatique ou non (par exemple mise en place d'index à partir d'une recherche statistique en texte intégral) ;
- *recherche des documents* : gestion des requêtes et de la fonction de correspondance entre celles-ci et les index de documents, ordonnancement éventuel des documents proposés en fonction de leur pertinence ;
- *présentation des documents* : celle-ci peut être éventuellement calculée à partir de la requête, se présenter sous la forme de résumés, *etc.* ;
- *édition de documents* : génération automatique ou gérée par l'utilisateur de nouveaux documents formés à partir de documents du système, modification de documents et gestion de leur cycle de vie ;
- *navigation* dans la base de documents et dans les documents eux-mêmes : participe également de la recherche, mais est souvent présentée à part, liée à l'interface graphique générale du système ;

- *aide à l'utilisateur*: gestion de connaissances sur la tâche de l'utilisateur, aide en cours de sessions de recherche, *etc.*

Nous reviendrons par la suite sur les fonctionnalités les plus importantes présentées ici. En effet, si le découpage est grossier, il permet cependant de présenter les aspects auxquels nous allons nous intéresser, et que nous raffinerons notamment dans le chapitre consacré aux systèmes d'information audiovisuelle.

Nous définissons pour toute la suite de ce mémoire la *représentation* d'un document comme la manière dont celui-ci est modélisé dans le système documentaire, et la *présentation* d'un document comme la manière dont celui-ci est présenté à l'utilisateur.

Remarquons alors que le format de représentation des documents est un des aspects les plus fondamentaux et les plus critiques d'un système d'information documentaire, car il contraint au moins tout à la fois l'indexation, la recherche, la présentation, la navigation documentaires. La *modélisation d'un document* dans le système, sa structuration seront donc les points-clés, et doivent être soigneusement prises en compte dès la conception d'un système.

Recherche d'information

La *Recherche d'Information* (RI) est un champ d'études historiquement organisé autour des documentalistes et des institutions chargées de gérer un grand nombre de documents, principalement textuels (tandis que le champ des bases de données se structurait principalement autour des informaticiens et de la gestion des systèmes d'information de l'entreprise). Les méthodes et les concepts en vigueur dans la Recherche d'Information dépendent fortement de ses origines historiques, et sont plus adaptées aux systèmes d'information documentaires que les méthodes issues de bases de données, car plus centrées sur les besoins des utilisateurs. Par exemple les notions de reformulation de requêtes, de pertinence utilisateur, de *besoin d'information* proviennent de la RI.

Van Rijsbergen, héraut de la recherche d'information la définit ainsi [van Rijsbergen, 1986] : « The user expresses his information need in the form of a request for information. Information retrieval is concerned with retrieving those documents that are likely to be relevant to his information need as expressed by his request. It is likely that such a retrieval process will be iterated, since a request is only an imperfect expression of an information need, and the documents retrieved at one point may help in improving the request used in the next iteration ».

Le premier point mis en exergue dans cette définition concerne la manière d'indexer un document (principalement textuel), d'exprimer une requête, et de calculer comment répondre à cette requête. Le deuxième point fait appel aux notions de cycle de recherche d'information et d'interaction avec l'utilisateur.

Indexation et calcul de correspondance dans les SRI. [Smaïl, 1994] définit quatre dimensions de l'indexation dans un Système de Recherche d'Information (SRI). Les descripteurs⁶ considérés sont ici de simples mots-clé :

- *l'unité d'indexation* concerne le choix de ce qu'il y a à indexer. On peut ainsi indexer un document de façon individuelle, ou bien un ensemble de documents (par exemple des séries d'images) quand le fond documentaire est trop grand. La prise en compte de la structure du document, sous-jacente encore récemment dans un texte numérisé, mais désormais explicitée se révèle alors d'importance [Chiaromella, 1997].

6. Comme tout descripteur d'un document peut en servir d'index, nous emploierons indifféremment ces deux termes dans la suite.

- le *degré de structuration* décrit l'organisation des descripteurs du document : cela peut aller de l'indexation par juxtaposition (ou indexation à plat), qui consiste à placer des mots-clés les uns à la suite des autres, jusqu'à la mise en place de structures de représentation de connaissances, en passant par de simples rubriques de différents types. Ces types d'indexation peuvent être appelés à rôle ou structurées.
- le *degré de relief* part du constat qu'un document pouvant illustrer certains concepts plus fortement que d'autres, il peut être souhaitable de mettre en relief certains descripteurs en leur affectant un poids correspondant à leur importance. L'indexation est alors dite pondérée.
- les *vocabulaires d'indexation* peuvent être classés en deux grandes catégories qui sont les vocabulaires *libres* (construits *a posteriori*) et les vocabulaires *contrôlés* (thesauri), ces derniers étant construits *a priori* afin de remédier en apportant de la cohérence d'ensemble aux problèmes issus d'une part de la polysémie des mots de la langue, et d'autre part de l'accroissement anarchique de la taille du vocabulaire d'indexation. Une procédure régulière de mise à jour du thésaurus est généralement prévue.

On pourrait rajouter à ces dimensions une cinquième qui tiendrait au *type de l'analyse* effectuée afin de mettre en place l'indexation : des systèmes manuels aux systèmes automatiques qui ne gardent que les mots du titre ou du résumé d'un document, il est possible d'ajouter par exemple des termes fondamentaux non présents dans celui-ci (par exemple le domaine scientifique d'un article pointu non cité dans celui-ci). L'indexation automatique, utilise des méthodes mathématiques afin d'extraire les termes statistiquement les plus représentatifs d'un texte, alors supposés les plus pertinents (TF/IDF — Term Frequency, Inverse Document Frequency — par exemple [Salton et Mac Gill, 1983].).

Une fois la base d'indexation construite, l'interrogation est la fonction principale des SRI. Elle offre à l'utilisateur les moyens d'exprimer son besoin selon un modèle de requête. L'étape suivante est alors la *mise en correspondance* entre la requête d'une part, et un document d'autre part. Une requête contient des critères décrivant les caractéristiques souhaitées des documents recherchés et le modèle de requête n'est pas indépendant du modèle de document choisi. La *fonction de correspondance* a pour rôle d'établir une base de comparaison entre les deux.

Passons ici très rapidement en revue quelques modèles classiques de la recherche d'informations. Malgré leur simplicité, ces modèles restent les plus utilisés, en combinaison avec des mécanismes permettant de pallier leur pauvreté d'expression (pour des études plus détaillées, voir à ce sujet [Salton et Mac Gill, 1983, Salton et Buckley, 1990]).

Le *modèle vectoriel* représente un document D_i par un vecteur de dimension n représentant un ensemble de descripteurs ou mots-clé : $D_i = (d_{i1} d_{i2} \dots d_{in})$ où n est le nombre de descripteurs connus et d_{ij} représente le poids affecté au descripteur j dans le document D_i . Une requête est de la même manière exprimée par un vecteur dans l'espace des descripteurs : $R = (r_1 r_2 \dots r_n)$. La mesure dans l'espace des descripteurs (supposé euclidien) du cosinus entre un vecteur document et le vecteur requête est une mesure de similarité typique liée à ce modèle.

La puissance du modèle vectoriel réside dans sa simplicité conceptuelle et de mise en œuvre. Documents et requêtes sont exprimés de la même manière, et la mesure de similarité permet de classer simplement les documents retrouvés en fonction de leur pertinence vis-à-vis de la requête. Plusieurs problèmes importants subsistent : l'orthogonalité implicite de l'espace de représentation suppose une indépendance entre les termes, ce qui est une hypothèse très forte. D'autre part, la limitation de l'expressivité du modèle reste très gênante.

Le *modèle booléen* est basé sur l'utilisation de la logique de Boole pour proposer une représentation des requêtes, et repose sur une représentation classique des documents à base de mots-clé. Trois types d'opérateurs (*et*, *ou*, *non*) servent à lier les critères de recherche formant une requête, ce qui permet d'y répondre en appliquant simplement ces opérations logiques sur des ensembles de documents extraits de listes-inverses. L'expressivité du modèle est supérieure à celle du modèle vectoriel, puisqu'il permet de retrouver tout sous-ensemble particulier d'une collection de documents. Le modèle booléen peut être pondéré (c'est par exemple le modèle utilisé par Altavista⁷).

Dans les *modèles logiques*, un document est considéré comme pertinent s'il implique logiquement la requête (ce qui est par exemple trivial dans le modèle booléen). Les modèles logiques fournissent un cadre unificateur pour la recherche d'informations [van Rijsbergen, 1986] et permettent de prendre en compte toutes sortes de connaissances structurelles sur les documents, les contenus multifacettes, les représentation des connaissances, l'inférence, *etc.* Cette approche associe une grande puissance d'expression à une gestion uniforme des connaissances, mais souffre de limitations liées à sa complexité théorique et pratique, ainsi qu'à la difficultés qu'il y a à mettre en place des modèles opérationnels liés à des indexations symboliques complexes (graphes conceptuels par exemple)⁸.

Cycle de recherche d'information. La forme de représentation de l'indexation, des requêtes et la fonction de correspondance permettent de fournir un certain nombre de documents candidats à l'utilisateur. Un *cycle de recherche d'information* présente les différentes étapes d'une recherche :

- la *formulation* de la requête est l'étape d'expression par l'utilisateur de son besoin d'information dans la forme imposée par le système, cette étape peut être aidée par le système, par exemple si le vocabulaire est contrôlé, il y a lieu de fournir à l'utilisateur un accès au thésaurus ;
- la *recherche* est l'étape, dévolue au système, d'appariement de la requête avec les index de documents à sa disposition, les documents réponses peuvent être éventuellement classés suivant leur pertinence supposée ;
- la *présentation* est le calcul par le système de la manière dont il va proposer à la visualisation de l'utilisateur les documents retrouvés, soit directement, soit en lui donnant accès aux identifiants.
- la *reformulation* de la requête en fonction des résultats obtenus permet à l'utilisateur de préciser l'expression du besoin d'information.

Evaluation. C'est dans le domaine des SRI textuels que se sont mis en place les concepts liés à l'*évaluation* des systèmes d'informations.

La *pertinence* recouvre des notions différentes selon que l'on se place du point de vue de la machine ou de l'utilisateur. Ainsi, du point de vue du système, la pertinence est la correspondance entre l'énoncé d'un besoin d'information (une requête) et un document, c'est-à-dire le point auquel le document recouvre la matière de l'énoncé du besoin. Le problème pour le concepteur du système est alors d'anticiper tous les besoins auxquels le SRI devra répondre.

Du point de vue de l'utilisateur, la pertinence dépend de l'utilité de chaque document que lui présente le SRI. Ainsi, la pertinence système et la pertinence utilisateur peuvent différer quand

7. Moteur de recherche bien connu <http://www.altavista.com>.

8. Y. Chiaramella, *Approche et modèles en Recherche d'Information*, cours Inforsid'99, Toulon, Juin 99.

un document correspond — du point de vue du système — parfaitement à la requête, tandis que l'utilisateur peut n'en avoir que faire (par exemple parce qu'il en connaît déjà parfaitement le contenu). L'utilité d'un document pour l'utilisateur ne peut être mesurée qu'à travers les jugements que celui-ci émet lorsque le SRI lui présente celui-là. Elle dépend naturellement du contexte, *i.e.* de facteurs aussi variés que le but poursuivi par l'utilisateur ou que le contexte socioculturel dans lequel est menée la recherche.

Plusieurs concepts de mesure ont été mis en place afin d'évaluer la pertinence du système de recherche d'information :

- le *rappel* mesure la proportion de documents pertinents retrouvés par rapport au nombre total de documents pertinents dans la collection.
- le *silence* est la proportion complémentaire du rappel, c'est-à-dire la proportion de documents pertinents que le système n'a pas retrouvés. Un silence important peut par exemple être dû à une mauvaise indexation des documents, ou encore à une recherche trop stricte ou peu adaptée à la requête.
- la *précision* mesure la proportion de documents pertinents retrouvés parmi l'ensemble des documents retrouvés.
- le *bruit* est la mesure complémentaire de la précision, et donne la proportion de documents non pertinents parmi ceux proposés à l'utilisateur par le système.

Plusieurs remarques doivent venir compléter ces définitions. Ainsi, le calcul de ces mesures nécessite de connaître *a priori* quels sont les documents pertinents pour une requête, ce que seul un utilisateur est capable de dire. De plus, l'accord de deux utilisateurs différents n'est pas obligatoire : une requête que l'un estime satisfaite peut être considérée différemment par l'autre. Également, si la précision reste mesurable à travers les jugements de pertinence de l'utilisateur, le rappel est plus problématique : comment évaluer quels sont les documents pertinents non fournis par le système ? Il s'agit alors de mettre en place des bases de tests et des jeux de requêtes parfaitement connus afin de pouvoir tester les performances de systèmes différents.

Interaction La gestion de l'interaction de l'utilisateur avec les systèmes de recherche d'information en est devenue une composante naturelle et obligatoire.

Une première interactivité consiste en la possibilité de reformulation conduisant éventuellement à des mécanismes de *bouclage de pertinence* (*relevance feedback* en anglais) [Harman, 1992, Salton et Buckley, 1990]. Celui-ci est né de la constatation que d'une part l'indexation était en général imparfaite et d'autre part que l'utilisateur avait de grandes difficultés à formuler dès la première tentative la « bonne » requête, ce qui se traduit par un décalage entre la fonction de pertinence du SRI et celle de l'utilisateur. L'idée est alors de prendre en compte la pertinence utilisateur pour améliorer les performances du système tout en tenant compte de ses performances passées. La recherche d'information passe alors par une suite d'étapes indépendantes au statut de processus *itératif*, dans lequel se met en place une véritable coopération permettant par un jeu de reformulations de la requête d'aboutir à un résultat satisfaisant pour l'utilisateur. À partir d'une première requête, le système fournit à l'utilisateur un ensemble de documents dont l'utilisateur évalue la pertinence (par exemple *bonne*, *mauvaise* ou *sans opinion*), ce qui conduit à une reformulation automatique de la requête qui tient compte de ce retour de l'utilisateur. Cette nouvelle requête fournit alors un nouvel ensemble de documents, à nouveau évalué, et ainsi de suite jusqu'à satisfaction de l'utilisateur.

On remarquera qu'alors que dans le bouclage de pertinence la reformulation de la requête est automatique, certains systèmes utilisent une autre forme de formulation interactive mais non automatique qui consiste à fournir à l'utilisateur toutes indications utiles à la (re-)formulation (nombre d'occurrence du terme dans la base, utilisation du terme, termes proches dans un thésaurus, *etc.*). A noter également que si [Chua et Ruan, 1995] considèrent que cette technique est utile pour améliorer les résultats d'une session de recherche, elle reste peu concluante dans son utilisation pour modifier les représentations des documents dans la base, ce qui correspondrait à une utilisation à long terme, probablement à cause du caractère très subjectif des jugements des utilisateurs.

Un autre niveau d'interaction, rendu possible par l'apparition du multimédia (tel que nous l'avons défini), concerne la *navigation* dans un espace documentaire. D'abord apanage de la communauté hypertexte, la navigation a été popularisée, et est désormais intégrée à la Recherche d'Information.

L'espace documentaire de navigation peut être construit *a priori*, par exemple, un modèle de recherche d'informations extrême est celui qui a cours dans les systèmes hypermédia : les requêtes en sont absentes, tandis que toute la recherche se fait lors de la navigation dans la base de documents (le Web ou un système de fichier sur une machine). Certains documents peuvent également être *construits au besoin*, le plus souvent comme manière d'organiser et de présenter les résultats d'une requête à l'utilisateur, l'accès aux documents numériques étant ainsi direct et immédiat. Il est également possible de considérer la mise en place automatique d'hyperliens dans les documents. La construction d'un ensemble de liens peut alors s'apparenter à un type d'indexation particulier permettant d'associer de façon explicite à un document un *voisinage*. [Dalamagas, 1998] met par exemple en place un système de reconstruction de liens entre articles de journaux sur un certain nombre de similarités.

Recherche d'informations et connaissances

Il est possible selon [Simonnot, 1996] de regrouper les connaissances prises en compte dans un SRI en trois classes principales : connaissances sur les documents d'abord, sur les concepts du domaine de l'application ensuite, sur les utilisateurs enfin.

Les *connaissances sur les documents* sont en fait les index de ces documents, et *explicitent* des connaissances contenues dans et sur les documents, telles qu'elles ont été interprétées pendant l'indexation.

Les *connaissances sur les concepts du domaine de l'application* concernent le plus souvent le vocabulaire d'indexation et la manière d'organiser celui-ci en indexation structurée. L'organisation des connaissances de description est en effet nécessaire afin de guider la description des documents, que ce soit en phase d'indexation ou de recherche. On mettra dans cette catégorie essentiellement les thésaurus, qui permettent dans le cadre d'un vocabulaire contrôlé de regrouper les différents termes utilisés ainsi que certaines relations entre ces termes, telles que la *synonymie*, les relations *spécifique/générique* ou encore la *méréonymie* (relation partie-tout). Certains systèmes, tels *I³R* ou encore *RUBRIC* (on trouvera les références les concernant dans [Simonnot, 1996]) ont une connaissance du domaine plus précise et représentent un certain nombre de concepts par des *schémas*, lesquels permettent d'inférer un certain nombre de connaissances de la requête de l'utilisateur. L'organisation des connaissances du domaine peut donc varier d'une simple organisation de termes en thésaurus à un véritable organisation en base de connaissances (réseaux sémantique ou treillis de Galois [Carpineto et Romano, 1996]) en vue d'inférences préétablies à la conception.

Les *connaissances sur les utilisateurs du système* concernent tout d'abord ce qu'il est possible de savoir sur les besoins d'information des utilisateurs auxquels le système va avoir pour objectif de répondre. Ensuite, pour chaque utilisateur particulier, des connaissances peuvent être mises en place comme profils ou modèles d'utilisateurs que le système pourra alors créer (incrémentalement ou par questionnaire) et utiliser afin de répondre au mieux aux requêtes. Les connaissances sur le besoin d'information de l'utilisateur peuvent également être considérées. Par exemple [Cluzeau-Ciry, 1988] définit différents besoins d'information dans une banque d'images correspondant à différentes stratégies de recherche :

- la *demande précise* naît quand l'utilisateur sait parfaitement ce qu'il cherche, voire connaît le document dont il a besoin. Le documentaliste doit alors chercher précisément.
- la *demande exploratoire* naît quand l'utilisateur veut se faire une idée d'une collection donnée sans *a priori*. Il s'agit alors de lui proposer des extraits jugés représentatifs de la base.
- la *demande thématique* est destinée à illustrer un thème. Le type de raisonnement alors suivi par l'utilisateur est un raisonnement par association d'idées stimulé par la visualisation des documents.
- le *besoin connotatif* concerne la recherche par exemple d'une certaine ambiance. Les termes utilisés sont alors fortement abstraits et subjectifs. la requête doit alors être reformulée en remplaçant ces termes par d'autres plus concrets convenant à la recherche dans la collection.

1.3 Vers des systèmes d'information documentaire idéaux

Nous avons jusqu'ici présenté rapidement un grand nombre de concepts liés aux documents et aux systèmes d'information documentaires numériques. Nous avons notamment vu que les documents les plus étudiés étaient — pour des raisons historiques — les documents textuels, et qu'un grand nombre de concepts découlaient directement de ce type de documents. Nous avons évoqué ici ou là l'utilisation de documents hypermédias et les différentes évolutions qu'il était possible de considérer, en termes de connaissances ou d'aide à l'utilisateur.

Il convient maintenant d'essayer de tracer un tableau synthétique de ce qu'est ou de ce que peut être un système d'information documentaire multimédia « idéal », regroupant les récentes évolutions liées à la recherche d'informations en général, mais faisant en fait appel à un certain nombre de domaines de recherche d'une manière ou d'une autre en relation avec les documents multimédia (et potentiellement, il y en a beaucoup). Nous analyserons d'abord l'évolution des notions d'indexation et des requêtes liées, puis situerons la recherche de documents dans le cadre de l'utilisation d'information et de la coopération personne-machine dans le cadre d'une tâche. Nous concluons enfin en définissant ce que serait une « indexation intelligente ».

1.3.1 Indexation de documents multimédia

Nous nous plaçons tout d'abord dans le cadre d'un document numérique multimédia faisant au moins appel à une forme d'appropriation non textuelle (image ou son par exemple).

L'indexation doit alors prendre en compte différents médias : texte, image, musique par exemple. Dans la lignée de l'indexation de documents textuels, l'approche la plus standard consiste à décrire un document quelconque avec une notice bibliographique. Celle-ci peut être mise en place pour répondre aux visées de l'institution (par exemple une agence de presse indexera ses images suivant son propre format). Il est également possible d'utiliser les fiches de

descriptions standard mise en place par la communauté des documentalistes, par exemple le Dublin Core⁹ et ses 15 champs (Titre, Créateur, Sujet, *etc.*) permettant de décrire minimalement tout document trouvé sur le Web¹⁰.

Il est également possible de considérer le document comme un signal sur lequel sont calculées des caractéristiques, exprimées dans un langage de description. Répondre à une requête posée dans le même langage revient à calculer une *similarité* entre la requête et les index.

Sur une image par exemple, calculer un histogramme de couleur revient à extraire du signal brut un ensemble de composantes couleurs, lesquelles sont considérées comme description de l'image. Une requête consiste alors en la description d'un histogramme par l'utilisateur, et en la comparaison de celui-ci avec ceux-là, un ordre sur les résultats de la fonction de similarité permettant de proposer une suite ordonnée d'images solutions.

Trois remarques sont ici nécessaires. En premier lieu, on ne traite que du signal, c'est à dire que le niveau de sens atteint par les descripteurs est celui d'un *résultat de calcul*. Il est alors nécessaire de connaître l'algorithme d'extraction pour pouvoir les interpréter. Deuxièmement, il devient possible de fournir comme requête un document (de la même forme que celui que l'on cherche, par exemple une image), à charge pour le système d'en extraire les descripteurs pour former une requête. La similarité entre document requête et documents réponses ne résulte toujours alors que d'une similarité calculée, laquelle peut correspondre plus ou moins bien à une similarité au sens de l'utilisateur, dans le cadre d'une tâche donnée. Troisièmement, le fait d'extraire des mots-clé d'un texte (comme déjà évoqué en 1.2.2) procède de la même démarche, c'est à dire que le texte est considéré comme signal sur des éléments duquel (les mots) un traitement statistique est réalisé. Même si le niveau symbolique n'est pas atteint (on en reste au calcul), l'adéquation du système fonctionnel de la langue (les lettres, les mots, les textes) à la représentation machine fait qu'il est possible d'obtenir de bons résultats, car la machine manipule les même éléments que l'être humain.

Nous reviendrons plus largement sur ces question dans le chapitre 4, retenons pour l'instant que deux modes d'indexation peuvent être distingués :

- l'indexation manuelle ou semi-automatique, dans laquelle un utilisateur (un documentaliste par exemple) réalise ou valide l'indexation, c'est à dire *interprète* le document (niveau « sens »). Les requêtes sont alors également au niveau symbolique.
- l'indexation automatique, dans laquelle les descripteurs résultent d'un calcul, et n'ont pas été interprétés (niveau « signal »). Les requêtes peuvent être de deux types : niveau symbolique *du calcul* (interrogation par l'utilisateur dans les termes de description des caractéristiques), ou bien *par l'exemple*, en fournissant un document duquel la requête va pouvoir être extraite automatiquement.

L'organisation d'un document multimédia consiste à y définir sa structure, c'est à dire à décrire la « mise ensemble » des différents éléments qui le composent. Les langages documentaires, d'abord prévus pour des documents principalement textuels (enrichis avec l'image, *i.e.* destiné à l'impression papier), décrivent la structure matérielle du texte et/ou sa structure logique. Une liste peut être par exemple une structure logique d'organisation d'un ensemble de faits (niveau logique), ou bien simplement un ensemble de textes précis qu'il convient de placer les uns sous

9. Voir <http://purl.org/dc/>.

10. Le moteur de recherche Altavista indexe également sous forme d'ensemble de mots-clé pondérés tous les documents du Web, en tenant désormais compte de leurs types (web, images, vidéos, sons), et l'utilisateur peut spécifier le type de document qu'il cherche.

les autres, légèrement décalés vers la droite, avec des puces bien définies (niveau matériel). SGML [Goldfarb, 1990] est un métalangage permettant de définir de tels langages et autorisant virtuellement toutes descriptions de documents.

Avec le multimédia, destiné à lecture sur un poste de travail multimédia, un niveau de structuration supplémentaire est mis en place pour décrire les interactions entre les éléments du documents, leurs mouvements par exemple, ainsi que les interaction possibles entre le document et l'utilisateur. A la structuration de présentation s'ajoute donc une structuration d'utilisation interactive de navigation, que nous pouvons appeler *structuration multimédia*.

S'ajoutent à cela les structurations de documents multilingues, par exemple les points de correspondance entre parties équivalentes dans des langues différentes, ainsi que la problématique de la multiplicité des sources, c'est à dire de la *répartition* du document.

Un certain nombre de problèmes se posent quant à la prise en compte de cette structuration au niveau de l'indexation. Si le problème par exemple de l'accès à une *partie* d'un document, pouvait se poser déjà pour les documents textuels, c'est bien le développement des documents multimédias et des normes de structuration qui a rendu son étude souhaitable et effective (car l'opérationnalisation devenait possible). Il faut en effet pouvoir désigner une partie de document (avec un identifiant) pour pouvoir l'indexer en tant que telle, donc faire appel à une structure rendant possible cette désignation.

La structuration multimédia peut être prise en compte à plusieurs niveaux.

Le premier niveau consiste en une mise en place de l'indexation guidée par la structure : en fonction de la position structurelle d'un morceau de document textuel, on accordera par exemple plus ou moins de poids aux mots-clé qui en sont extraits. On pourra également décrire une image à l'aide de termes d'indexation extraits de textes auxquels elle est liée, ce qui est par exemple le cas dans Altavista.

Le second niveau consiste en une utilisation de la structure pour l'indexation : on peut décider que toute partie de document devra être indexée pour elle-même, c'est à dire bénéficier d'une description, mais aussi que l'indexation devra prendre en compte les relations entre les morceaux décrits, c'est à dire que la structure elle-même sera décrite dans l'index. Ainsi un document sera décrit par les descriptions de ses parties, dont les indexations bénéficieront de la description du document les contenant.

Enfin, si l'indexation du document dépend de et représente la structure d'un document, alors il devient possible de mettre en place des requêtes qui interrogent également cette structure, par exemple « Trouver un paragraphe décrit par tels mots-clé contenu dans un chapitre contenant également une image dont l'histogramme se rapproche de celui-ci ».

En conclusion, il apparaît que le document multimédia entraîne un changement dans l'indexation par rapport aux textes au niveau des différents médias qui le composent, puisque l'indexation peut être symbolique ou *calculée* ; ainsi qu'au niveau de la *gestion de la structuration* des documents, aussi bien pour l'indexation que les requêtes. Les modèles d'indexation et de recherche d'information appartiennent alors de fait à la classe des modèles logiques, seuls à même de prendre en compte la complexité de l'information. [Chiarabella, 1997] utilise par exemple des graphes conceptuels pour représenter l'information d'indexation .

1.3.2 Utiliser l'information

La recherche de documents se passe dans la plupart des systèmes sous les deux formes des requêtes/réponses, et de la navigation, exploratoire ou non, dans la base, mêlant les acquis de la Recherche d'Information et des hypermédias. [Chiarabella, 1997] considère que ce sont là deux modes d'interaction de l'utilisateur et du système, la première consistant en une mise à disposition

d'un ensemble de documents répondant à une requête (décrite par une « fenêtre »), tandis que la navigation au contraire donne accès au voisinage d'un document (figure 1.2). Si l'on considère de plus les liens entre parties de documents indexées, la base de documents apparaît comme une énorme structure, dans laquelle il est possible de naviguer, en suivant les liens implicites, mais aussi les liens structurels (explicites), et aux noeuds de laquelle il est possible d'accéder à partir de requêtes sur l'ensemble des parties de documents indexées.

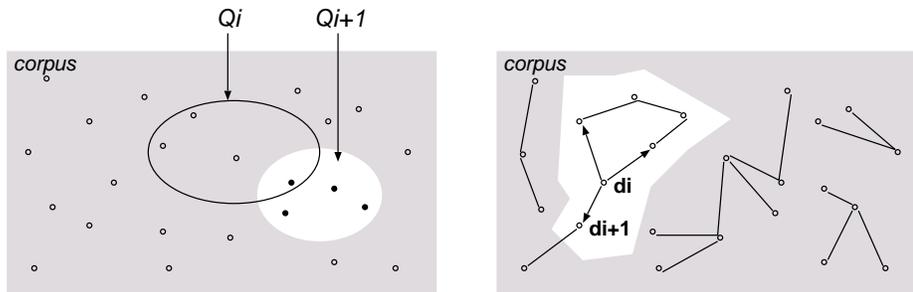


FIG. 1.2 – Recherche par requêtes et navigation : ouvrir des « fenêtres » dans le corpus de parties de documents d'après [Chiararella, 1997]

Cela pourrait correspondre à une appréhension de la base à la fois d'un point de vue *global*, fournissant à l'utilisateur un moyen de résoudre ses problèmes pour embrasser le contenu de celle-ci (désorientation et surcharge cognitive); mais aussi d'un point de vue local puisqu'autorisant un accès libre et sans limites à l'exploration (moins dirigé vers le but à atteindre donc).

Le processus de recherche d'information devient alors une suite de requêtes et d'exploration des solutions, c'est à dire une séquence de recherche interactive avec le système. Plus encore, il convient de considérer que la recherche d'information n'est qu'une partie du processus beaucoup plus général d'*utilisation d'information* [Card *et al.*, 1991, Yeo et Yeung, 1997], d'interaction coopérative entre l'utilisateur et le système; l'objectif du second étant d'aider le premier à mener à bien une tâche d'utilisation d'information.

C'est pourquoi il est nécessaire de mettre en place des cadres permettant une réelle intégration, et non seulement une combinaison entre ces deux modalités de recherche [Chiararella, 1997, Dunlop et Rijsbergen, 1993] laquelle est, comme nous l'avons vue, favorisée par le développement de la structuration des documents. Il s'agit encore d'autoriser une « interaction riche » entre l'utilisateur et le système comme le proposent par exemple [Rao *et al.*, 1995], qui considèrent qu'un « espace informationnel » doit prendre en compte :

- le *raffinement itératif des requêtes* puisqu'il apparaît majoritairement que c'est au cours de la session de recherche que l'utilisateur précise ce qu'il cherche réellement. Il faut également offrir la possibilité de naviguer interactivement dans des sous-ensembles de documents, de visionner le contexte dans lequel une requête a été traitée (avoir une idée de l'organisation de la base *autour* des documents retrouvés, ou encore de visualiser certains passages à l'intérieur d'un document).
- l'*hétérogénéité des sources* (par exemple plusieurs institutions), qui impose que l'utilisateur ait conscience du fonctionnement et de l'organisation de chaque type de source. Celles-ci doivent donc être modélisées de façon explicite.
- l'*accès parallèle* à l'information qui conduit à des temps de réponse variables suivant les documents visés, voire à l'intérieur d'un document. Il faut alors que l'utilisateur soit au

fait de ce qui se passe (temporellement parlant), et puisse travailler sur plusieurs requêtes en même temps pour s'affranchir d'attentes trop longues. L'interface devra alors refléter clairement le coût temporel de l'interaction, et autoriser plusieurs sessions de recherche parallèles.

- l'intégration d'un *processus de travail large* dans le système : techniques de recherche, de navigation, analyse des résultats doivent pouvoir être menées dans un espace de travail intégré.

Si nous poussons le dernier point à son paroxysme, il est en fait clair que le système informatique doit aider l'utilisateur à mener à bien une certaine tâche nécessitant l'utilisation d'information, la recherche de celle-ci étant intégrée au système lui-même. Ainsi, un logiciel idéal de traitement de texte permettrait de façon naturelle de mener des recherches sur des documents, de visualiser et de lire ceux-ci, de créer de nouveaux documents, de mettre ceux-ci à disposition sur le réseau, *etc.* Tout le problème réside alors dans la définition du système : par exemple un système d'exploitation associé à quelques logiciels permet de gérer toute l'information sur une machine (recherche, édition, visualisation...), mais un logiciel seul ne le permet pas (quoique qu'une suite bureautique comme Star Office s'en rapproche).

Nous ajouterons aux quatre caractéristiques proposées par [Rao *et al.*, 1995] la possibilité de mener des sessions d'exploitation longues, ce qui implique que celles-ci soient interruptibles, et qu'il soit possible d'en sauvegarder le contexte afin de pouvoir les reprendre après interruption.

Nous venons de présenter quelques-unes de caractéristiques des systèmes d'information documentaires « modernes ». Nous avons vu notamment que la recherche d'information, en tant que simple phase d'un processus plus général d'utilisation de documents devait faire appel à toutes les possibilités de description offertes par les documents structurés. Il convient maintenant d'interroger à nouveau la notion de connaissance (au sens général du terme) dans le contexte de tels systèmes.

Par exemple les connaissances sur les documents sont désormais explicitées et enrichies par les données de leurs structures, les parties de documents sont elles-mêmes indexées, et peuvent être manipulées et réutilisées dans de nouveaux documents.

Les connaissances du domaine *incluent* les connaissances de structuration des documents de ce domaine, ce qui permet de s'en servir pour décrire et pour interroger¹¹. Les connaissances de description peuvent également être utilisées pour naviguer directement vers certaines parties de documents comme déjà proposé il y a longtemps pour les hypermédias [Nanard et Nanard, 1991, Bruza, 1990, Nanard et Nanard, 1995], et sont de fait connaissances du domaine.

Toutes ces connaissances doivent être partageables, partagées et évolutives. Etudiant les systèmes multimédia, [Gabrielli, 1998] insiste sur le fait qu'il est nécessaire que les utilisateurs puissent accéder aux structures de métadonnées, à la fois pour rechercher des informations, mais aussi se faire une idée du domaine couvert par la base, ainsi les *connaissances de l'utilisateur sur le système* ne doivent pas être négligées. Par exemple, l'utilisateur peut plus ou moins bien connaître (ou deviner) l'organisation de la base de documents, l'organisation de ses descripteurs (exploration du thésaurus), et également les liens et les structures de descripteurs mis en place. Il s'agit donc de lui donner un aperçu de la manière dont sont organisés aussi bien les index que les documents, afin qu'il s'adapte au mieux de ses capacités à celles du système.

11. Par exemple, l'utilisation d'une organisation conceptuelle des termes de description permet d'enrichir les requêtes <http://www.sunlabs.com/research/knowledge/>.

Le système documentaire devient de plus en plus complexe tandis qu'il gère des informations dont la structure n'est pas totalement connue (par exemple des liens). Il y a donc lieu que des connaissances sur *les tâches des utilisateurs* soient considérées, lesquelles permettront de fournir une *assistance*¹² aux utilisateurs au cours de sessions d'utilisation.

La collaboration entre l'homme et la machine pour une tâche impliquant une utilisation d'information documentaire nécessite en effet que le système fournisse un ensemble de services de recherche à l'utilisateur, mais aussi soit capable d'essayer d'aider celui-ci en cas de besoin. Il s'ensuit qu'il y a lieu de modéliser (plus ou moins finement) l'ensemble des tâches connues d'utilisation du système, afin de pouvoir stocker sous la forme d'*expérience* des traces d'utilisation qui pourront être exploitées, d'une part afin d'apprendre de nouvelles connaissances (par exemple statistiques), qui permettront de modifier l'accès à l'information [Smail, 1994, Delanoy, 1996], voire les informations elle-mêmes (plus rarement); d'autre part dans l'objectif de pouvoir réutiliser cette expérience au besoin en y cherchant une situation équivalente à la situation présente d'un utilisateur, et en proposant des solutions plus ou moins similaires. A partir du moment où toute utilisation d'un système documentaire peut-être tracée, il devient possible à partir de ces traces d'aider l'utilisateur dans ses tâches, l'indexation elle-même pouvant être aidée par le système. Il apparaît en fait que indexeurs (documentalistes ou non) et les utilisateurs « quelconques » ont à leur disposition les mêmes outils, et que seule la tâche d'utilisation — la pratique — varie. Ceci ouvre la voie à d'intéressantes possibilités, par exemple aider un utilisateur à décrire comme un documentaliste a indexé, ou un documentaliste à indexer comme des utilisateurs ont décrit.

1.3.3 Vers une indexation intelligente

Un système d'information documentaire idéal manipule des documents structurés, des structures de documents (organisées par les liens), des structures de connaissances d'indexation, des structures d'index et enfin des structures décrivant ses propres connaissances d'utilisation.

Il s'agit donc idéalement de décrire toutes ces connaissances comme des documents, lesquels sont décrits, indexés et manipulables dans le système y compris au niveau de leur relations. Les index prennent alors une valeur fondamentale, puisqu'ils sont à la fois informations sur les informations, et sujets de description par d'autres index, et ont donc le statut de données et méta-données.

Nous définissons la notion d'*indexation intelligente* comme une indexation offrant *dans son mode de fonctionnement/représentation* même la possibilité de l'interroger sur elle-même, en tant qu'indexation et métadonnée (*cf.* par exemple [Davenport, 1996]), et plus seulement en tant qu'index à traverser vers une information « brute ». Tout document ainsi indexé offre par la description de son contenu structuré et indexé la possibilité d'inférences, de calcul sur ses index, que ceux-ci soient décrits en dehors de lui ou bien en son sein. Toute utilisation du document passe par conséquent par l'un de ses index, et toute utilisation peut être considérée comme ensemble de connaissances — comme le sont également les index utilisés comme tels en situation d'action —. Le système est alors entièrement géré par les connaissances, puisque pour toute utilisation, tout est connaissance.

Conclusion de cette partie. Nous avons dans cette partie introductive présenté un certain nombre de notions que nous pourrions ainsi réutiliser ou interroger dans la suite du mémoire :

12. Espérons-le, intelligente.

documents, documents numériques, documents multimédias, systèmes d'information documentaires, d'indexation, recherche d'information, connaissances documentaires, *etc.*

Nous avons ensuite présenté ce que serait pour nous un système d'information documentaire multimédia idéal, et vers quoi il y a donc lieu de tendre, c'est à dire vers un système prenant en compte une description plus ou moins structurée de documents, en permettant toute exploitation (et non seulement la recherche), celle-ci se basant sur des index généralisés (*i.e.* tout ce qui est utilisé comme tel), lesquels prennent le statut de connaissances lorsqu'ils sont explicités et utilisés. Un tel système gèrera l'ensemble de ses connaissances comme des documents, et fournira une assistance à l'utilisateur dans le cadre des ses tâches, fondée sur l'expérience d'utilisations passées.

Que le lecteur ne saute pas tout de suite à la dernière page, un tel système n'existe pas encore. Il nous semble toutefois que c'est cet idéal qu'il convient d'avoir en vue lorsque l'on met en place un système d'information documentaire « moderne ».

Première partie

Modélisation de documents
audiovisuels numériques

Chapitre 2

Exploitation de documents audiovisuels numériques dans un système d'information audiovisuelle

Sommaire

2.1 Document audiovisuel	50
2.1.1 L'image	50
2.1.2 Le flux d'images	51
2.1.3 Le son	52
2.1.4 Le montage audiovisuel: la véritable fabrique du document	52
2.1.5 Document audiovisuel	54
2.2 Document audiovisuel numérique	54
2.2.1 Formats de représentation de documents audiovisuels	54
2.2.2 Les apports du numérique	55
2.3 Systèmes d'information audiovisuelle	57
2.3.1 Utilisations de documents audiovisuels	57
2.3.2 Archives audiovisuelles	58
2.3.3 Fonctionnalités d'un SIAV	60
2.3.4 Annotation de documents audiovisuels numériques	61

Nous faisons dans cette partie connaissance avec le médium auquel nous allons nous confronter. Les documents audiovisuels comptent en effet parmi les derniers invités au banquet du tout numérique, et leur exploitation par l'ordinateur n'en est qu'à ses balbutiements, si on la compare par exemple avec celle des documents textuels, qu'on pourrait dire consubstantiels à l'ordinateur depuis que l'homme a cessé de parler à ce dernier à l'aide de fiches perforées. Il paraît donc nécessaire de s'intéresser d'un peu plus près au médium, dans ses particularités, mais aussi en comparaison avec d'autres, ainsi qu'à la manière dont celui-ci est généré, notamment au travers du montage, avant d'étudier les conséquences de la numérisation des documents audiovisuels sur ceux-ci. Nous étudierons ensuite les usages des documents audiovisuels, ainsi que les contraintes introduites par le numérique sur les fonctionnalités d'un système d'information audiovisuel. La nécessité d'une « écriture » sur le document pour en autoriser la manipulation est ensuite abordée, et nous terminons cette partie en présentant ce qu'est l'annotation d'un document audiovisuel, concept qui sera central au restant du mémoire.

2.1 Document audiovisuel

Le document audiovisuel par excellence mêle images et sons, et fait donc appel aux modalités d'appréhension visuelle et auditive. Nous appellerons composantes vidéo et audio les deux flux qui, superposés, forment un flux audiovisuel. Le terme vidéo nous semble en effet trop restrictif et limité pour un médium qui existe quand même depuis qu'une bande son a été synchronisée avec des images animées pour donner le cinéma parlant, et ce bien avant l'existence de la télévision, des magnétoscopes et des caméscopes. De plus, il nous semble que l'étude de flux audiovisuels ne peut — ou ne devrait — pas faire l'économie de la bande son. Le terme audiovisuel¹ est donc le terme idéal, en témoigne par exemple le nom Institut National de l'Audiovisuel (INA), où sont archivés aussi bien des images issues des actualités (format cinéma), de la production télévisuelle, ou que des émissions radio.

Les documents audiovisuels auxquels nous nous intéressons sont donc des flux temporels composés de flux superposés. En général, on aura un flux d'images, et un ou plusieurs flux de sons. La décomposition de la bande sonore en plusieurs flux peut découler du montage (par exemple un flux contenant musique et bruits, et un autre pour les dialogues, qui peut être changé pour des raisons de traduction), ou bien encore des effets que l'on cherche à obtenir (un son stéréo aura par exemple deux pistes différentes). Intéressons nous tout d'abord au flux vidéo, qui est un enchaînement temporel d'images.

2.1.1 L'image

Une image est délimitée par un cadre explicite, le plus souvent rectangulaire, qui fixe l'attention sur son contenu : le lecteur sait où regarder, et sait que ce qui est à l'intérieur du cadre a été mis en place par un être humain dans le but d'être regardé et interprété. Il projette alors cette volonté d'interprétation sur le *contenu visuel physique* (l'arrangement des points de couleur), et interprète l'image en signes. Rappelons que l'on considère en général trois types de signes visuels, selon leur degré de relation au « réel » : les signes concrets iconiques, abstrait iconiques (un panneau prévenant de la traversée d'animaux), et symboliques (une lettre, un panneau de sens interdit). L'image fait le plus souvent appel aux signes concrets iconiques, lesquels réfèrent par exemple à des objets du monde réel (une voiture, un drapeau), ce qui correspond au caractère *analogique* de l'image [Metz, 1968]. Ce premier niveau correspond à un niveau simple d'objets, pour lequel on pourrait dire que les objets sont reconnus, mais l'interprétation n'est jamais limitée (une image représentant un drapeau français peut en fait représenter la liberté ou l'oppression). De plus les signes-objets reconnus sont en relation, participent de par leurs arrangements à leur reconnaissances mutuelles et contextuelles et à des interprétations globales en signes de plus haut-niveau d'abstraction référant à d'autres signes, et ce *ad libitum*.

L'image, et la vision des images est fort étudiée aussi bien en sémiologie (par exemple [Groupe Mu, 1992, Vaillant, 1997]) qu'en psychologie cognitive [Bonnet *et al.*, 1989] quand il s'agit de comprendre la construction par le cerveau de formes et de concepts précis à partir d'un ensemble de tâches colorées. Nous n'irons pas plus loin en ce qui concerne l'analyse de l'image.

Retenons simplement qu'une image est *interprétée* par un être humain, qu'elle *contraint* de façon plus ou moins forte son interprétation, mais que cette contrainte n'est jamais totale, que l'interprétation fait appel à des systèmes d'appréhension de l'image, à des systèmes de signes *appris* et culturels. Ainsi, l'image de la Terre prise d'un satellite n'évoquera pas grand chose à

1. Audiovisuel a également perdu son statut de « Nouvelles Technologies de l'Education » (NTE) d'il y a vingt ans.

qui ne connaît pas ce genre d'image, n'a pas appris à les lire et ne sais pas par des informations extérieures à l'image dans quelles conditions celle-ci a été créée. A plus haut-niveau, on notera que toute photographie utilisée dans un journal vient avec son commentaire expliquant comment il convient d'en interpréter le contenu et les informations véhiculées. Une figure d'un article scientifique enfin est ininterprétable correctement (*i.e* selon la volonté de son auteur) en l'absence de sa légende. Une image ne contenant pas en elle-même la description de ses conditions de conception et de lecture, qui pourraient prescrire son interprétation de façon non ambiguë, reste le plus souvent pas interprétable sans ambiguïté en dehors d'un entour fortement prescriptif disant ce qu'il y a lieu d'interpréter.

Toute image est interprétable, mais toute interprétation n'est pas forcément correcte ou utile en regard d'un contexte particulier.

En bref, une image est interprétée en situation d'action par un être humain, l'interprétation dépendant de l'image elle-même, de son entour, et de l'engagement du lecteur. Mais l'image elle-même, à la différence par exemple d'un texte n'est pas sémiotisée *a priori*: les signes qu'elle contient ne sont pas désignés, mais résultent d'un processus interprétatif beaucoup moins contraint que dans un texte (un mot existe dans ce dernier en tant que tel).

2.1.2 Le flux d'images

La vidéo est un *flux temporel d'images*. Si on filme une scène, c'est à dire qu'on enregistre à l'aide d'une caméra 25 ou 30 images toutes les secondes, et qu'on fait défiler ces images à la même cadence devant un téléspectateur à l'intérieur d'un cadre fixe, celui-ci ne sera pas à même de distinguer dans ce qu'il voit une suite d'images. En effet, la persistance des images sur sa rétine fait qu'une image est remplacée par une autre, de telle sorte que les zones qui ne changent pas sont perçues comme un continuum stable, tandis que les mouvements sont lissés par la perception. L'*illusion de mouvement* qui en résulte entraîne alors l'*illusion de la réalité* de ce qui est vu. Si l'objet, la personne bouge sur l'écran comme elle le ferait dans la réalité, alors elle est perçue comme réelle. Si la caméra de plus n'est pas fixe, le mouvement de l'ensemble du décor autour des objets en donne une perception qui se rapproche de celle qu'aurait le spectateur en mouvement²

La visualisation d'un *plan* — suite ininterrompue d'images filmées en continu — donne par conséquent l'illusion de voir et de percevoir ce qui a été filmé comme si cela était réel³. La dimension temporelle du flux vidéo est une de ses caractéristiques principales. Un flux vidéo est donc un ensemble d'images associé à une vitesse de défilement, mais n'est pas que cela: il nécessite d'une part un dispositif (projecteur, téléviseur) organisant la visualisation, d'autre part un spectateur, seul à même de le percevoir en tant que flux et non en tant qu'ensemble d'images.

Remarquons également que la temporalité du flux vidéo permet d'en différencier les signes. Un plan n'aura en général qu'un unique *focus vidéo*, lequel sera par exemple un objet en mouvement, sur lequel l'oeil se fixera sans s'attacher aux détails des images, car il n'en a pas le temps. L'interprétation d'une suite d'images différera fondamentalement de l'interprétation de chacune des images fixes, car le flux audiovisuel coïncide avec le flux de conscience dans une temporalité fixée [Bachimont, 1999]. Remarquons dès à présent que l'étude d'un document audiovisuel en

2. C. Metz: « Le mouvement donne aux objets une corporalité et une autonomie qui étaient refusées à leur effigies immobiles, il les arrache à la surface plane où ils étaient confinés, il leur permet de mieux se détacher comme « figures » sur un « fond »: libéré de son support, l'objet se « substantialise » » [Metz, 1968].

3. Mieux, si un personnage s'envole soudain, le spectateur, passant sans le savoir un « contrat de réalité » avec le réalisateur du flux, considérera cela comme un mouvement réel, avec l'effroi que l'on peut imaginer au début du XXème siècle.

passant par ses images s'attache à son objet en niant, ou au moins en en déformant la temporalité. Il reste qu'il est difficile de faire autrement, les seuls objets matériels manipulables, les seules unités accessibles facilement étant les images dont est composé le flux⁴.

2.1.3 Le son

À la composante simplement visuelle d'un flux vidéo, on peut superposer un flux sonore. Si celui-ci a été enregistré en même temps que les images par la caméra et qu'il est synchronisé avec le flux vidéo, alors l'impression de réalité peut être redoublée : le personnage remue les lèvres, parle, et je l'entends.

Mais il apparaît souvent que le son tourné en même temps que les images est de mauvaise qualité, perturbé par des éléments extérieurs à ce que l'on veut filmer. Il y a alors souvent post-synchronisation d'un flux audio créé en studio avec le flux image. Plusieurs conséquences en découlent.

D'une part, il est possible de rajouter des sons non naturels, et qui pourtant font « plus vrais » que les sons réels afin de renforcer les liens contextuels avec les images. Un coup de poing qui n'aurait pas fait de bruit dans la réalité peut être sonorisé par un bruit sourd qui amplifiera l'impression de puissance du coup. D'autre part, on se limitera en général seulement aux sons importants pour la volonté du réalisateur, les sons annexes étant réduits à zéro. Au cours du XX^e siècle est ainsi apparue, à force de tâtonnements un véritable répertoire des bruits de cinéma, qui sont régulièrement utilisés, et ont été appris par les spectateurs. On appelle « Foley sounds » les bruitages.

D'autre part, une composante musicale peut faire son apparition, un plan peut être illustré par une mesure de musique, dont le volume baissera quand un personnage parlera.

Enfin, il est possible de doubler — plus ou moins bien — les personnages, sans que cela pose de problèmes au spectateur normal. Le cas du direct interdit bien entendu ce genre de montages.

Nous avons donc vu les trois types de sons de l'audiovisuel : les dialogues, la musique, et les bruitages, et qu'il était aisé de mettre en place une bande-son reconstruite pour accompagner les images, pour même au simple niveau du plan, construire un effet réellement audiovisuel, globalement supérieur à la juxtaposition des deux flux. Il y a donc déjà à ce niveau, montage.

2.1.4 Le montage audiovisuel : la véritable fabrique du document

L'art du montage audiovisuel consiste à mettre en place à l'aide des éléments à disposition (un ensemble de plans vidéo, des musiques, des bruitages, des dialogues, *etc.*) un véritable document exprimant les visées du réalisateur.

Pour cela, le réalisateur doit faire des choix à tous les niveaux.

Cela commence dès le filmage : choisir en effet de filmer telle ou telle partie du monde qui entoure la caméra relève d'une volonté de mise en scène.

Cela concerne surtout le choix des plans qui vont être découpés dans les rushes et utilisés, montés et collés ensemble afin de raconter visuellement ce que le document doit exprimer, ainsi que celui de la bande-son, elle-même composée de flux superposés. Les documents sans travail de montage différé (tels que les émissions en direct, les interviews, le théâtre filmé) se différencient de la production vue plus haut en ce sens que le montage, bien que prévu avant la prise de vue ou

4. Ceci pourrait changer avec de nouvelles technologies, telles MPEG4 (voir 2.2.1), dans lesquelles ont décrit les contenus audiovisuels à l'aide d'objets vidéo dont les représentations sont superposées à l'écran et sur les haut-parleurs : il devient possible d'accéder aux objets individuellement.

de son est réalisé en direct et ne peut être modifié. Mais le travail de montage y existe cependant ; les documents en direct, sans montage, sont très rares.

Un exemple célèbre illustre l'importance du montage dans l'audiovisuel, que l'on doit au réalisateur soviétique Kouleshov. On étudie les réactions du public à trois montages différents : un plan immobile d'un visage inexpressif suivi d'un plan sur un bol de soupe fumante ; le même plan du visage suivi d'un plan d'une femme morte dans un cercueil, et enfin toujours le même plan de visage suivi du plan d'un enfant en train de jouer⁵. Interrogé sur qu'il a vu, le public répond : un homme affamé, puis triste, puis heureux. Cette expérience montre qu'un plan ne signifie pas grand chose en lui même, et que la majeure partie de sa signification provient de son entour, aussi bien visuel (les autres plans) que sonore (le commentaire). Il n'y a pas là à s'étonner, de la même manière que l'interprétation d'une image dépend des indices qui la contraignent, l'interprétation d'un plan dépend d'autres indices situés dans son entour. Mieux, une séquence (que nous définirons ici comme une « suite de plans consécutifs constituant un tout sous le rapport d'une action dramatique déterminée » [Haddad *et al.*, 1996]) fait disparaître même la perception du plan chez le spectateur⁶.

Si l'unité de montage vidéo est le plan⁷, l'unité de montage audio est beaucoup plus floue, et ne saurait se résumer à un seul élément. Cependant, pour le flux vidéo comme pour le flux audio, mais aussi pour leur superposition existent un certain nombre de règles qui ont petit à petit été mises en place depuis que le médium existe. Ainsi il ne convient pas de superposer au même niveau sonore musique d'ambiance et voix, ainsi dans un dialogue il faut respecter la règle des 180° (deux personnes qui se parlent de doivent pas être prises suivant le même profil), ainsi encore les raccords sur les mouvements doivent-ils être travaillés (il ne doit pas y avoir d'impression de saut temporel pour le spectateur). Ces règles de « bon sens » cinématographiques visent à ne pas troubler le spectateur et à lui permettre de s'installer dans le document.

Le document audiovisuel lui-même appartient également à un genre, ce pourra être un reportage, un film publicitaire, une explication de montage de circuit électronique, un programme scientifique, un film, un journal télévisé. Chacun de ces genres introduira des règles particulières sur la manière de raconter⁸, de faire progresser une intrigue, d'expliquer, de présenter des faits, *etc.* Chacun respectera également des canons dans la manière de filmer (en plans fixes ou tout en mouvements), de monter (fondus ou cuts), d'utiliser la musique...

Il apparaît donc que le réalisateur d'un document audiovisuel dispose d'une liberté certaine dans ses choix, mais que ceux-ci sont naturellement contraints par le média.

Ainsi, de la même manière que la possession d'un dictionnaire ne suffit pas à écrire des textes, il convient de connaître les règles qui rendront le document « lisible » par le spectateur. Ce qu'on pourrait appeler le langage audiovisuel.

Ainsi encore, de la même façon que les textes répondent à des genres, qu'il existe des manières de communiquer des informations ou des histoires, l'audiovisuel met en place ses propres codes. Il va de soi que ceux-ci dépendent également des moyens de diffusion : les codes des émissions de télé-achat ne sont pas les mêmes que ceux des films de science-fiction. On notera ainsi par exemple que dans les seconds l'alliance de l'image et du son ont pour objectif de clouer le spectateur à

5. Les descriptions varient suivant les sources : l'enfant qui joue peut être remplacé par une « femme lascive ».

6. C. Metz : « Le film, que l'on croirait susceptible de donner lieu à une lecture transversale, par l'exploration à loisir du contenu visuel de chaque plan est presque à tout coup l'objet d'une lecture longitudinale, précipitée, déplacée vers l'avant et anxieuse de la « suite ». La séquence n'additionne par les plans, elle les supprime » [Metz, 1968].

7. Auquels on peut rajouter les objets vidéo insérés, par exemple les incrustations de commentateurs météo sur les photos satellites.

8. Les règles du montage et de la narratologie ont été fort étudiées pour le cinéma [Chevrier, 1995, Gardies, 1993].

son siège, ce qui n'est pas exactement l'objectif des premiers⁹.

Enfin, de la même manière que pour les textes, la personnalité de l'auteur peut s'épanouir au sein de ces règles, ou bien en les remettant en cause¹⁰, et on est du côté de l'art, ou bien s'accomoder de règles établies et validées¹¹.

2.1.5 Document audiovisuel

Un document audiovisuel peut être perçu comme une superposition de flux — techniquement parlant — mais la manière dont il a été composé, la mise ensemble des éléments, les règles auxquelles l'auteur a obéi, les contraintes d'interprétation qu'il a mises en place, ne sauraient se réduire à une superposition de flux. Un document audiovisuel possède, pour reprendre la terminologie vue dans la partie 1.1.1 une structure physique minimale, décrivant la mise ensemble des images et la synchronisation d'une bande-son plus ou moins complexe, tandis que sa structure logique, non explicitée, recouvre toutes les analyses possibles du document. Enfin, la diversité des types de documents audiovisuels, liée à leurs production dans des objectifs, pour des publics et sous des formes différents en font un médium difficile à appréhender en soi, globalement, en tant que document.

C'est pourtant à cet objectif que s'attachent un certain nombre de scientifiques, depuis que le document audiovisuel est devenu numérique, c'est à dire propre au calcul.

2.2 Document audiovisuel numérique

Les progrès dans les vitesses de traitement de l'audiovisuel permettent désormais de manipuler des documents audiovisuels numériques avec des tailles d'image et une qualité satisfaisantes pour de nombreuses applications. L'arrivée sur le marché des DVD (Digital Versatile/Video Disk) en est la preuve, qui doivent remplacer les supports analogiques chez les particuliers. Ajoutons que les projecteurs numériques sont seulement à leur début de carrière dans les salles de cinéma. En ce qui concerne les réseaux, si les débits augmentent, la transmission de données audiovisuelles ne semble pas encore être une réalité largement mise en place, si ce n'est à titre d'essai sur des réseaux locaux. Il n'empêche que les progrès technologiques sont là, et qu'il convient d'envisager le statut du document numérique comme une réalité qui ne peut que prendre de l'ampleur. La présente partie étudie le document audiovisuel numérique en soi, étudiant d'abord les formats de description, avant d'envisager les apports du numérique pour l'exploitation des documents. La partie suivante est réservée à ses utilisations et aux fonctions des systèmes d'information audiovisuelle.

2.2.1 Formats de représentation de documents audiovisuels

Un document audiovisuel étant considéré comme superposition de flux audio et vidéo, un certain nombre de formats ont été mis au point afin de stocker son et image dans la forme la plus

9. La bande-son y prend une importance prépondérante sur l'image — comme dans une interview. Comme le montrent de façon convaincante Sorlin [1992] ou Chion [1994], le rôle de la télévision dans la plupart des foyers est souvent pendant la journée celui qui était assuré par les radios dans le passé: celle d'une présence, essentiellement sonore, qui permet de vaquer à d'autres occupations, par exemple pousser le siège pour passer le balai, les programmes étant souvent conçus en conséquence.

10. Voir par exemple la Nouvelle Vague en France.

11. Il n'est pas sûr qu'un réalisateur d'une émission de télé-achat ait plus de liberté qu'un rédacteur d'horoscope dans un magazine.

compacte possible. Les standards MPEG-1 (destiné aux CD-ROM), MPEG-2 (télévision numérique, DVD); les formats AVI, QuickTime, RealVideo... résultent des efforts de la communauté internationale ou d'entreprises dans la mise au point des formats de représentation. Certains de ces formats sont avec pertes ou sans perte (ce qui est important quand il s'agit d'éditer des flux audiovisuels), et peuvent faire cas de la temporalité du flux ou non (par exemple prendre en compte la redondance d'information entre images qui se suivent, comme dans MPEG).

Le format MPEG-4 [Koenen, 1999] se donne des objectifs plus ambitieux, tournés vers la télévision numérique et la vidéo à la demande, les applications graphiques interactives (avec un contenu synthétisé), et de façon plus générale liés au multimédia et au Web. Il s'agit donc principalement d'intégrer en un seul format tout ce qui est lié à la production, à la distribution et à l'accès aux contenus audiovisuels. La norme définit ainsi des unités de contenu sonore, image ou audiovisuel appelées *objets média*, donc l'origine peut être quelconque (enregistrement ou calcul pur). Les objets média sont regroupés en *scènes* audiovisuelles, par exemple un objet fond calculé, sur lequel on ajoute un objet personnage et un objet son. MPEG-4 définit également la manière de multiplexer et de synchroniser les données associées avec les objets, en gérant la qualité de service de la transmission. Enfin, l'utilisateur final, qui visualise les scènes arrivées par morceau jusqu'à son poste peut interagir avec elles, par exemple cliquer sur des objets pour les désigner à un service de télé-achat.

Retenons de MPEG-4 la notion d'objets média qui sont multiplexés et synchronisés au besoin pour former des documents audiovisuels. Si bien entendu seules les scènes artificielles fonctionnent réellement aujourd'hui (*i.e.* donnent des résultats de construction acceptables), les bases n'en sont pas moins jetées d'une construction explicite de documents par objets, c'est à dire d'une sémiotisation *a priori* du document définissant dans le document lui-même des contraintes fortes d'interprétation.

MPEG-7 lancé il y a peu de temps, est consacré à la représentation de documents audiovisuels en vue de leur recherche. Nous reviendrons en détail sur cette norme en cours de conception lorsqu'il s'agira de lui comparer notre propre approche de représentation.

2.2.2 Les apports du numérique

Considérant les conséquences de la numérisation des documents textuels, Virbel [1994] note quatre possibilités amenées par le numérique :

- les explorations linguistiques fines, puisque le numérique permet d'instrumenter le travail manuel d'analyse en gardant en mémoire des parcours interprétatifs ;
- la représentation de l'organisation textuelle dans le texte lui-même, la définition d'unités textuelles (*cf.* SGML, [Kimber, 1995]) ;
- les opérations de segmentation et de regroupement d'unités textuelles, génération par exemple de nouveaux textes suivant un certain nombre de critères [Alaya *et al.*, 1995] ;
- la continuité technique entre la visualisation et la saisie de données textuelles, permettant une relation entre le texte et le lecteur enregistrant sa lecture : la possibilité d'annotation.

Pour les documents audiovisuels, la nouveauté fondamentale et première introduite par le numérique est la possibilité d'*accès direct* (random access) à une quelconque des parties du document. Si un document audiovisuel a par exemple une durée d , alors il est possible de jouer celui-ci à partir de l'instant t avec $0 \leq t \leq d$. Le repère temporel t joue alors le rôle d'identifiant

pour la partie de document commençant à cet instant, voire d'un index s'il est repéré par exemple sur une glissière permettant d'accéder au document.

L'accès direct autorise la navigation fine dans le document, au sens où il est possible de sauter d'une quelconque de ses parties à une autre autant de fois que nécessaire pour la tâche d'analyse en cours. D'autre part, le dispositif matériel de visualisation vidéo se simplifie (l'écran de l'ordinateur peut faire l'affaire), l'arrêt sur image, les ralentis et les accélérés sont simples à réaliser et à manipuler.

La manipulation de parties de document est grandement facilitée par le numérique. Le montage a certes toujours été affaire de coupe et de collage, mais il devient ici très facile et rapide d'expérimenter un montage, en visualisant plusieurs flux en même temps, de superposer plusieurs flux sonores, *etc.* S'y ajoutent la possibilité de tracer les opérations de montage réalisées, mais aussi celle de générer des images artificielles qui permettent d'une part de créer des documents entièrement artificiels, mais aussi autorise toute sorte d'effets de transition entre les plans qui n'étaient pas imaginables avec les appareils de montage classiques.

Il devient également possible d'organiser le document en attachant des index à ses différentes parties. Les tables de matières permettent d'accéder à des parties *nommées*, par exemple les différents « chapitres » d'un film en DVD. Le document peut donc être considéré de façon *délinéarisée*, c'est à dire qu'il devient d'une certaine manière document structuré. Dès qu'une partie est décrite et accessible, elle peut être réutilisée et montée.

Le nommage d'une partie du flux audiovisuel correspond en fait, de façon plus générale, à sa *documentation*. Il est ainsi possible de décrire un document avec des caractéristiques telles que les droits qui y sont attachés, le nom de son auteur, au sein même du format de représentation numérique du document. Certains formats permettent d'ores et déjà ce genre de marquages, par exemple QuickTime 3 d'Apple. En fait la documentation du document est un pas vers ce qui est appelé *numérique intégral* à l'INA, c'est à dire que la documentation doit être présentée et intégrée aux outils numériques, de la conception du document au filmage des rushes¹² et au montage, jusqu'à son exploitation et sa réutilisation.

Une autre particularité du numérique est que les documents peuvent également être générés *au besoin*, à partir de descriptions. De la même manière qu'un fichier MIDI est une description d'un morceau à construire, et non un enregistrement qu'il s'agit de jouer, un document VRML ou MPEG-4 peut être généré au moment de la lecture. On notera que les objets média de MPEG-4 de type vidéo peuvent être mêlés à des objets construits, par exemple un fond vidéo sur lequel on plaque une incrustation de publicité pour construire un nouveau document audiovisuel. On touche là à une dimension intéressante de l'audiovisuel numérique : à partir du moment où une partie de document audiovisuel est manipulable en machine, il est possible de l'utiliser dans d'autres documents, non simplement audiovisuels, mais multimédia au sens où ils mêlent par exemple texte et médium temporel.

Inversement, un document audiovisuel est également un document multimédia — appauvri, puisque l'on ôte à celui-ci son interactivité explicite. En effet, mobilisant les deux modalités auditives et visuelles, il est possible de retrouver dans un document audiovisuel simple à la fois du texte (un générique), et d'autres morceaux de documents, y compris audiovisuels¹³. Ajouter des possibilités d'interaction basées par exemple sur une table des matières rend le document numérique pleinement multimédia au sens où nous l'avons défini.

12. Les rushes sont les plans « bruts » tels qu'ils ont été tournés. Les caméras numériques permettent de documenter les conditions de tournage, automatiquement (heure, température) ou manuellement. Cette opération est appelée *logging* en anglais [Davenport *et al.*, 1991].

13. Voir par exemple les expérimentations de Peter Greenaway dans le film *The pillow book* (1996).

Il apparaît en conclusion de cette partie que l'apparition du numérique autorise à accéder à n'importe quelle portion de contenu d'un document audiovisuel, et que la définition même de ces parties permet de les documenter (de les annoter, nous y reviendrons) et d'en considérer une structuration, au sein du document même, mais également pour une utilisation dans d'autres documents. Nous pouvons alors toucher du doigt le fait que la possibilité de décrire un document et sa structure afin de l'exploiter numériquement (approche d'indexation MPEG-7), peut être à la limite — et paradoxalement — renversée en la possibilité de décrire un document et sa structure et ses objets pour en permettre la génération numérique (*cf.* MPEG-4).

2.3 Systèmes d'information audiovisuelle

Après avoir rapidement décrit ce que sont les documents audiovisuels numériques et quelles utilisations on est en droit d'en attendre dans le futur, il convient d'interroger la notion de *système d'information audiovisuelle* (SIAV) que nous définissons comme tout système informatique permettant d'exploiter des documents audiovisuels numériques. Nous commencerons par nous interroger sur les différentes utilisations passées et futures des documents audiovisuels, avant de décrire les fonctionnalités qu'on peut attendre d'un SIAV. Nous concluons ce chapitre en évoquant la notion d'*annotation*, qui est le principe au travers duquel nous allons mener toute notre étude de la représentation audiovisuelle.

2.3.1 Utilisations de documents audiovisuels

Le cycle de vie d'un document audiovisuel peut sensiblement être considéré comme le suivant :

- conception, écriture du scénario, mise en place d'un synopsis s'il y a lieu (pré-production) ;
- tournage des images, enregistrement de la bande sonore, recherche d'archives ;
- montage et post-production ;
- reproduction et/ou diffusion ;
- archivage ;
- hors du cycle, éventuellement réutilisation de parties dans d'autres documents.

Si l'on ne s'intéresse qu'à tout ce qui suit la création du document, son utilisation première en est la simple visualisation, l'appropriation de son contenu.

Cette appropriation peut être faite dans divers objectifs : se distraire, s'informer, apprendre. Le genre du document dépend bien entendu de l'objectif premier ayant présidé à sa création. On ne crée pas un film de la même manière qu'un documentaire animalier ou qu'un manuel d'utilisation d'un appareil, de même qu'on ne les regarde pas de la même façon. Historiquement, le document audiovisuel a d'abord été le film, de distraction ou d'actualité, ainsi que les bandes-annonce et les publicités. Avec l'apparition de la télévision sont apparus les journaux télévisés, les émissions régulières, les reportages, *etc.* L'audiovisuel est également utilisé pour la formation technique, par exemple l'apprentissage de connaissances procédurales [Lieberman, 1994], ou bien la surveillance (supervision, contrôle qualité, zones sensibles, *etc.*). Le document n'est alors qu'objet de visualisation.

Avec l'avènement du super-8, puis du magnétoscope, les particuliers ont été à même de créer leurs propres documents audiovisuels (c'est à dire de passer au niveau de la conception), dans

le même temps qu'a été rendue possible la diffusion de masse de programmes ayant déjà été diffusés dans les cinémas et sur les ondes. Les documents audiovisuels sont donc devenus partie du quotidien, et tout un chacun peut d'ores et déjà s'essayer à la création.

En ce qui concerne le champ universitaire, nous avons distingué avec Gwendal Auffret dans [Auffret et Prié, 1999] plusieurs façons de considérer les documents audiovisuels dans les Sciences Humaines. Ces utilisations ont pour point commun une visualisation active des documents audiovisuels, une analyse de ceux-ci, et on peut donc les rapprocher d'autres utilisations dans d'autres domaines tels que celui de la supervision :

- *outils pédagogiques* : les documents peuvent être utilisés dans les cours ou les conférences comme illustrations (auquel cas on utilisera des documents existants), ou bien comme outils pédagogiques spécifiques [Jacquinot, 1985]. Par exemple dans un cours de communication ou de théâtre, il est possible de filmer une performance des élèves et de la commenter en faisant une large utilisation des fonction d'avance et de retour arrière.
- *témoignages du passé* : les historiens par exemple [Ferro et Planchais, 1997] considèrent les documents audiovisuels comme des témoignages de valeur ou comme des miroirs de nos sociétés¹⁴. L'analyse utilise alors le magnétoscope et une prise de notes intensives.
- *œuvres d'art* : considérés comme témoignage de la créativité d'un auteur ou des ses choix esthétiques et thématiques [Bordwell, 1993], les documents audiovisuels peuvent être étudiés comme œuvres d'art.
- *notes personnelles* : les anthropologues, les psychologues, les sociologues ou les spécialistes en sciences de l'éducation utilisent souvent la caméra pour conduire une analyse *a posteriori* du comportement humain. Dans un tel contexte, les documents audiovisuels sont un moyen de mémoriser des faits passés de la vie réelle, ce sont les notes personnelles du scientifique [Aguierre-Smith, 1992], qui nécessitent de pouvoir être classées et manipulées.
- *actes de communication* : la sémiologie des documents audiovisuels — telle qu'elle a été définie par C. Metz [Metz, 1968, Metz, 1973] pour le cinéma et étendue aux documents télévisuels [Jost, 1992, Jacquinot, 1977] — étudie ceux-ci comme actes de communication. Les sémioticiens analysent la manière dont le médium audiovisuel est utilisé pour transmettre de l'information et des émotions, c'est à dire la manière dont apparaît le sens d'un document et quels sont les éléments utilisés pour son interprétation. Il s'agit alors d'étudier des corpus de documents, de mener des analyses du discours, d'évaluer les conditions de production et de réception et de mettre en relation la forme et la signification du matériau audiovisuel en utilisant des modèles de stratégies de communication audiovisuelle.

Il apparaît au travers de cette petite étude que les utilisations de documents audiovisuels peuvent se diviser en trois catégories principales : la *visualisation simple* suivant la temporalité du flux, relativement passive ; la *lecture active* qui fait une large part à l'annotation personnelle et à un accès direct au flux ; et enfin la réutilisation de parties de documents dans d'autres, c'est à dire l'*édition*.

2.3.2 Archives audiovisuelles

Toutes ces utilisations documentaires nécessitent au préalable de disposer des documents, c'est à dire que leur archivage et leur recherche représentent des passages obligés. C'est pourquoi

14. On notera que la manière dont un document est diffusé, par exemple comme partie d'un journal télévisé est souvent aussi pertinente que ce qui a été filmé .

des archives audiovisuelles ont petit à petit été mises en place.

Les archives dépendent en fait des institutions qui les abritent. Celles-ci comptent principalement des chaînes de télévision, qui gardent leur production (par exemple d'actualité) en vue de leur réutilisation, les producteurs (par exemple de cinéma) qui gèrent des fonds, les fondations attachées par exemple à conserver une partie du patrimoine artistique d'un auteur¹⁵, les dépôts légaux comme l'Inathèque en France¹⁶, les entreprises, la publicité, les hôpitaux, *etc.*

Les archives audiovisuelles traditionnelles possèdent leurs documentalistes, spécialisés dans l'indexation/recherche des documents et dans la restauration de documents anciens. Avec le développement du numérique (au niveau des documents et non des notices s'entend), on assiste à une remise en questions du métier de documentaliste audiovisuel, le statut de l'archive même change. Ainsi, Michel [1997], appliquant l'analyse de la valeur à la « tradition audiovisuelle » considère que les processus de travail sont remis en cause, que les objets audiovisuels eux-mêmes « explosent » (n'ont plus l'unité qu'on pouvait leur trouver quand ils étaient sur bande), que les rapports producteurs/consommateurs changent également. Il distingue quelques processus en œuvre dans l'exploitation de documents audiovisuels : création/production/diffusion de biens matériels ; processus utilisation/exploitation/consommation qui reflètent les usages culturels d'utilisation de ces biens matériels ; inter-médiation/information/communication ; et enfin conservation/préservation. Les documentalistes, les archives doivent alors s'interroger sur leur rôle dans la société de l'information audiovisuelle¹⁷. Les pratiques documentaires et les systèmes d'information mis en place s'en ressentent alors dans les services qu'ils offrent aux utilisateurs.

Il n'est en effet plus seulement question de documents décrits avec de simples notices que l'on recherche avant de les télécharger. Il s'agit de gérer la documentation en même temps que le document, par exemple les droits, la diffusion de ceux-ci au travers de réseaux, voire la génération de nouveaux documents comme réponse à des requêtes (par exemple un document contenant toutes les apparitions d'un homme politique au cours d'une journée de télévision). D'autres part, la création numérique de documents audiovisuels entraîne qu'il devient possible pour le créateur de documenter celui-ci directement, sans passer par la médiation de documentalistes, dont l'intervention doit alors être ajoutée en fonction des objectifs de la collection. Enfin, la mise à disposition sur le réseau de productions « légères » (dues aux particuliers) est possible, qu'il faut documenter.

Les archives audiovisuelles doivent donc dépasser le stade de la mise à disposition de notices que l'on a vu apparaître avec le développement de l'internet. Idéalement, les documents doivent pouvoir être recherchés, manipulés, mis à disposition et téléchargés, réutilisés, de la façon la plus naturelle possible pour toutes les tâches qui le nécessite, de la même manière que les documents hypermédias disponibles sur le réseau. La difficulté est bien entendue accrue par la taille des archives. Par exemple l'INA stocke 3 millions de documents (400 000 heures de vidéo et 500 000 de radio). Si on met ces chiffres en rapport avec la taille des fichiers numériques correspondant à des flux audiovisuels de qualité, on atteint des tailles gigantesques de données numériques.

15. Signalons au passage la volonté récente du gouvernement des Etats-Unis de financer la restauration et la préservation de quelques films anciens « témoignages du génie américain ».

16. Créée en 1995, ce qui correspond à la reconnaissance officielle par l'Etat français des documents audiovisuels produits et diffusés en France comme partie de l'héritage culturel national [Denel *et al.*, 1994]. l'Inathèque enregistre systématiquement les programmes des six principales chaînes françaises et de Radio France, et met les documents à disposition d'un public de chercheurs. Les notices sont consultables à <http://inatheque.ina.fr>.

17. Par exemple dans un rapport récent à l'UNESCO [Edmondson, 1998], les archivistes audiovisuels cherchent à définir leur métier.

2.3.3 Fonctionnalités d'un SIAV

Nous passons rapidement en revue les fonctionnalités globales souhaitables d'un système d'information audiovisuelle numérique, tirant partie des apports du numérique entrevus dans ce chapitre et le précédent.

Stockage — accès

Un système d'information audiovisuelle doit être capable de stocker des documents audiovisuels et de les servir en réponse aux requêtes des utilisateurs. La « vidéo à la demande » dépend fortement de ce genre de requêtes, dans un premier cas il s'agit de fournir avec des débits suffisants des fichiers documents ou parties de documents audiovisuels. Dans un deuxième cas, et c'est l'objectif principal, il est crucial de fournir un flux assurant une qualité de service suffisante pour que l'utilisateur visualise le document au fur et à mesure qu'il reçoit les informations. Il s'agit alors par exemple d'organiser les documents — ou les fragments — de façon optimale sur les disques, de dupliquer éventuellement l'information [Mostefaoui *et al.*, 1999].

Indexation — recherche — présentation

Ces fonctionnalités sont bien entendu la clé de voûte des SIAV, l'indexation consiste à mettre en place les descripteurs de documents, qui seront ensuite utilisés pour la recherche. La présentation concerne la manière de donner à l'utilisateur un aperçu du document qu'il a retrouvé. La temporalité d'un document étant celle de la vie réelle, il faut en effet un temps égal à sa durée pour avoir un aperçu complet de celui-ci, ce qui semble rédhibitoire même si l'on doit simplement vérifier lequel parmi une dizaine de réponses à une requête est pertinent pour sa tâche, alors que cela peut prendre relativement peu de temps sur le Web pour des documents principalement textuels. Il s'agit donc de présenter de façon *détemporalisée*¹⁸ un document audiovisuel, de la façon la plus synthétique possible et adaptée à la recherche de l'utilisateur. La navigation par exemple dans la base pourra se faire à partir des présentations calculées de ses documents, par exemple un ensemble d'images-clé censées représenter des films.

La recherche de documents peut, nous l'avons vu se faire dans des objectifs très variés. Il est possible de rechercher des documents possédant des informations factuelles, des images ou des sons à réutiliser, des documents précis, des ambiances, *etc.* On recherchera par exemple des plans avec un personnage de dos devant une foule, ou bien des couchers de soleil, ou encore une description du fonctionnement du cerveau. La recherche peut être précise si l'utilisateur a les références du document ou de sa partie, ou bien imprécise au sens où l'utilisateur sait ce qu'il cherche car il l'a déjà vu ou entendu, ou enfin totalement imprécise quand il a une vague idée de ce qu'il cherche [Bolle *et al.*, 1998].

L'importance des interfaces de consultation et de navigation dans la base de documents est grande, de par l'importance que prend celle-ci dans les systèmes d'information documentaires en général, mais aussi du fait du média, temporalisé, qui oblige à construire des vues détemporalisées des documents afin de pouvoir en prendre connaissance sans les jouer entièrement. Les vues de présentation des documents peuvent être calculées au besoin, mais dépendent également du modèle de représentation documentaire choisi. Par exemple, s'il s'agit de « feuilleter » rapidement un document audiovisuel, alors ce feuilletage pourra être régulier (une image toutes les 3 secondes), ou bien irrégulier, par exemple passer d'un plan à l'autre, ce qui suppose une description du document prenant en compte les plans.

18. Mais non forcément *atemporalisée* : la temporalité peut être en partie conservée, par exemple si on conserve des extraits de bande-son pour accompagner une image représentative d'un plan.

Navigation — annotation

Dès que l'utilisation du document dépasse la simple visualisation, c'est à dire que le document est analysé (soumis à une lecture active) ou réutilisé ou généré, le modèle de représentation permettra d'accéder aux parties intéressant la tâche en cours.

Ainsi il doit être possible de naviguer dans un document, d'une partie à une autre, en suivant des liens décrits par avance ou bien calculés au besoin. Ensuite, il doit être possible d'annoter soi-même le document, c'est à dire d'en décrire des parties, de pouvoir relire ces descriptions, et les mettre en relation [Virbel, 1994]. il s'agit alors d'avoir une aide automatique à l'analyse, ou simplement à la prise de note, comme marquer un plan comme étant à réutiliser plus tard, par exemple pour générer un nouveau document.

Manipulation — génération

La manipulation et la génération concernent les possibilités de découpe de parties de document et de montage de ces parties en de nouveaux documents. Cette génération peut être automatique, par exemple afin de répondre à une requête, ou bien correspondre à une aide à la création audiovisuelle. Si des bancs de montage numériques (Avid par exemple) ou bien des produits grand-publics tels Adobe Première¹⁹ ou Macromedia editor²⁰ existent déjà, tout système d'information audiovisuelle devra disposer de possibilités minimales de montage et d'édition afin de conserver le travail effectué.

Assistance

L'assistance à l'utilisateur dans sa tâche a la même importance que dans un système d'information documentaire classique tel que décrit dans le premier chapitre. Les tâches étant liées à la manipulation de documents audiovisuels étant par nature complexes, l'assistance en est presque obligatoire. Il convient d'offrir dans les fonctionnalités d'un SIAV la possibilité de gérer cette assistance.

2.3.4 Annotation de documents audiovisuels numériques

Nous nous sommes attachés au long de cette partie à étudier ce que l'arrivée du numérique dans l'audiovisuel permettait d'imaginer au niveau de l'utilisation et des fonctionnalités des systèmes d'information audiovisuel. Cette étude est cependant partielle, nous ne pouvons bien entendu imaginer toutes les évolutions introduites dans l'audiovisuel, au niveau de son appréhension sociale ou de ses évolutions en terme de document multimédia. La présente recherche participe en effet de cette évolution.

Nous considérons que le modèle de représentation de documents audiovisuels choisi dans un SIAV en représente le point-clé et la charnière. Sur ce modèle seront en effet construits les index qui serviront de support à la recherche. La navigation dans les documents prendra appui sur le substrat représenté par la représentation, tout comme la génération et la manipulation de documents. Le modèle de représentation doit enfin pouvoir être enrichi pour permettre l'annotation par l'utilisateur des documents qu'il étudie.

En fait, le document audiovisuel numérique existant à la base comme simple superposition de flux, ou — à la limite — comme suite d'images munie d'une vitesse de lecture, reste largement inorganisé. Il ne peut donc être accédé et manipulé autrement que par la description structurée,

19. <http://www.adobe.com>.

20. <http://www.macromedia.com>.

de ses parties. Sa représentation et sa structuration dans un modèle est donc une condition *sine qua non* d'utilisation numérique telle que la navigation ou la génération²¹.

S'il est certain qu'il est nécessaire de mettre en place une structuration explicite des documents audiovisuels afin d'être à même de les exploiter, nous ne faisons pourtant pas de pari sur la forme que prendra cette structure. Particulièrement, nous refusons de nous décider sur son éventuelle arborescence, nous ne pouvons donc pour l'instant qu'adopter une approche basée sur la seule chose dont nous soyons certain, à savoir que tout modèle de représentation associera une description à une partie de document audiovisuel, la première jouant le rôle d'index pour la seconde.

Nous appelons *annotation* l'acte qui consiste à mettre en place cette association, quelles que soit la partie de document considérée et la description. La description elle-même peut également prendre le nom d'annotation d'une partie de document. Si l'on parle d'annotation d'un document, alors on fait référence à l'ensemble des annotations de ses différentes parties.

L'annotation (comme processus ou résultat) est souvent considérée comme faisant référence au texte libre (ou à son écriture) qu'un utilisateur d'un système d'information documentaire ou autre a le droit d'utiliser afin d'écrire tout et n'importe quoi sur un document ou une de ses parties, se rapprochant en cela de la pratique de l'érudit qui annote le livre qu'il est en train de lire. Les travaux autour de la station de lecture de documents de la Bibliothèque Nationale de France ont ainsi eu pour objectif d'offrir une assistance informatique à la mise en place et à la gestion d'annotations [Chahuneau *et al.*, 1992].

Nous étendons cette notion à toute description pouvant être attachée à une partie de document, quelle que soit cette description, ce qui signifie qu'elle peut être automatique, et de tout type. Par exemple un entier, une image ou un mot-clé pourront représenter des annotations de parties de documents audiovisuels. Remarquons d'ores et déjà qu'il est possible de considérer l'annotation comme une véritable *écriture* à propos d'un flux audiovisuel et sur ce flux à l'aide d'objets informatiques qui sont alors temporalisés par rapport au flux.

S'interroger sur la matière avec laquelle cette écriture se réalise, et dans quels objectifs revient à interroger la modélisation des documents audiovisuels. Nous étudierons donc au chapitre suivant les diverses tentatives de modélisation de documents audiovisuels au travers de l'annotation et de ses propriétés.

21. Ces considérations font écho à celles d'usage au niveau des documents textuels, dont les modèles de structuration permettent la manipulation et l'accès. Si bien entendu les documents audiovisuels sont différents des documents classiques, il n'empêche que bien des études menées sur eux sont consciemment ou non redevables des acquis issus des recherches sur les documents textuels. Sans préjuger d'évolutions futures pouvant radicalement changer notre manière de voir les documents audiovisuels, il convient de reconnaître que l'analogie avec les documents textuels se révèle féconde, pourvu qu'on en reste conscient.

Chapitre 3

Etat de l'art de la représentation de documents audiovisuels

Sommaire

3.1	Des champs de recherche et des problématiques différentes . . .	64
3.2	Caractéristiques d'annotation de documents audiovisuels	66
3.2.1	De la granularité de l'annotation	66
3.2.2	Primitives	69
3.2.3	Caractéristiques calculées de haut-niveau d'abstraction	70
3.2.4	Caractéristiques interprétées	73
3.3	Modèles de structuration	74
3.3.1	Structuration des annotations	75
3.3.2	Structuration implicite de documents audiovisuels	77
3.3.3	Structuration hiérarchique	78
3.3.4	Structuration par graphes	80
3.4	Fonctionnalités de systèmes d'information audiovisuelle	80
3.4.1	Indexation textuelle	80
3.4.2	Recherche par similarités et évolutions	81
3.4.3	Présentation de documents	82
3.4.4	Génération de documents	88
3.5	Conclusion	89

Nous avons vu que l'annotation d'un document audiovisuel était le concept de base de notre étude de la modélisation et de la représentation de documents audiovisuels numériques. Il s'agit en effet d'attacher des descriptions à des parties de documents audiovisuels. Dans ce chapitre, nous allons tout d'abord rapidement présenter quels domaines de recherche se sont intéressés aux documents audiovisuels, afin de fournir un cadre permettant de comprendre pourquoi telles et telles analyses ont pu être proposées. Nous analysons ensuite quelles sont les diverses caractéristiques d'annotation utilisées, et quelles structurations de document (donc d'ensemble de caractéristiques) ont été proposées. Nous décrivons enfin quelques fonctionnalités de systèmes d'information audiovisuelle intéressantes et/ou paradigmatiques présentées dans la littérature, ainsi que quelques systèmes existants.

3.1 Des champs de recherche et des problématiques différentes

Les travaux que nous avons étudiés ont tous comme point commun de s'occuper de documents audiovisuels. Cependant, les approches proposées, les systèmes mis au point, la manière même d'appréhender les documents varient suivant la communauté de recherche des auteurs. De la même manière que les préoccupations d'une communauté varient, la considération de l'audiovisuel numérique n'a pas non plus été la même partout. Il n'existe donc pas de communauté de l'audiovisuel numérique stable et établie, bien que la tendance soit à la convergence, suivant en cela ce qui se passe autour du Web et du multimédia en général.

Nous présentons dans la suite quelles approches ont pu être développées en rapport avec les documents audiovisuels. Bien évidemment, les divisions que nous fournissons sont caricaturales, et de nombreux travaux transcendant les domaines existent. Nous espérons seulement que cette division permettra d'appréhender mieux l'état de l'art de la représentation audiovisuelle que nous présentons dans ce chapitre.

Traitement du signal. Le signal ici traité est le plus souvent le signal image, et l'objectif ancien et avoué est de calculer, de trouver le sens d'une image, c'est à dire quels sont les objets qu'on y trouve, leurs relations, objets que l'on pourra ensuite utiliser, par exemple pour choisir le chemin d'un robot, ou envisager la présence d'une tumeur. Avec la possibilité de créer de grandes bases d'images se pose la question de retrouver une image dans la base à partir d'une description. Cette description s'exprime dans un langage d'indexation, qui peut être basé sur des primitives de très bas-niveau, par exemple un taux de couleur, ou encore des formes et leurs relations. Au lieu d'interroger directement la base en fournissant les descripteurs de recherche, il est également possible de les calculer à partir d'une image présentée par l'utilisateur. Ils sont alors comparés aux descripteurs de la base en fonction de *mesures de similarité* [Aigrain *et al.*, 1996]. Les résultats ne sont pas binaires, mais peuvent être classés : la réponse à une requête n'est pas une partition de la base, mais un classement des documents de celle-ci en fonction de leur similarité [Santini et Jain, 1996]. S'ajoute également la possibilité de combiner différents opérateurs de similarité et d'utiliser du bouclage de pertinence [Picard *et al.*, 1996, Minka, 1996].

Avec l'image animée, plusieurs champs de recherche s'ouvrent, tenant compte des caractéristiques temporelles du flux. Il devient par exemple possible d'essayer de repérer des mouvements d'objets ou de caméra [Aigrain *et al.*, 1996] ; ou de segmenter le flux, c'est à dire de repérer des ruptures de plans. Enfin, considérant qu'une image bien choisie peut être représentative d'un plan (keyframe), il devient possible de retrouver des plans par similarité d'image.

La notion de document n'est en général pas prise en compte.

Bases de données. Les spécialistes des bases de données se sont attaqués aux documents audiovisuels suivant deux axes différents. Le premier axe concerne les flux audiovisuels eux-mêmes, il s'agit alors de considérer des bases de très grande taille, et d'organiser celles-ci de telle sorte que les requêtes sur les flux puissent être servies avec des débits de qualité suffisante. Cela peut par exemple se faire en organisant les données de telle sorte que les plus demandées soient les plus accessibles, et en parallélisant les accès [Mostefaoui *et al.*, 1999].

L'autre point de vue concerne les index, leur organisation et les requêtes qu'il est possible de mener dessus. Les schémas de description choisis sont en général des extensions de schémas classiques intégrant la composante temporelle des descripteurs. Certaines études prennent en compte la notion de document audiovisuel, c'est à dire qu'elles cherchent à définir des schémas de description adaptés aux documents et à leur décomposition [Hjelsvold *et al.*, 1995a, Adali *et al.*, 1996],

d'autres se contentent de décrire des objets temporels sans s'intéresser à la sémantique de ceux-ci [Decleir *et al.*, 1999].

L'intégration du multimédia aux bases de données est un processus difficile au sens où les extensions ne suffisent pas et qu'il convient de prendre en compte les composantes image et son des documents [Grosky, 1997], c'est à dire les descripteurs de bas-niveau. D'autre part, l'extension des bases de données à la recherche d'information, c'est à dire à un processus de recherche de données incomplètes, partiales, et ne répondant pas obligatoirement à des schémas figés n'est pas évidente, mais les bases de données semi-structurées [McHugh *et al.*, 1997] s'attaquent désormais à ces problèmes.

Recherche d'information. La recherche d'information classique s'est principalement intéressée aux documents audiovisuels en tant que ceux-ci sont décrits comme des documents quelconques, par exemple à l'aide d'un ensemble de mots-clé, ce qui permet d'appliquer des algorithmes stables et réputés pour leur recherche. On notera que les seuls systèmes opérationnels et utilisés à l'heure actuelle reposent sur de telles approches, en se basant sur l'utilisation des sous-titres ou de la reconnaissance vocale des paroles pour mettre en place des textes d'indexation qui pourront alors être traités par des approches standards [Brown *et al.*, 1995]. Ce genre d'approche est également très utilisé au Japon [Takeshita *et al.*, 1995, Ariki *et al.*, 1998].

Multimédia / synthèse. Le multimédia s'intéresse au document audiovisuel de deux manières. D'une part il s'agit d'intégrer des documents audiovisuels aux documents multimédias, en tant que « briques de construction », et donc d'en gérer la temporalité et les comportements [Sabry-Ismail *et al.*, 1997].

Il s'agit d'autre part d'accroître les capacités multimédia du document audiovisuel, principalement en y ajoutant la possibilité de navigation suivant des liens insérés dans les images [Nabeshima *et al.*, 1998, Geißler, 1995], ce qui peut conduire à créer des documents à structure non temporalisée de façon unique, puisqu'on pourra par exemple choisir comment se déroule un film, en intervenant directement sur celui-ci [Sawhney *et al.*, 1997].

La synthèse de documents audiovisuels artificiels, les formats permettant de décrire des mondes en trois dimensions sont des secteurs qui prennent de l'importance du fait du développement de MPEG4, puisqu'un flux peut alors être décrit en terme d'objets, de leurs mouvements et interactions.

Intelligence artificielle. Certains membres de la communauté d'intelligence artificielle se sont intéressés aux documents audiovisuels, dans le but de réutiliser par exemple des techniques d'organisation de termes de description en base de connaissances [Domeshek *et al.*, 1996]. D'autres travaux s'intéressent à l'édition automatique de documents, par exemple [Sack et Davis, 1994]. Plus généralement, il s'agit d'identifier les connaissances liées à l'exploitation d'un document audiovisuel, et les règles qui peuvent y être appliquées afin de réaliser une exploitation automatique (par exemple du montage).

Document structuré. Certains membres de la communauté des hypertextes et des hypermédias, c'est à dire des documents structurés tels que SGML, HyTime ou XML s'intéressent depuis peu aux documents audiovisuels, essayant de mettre en place des véritables définitions de types de documents (DTD) adaptées à l'audiovisuel [Auffret *et al.*, 1999]. L'objectif à long terme consiste à mettre en place des modèles de documents correspondant à des genres de documents, c'est à dire à décrire la structure d'un document et ses descripteurs dans des formats échangeables.

Toutes proportions gardées, il s'agit d'essayer qu'une communauté semblable à celle de la TEI fasse le même travail, mais cette fois sur des documents audiovisuels [Auffret *et al.*, 1999b].

Industrie. Terminons en évoquant un moteur de la recherche qui n'en est pas un acteur classique, mais se révèle extrêmement vivace. Les enjeux liés à la distribution de contenus audiovisuels numériques sont en effet très importants, et nombre de sociétés (industrie du logiciel multimédia, fournisseurs de contenu, diffuseurs) se réunissent pour essayer de définir les normes qui permettront ces échanges. Le consortium W3C¹ développe par exemple simultanément un grand nombre de standards, dans des délais relativement impressionnants.

En conclusion. Il nous paraissait nécessaire d'essayer de fixer grossièrement — et à la manière d'un contexte de lecture — quels étaient les enjeux de recherche, les objectifs de différentes communautés. Notons cependant que les travaux inter-communautaires existent et ne peuvent que se développer pour permettre d'avoir une vue unifiée sur les documents audiovisuels, notamment les bases de données, la recherche d'information structurée et l'intelligence artificielle (gestion des connaissances).

La suite du présent chapitre présente une analyse de l'annotation de documents audiovisuels en termes de caractéristiques de modélisation et de structuration.

3.2 Caractéristiques d'annotation de documents audiovisuels

Le terme *contenu* (*content* en anglais) apparaît plus souvent qu'à son tour dans les titres des articles consacrés aux SRI audiovisuels. Cependant, il importe de réaliser que pour beaucoup d'auteurs (pour la plupart issus du traitement d'images), ce terme ne regroupe le plus souvent que des caractéristiques extraites automatiquement des données brutes. Cela est particulièrement frappant par exemple dans [Aigrain *et al.*, 1996] : «*Not only content-based retrieval reduces the high variability among human indexers. . .*». Dans l'esprit des auteurs, la recherche basée sur le contenu ne s'appuie que sur une détection automatique.

Il nous semble que la notion de contenu — ou au moins de *contenu sémantique* s'il faut préciser — doit s'appliquer à des concepts de tout niveau. Ainsi un histogramme de couleur fait partie du contenu d'une image au même titre qu'un objet détecté y aurait sa place, et *a fortiori* toutes abstractions conceptuelles. Le contenu est donc simplement dépendant de la personne qui le met en place et qui le lit, et pourra être plus ou moins bien partagé.

3.2.1 De la granularité de l'annotation

Les caractéristiques d'annotation de document audiovisuel représentent des contenus de documents audiovisuels de tout niveau. Ces représentations de contenus sont attachés à des parties de documents. Il faut donc s'interroger sur la manière dont se fait le découpage de ces parties, et quelles en sont leurs caractéristiques.

Deux approches de segmentation se distinguent généralement : la *segmentation a priori* du document, et la segmentation *au besoin*, aussi appelée approche de *stratification*.

Segmentation *a priori*

La segmentation *a priori* suppose qu'il existe des unités du document audiovisuel qui peuvent être mises en place dans un premier temps, l'annotation, la description de ces unités se faisant

1. <http://www.w3c.org>.

dans un deuxième temps. L'unité de base considérée est en général le *plan*, car celui-ci correspond à une unité de montage, et — de façon présupposée — d'analyse.

Comme les plans peuvent de plus être détectés dans leur majorité de façon automatique², la segmentation *a priori* en plans semble désormais être un standard de fait de la représentation audiovisuelle.

Prenant en compte le document dans son ensemble, le niveau suivant de segmentation *a priori* concerne le regroupement des plans en *scènes*, et des scènes en *documents*, avec éventuellement des regroupements intermédiaires. Le problème principal de cette approche — nous reviendrons dessus — réside dans le fait que d'une part la décomposition primaire du document est forcément un arbre, que d'autre part si la détection et la définition des plans ne souffre pas de discussion, celles des scènes sont beaucoup plus ambiguës.

Stratification

L'approche de stratification [Davenport *et al.*, 1991] prend pour principe que toute mise en place d'une annotation correspond *en même temps* à la définition du segment audiovisuel annoté. C'est par exemple parce que le personnage « Chirac » est repéré dans un document que l'on va définir une strate, laquelle sera alors annotée par le mot-clé *Chirac*.

L'approche de segmentation *a priori* est en effet critiquée par un certain nombre d'auteurs, au nombre desquels Davis [1993], en se basant sur le fait qu'en dehors d'eux-mêmes, plans et séquences ne représentent pas grand chose, et que les effets de montage sont là pour nous le prouver (*cf.* 2.1.4). Le contenu doit donc être considéré de la façon la plus globale possible dans le document, par exemple un simple descripteur vidéo tel qu'un mot-clé attaché à un personnage peut se prolonger sur plusieurs plans, et on ne doit pas suivre à la lettre les limites « syntaxiques » imposées par la segmentation. La redondance de l'annotation de plans différents mais ayant un élément de contenu similaire est un autre élément en défaveur de la segmentation pour Davis, ainsi que la perte de linéarité qui s'ensuit (ce qui semble un comble pour un média temporel!).

En conséquence, suivant en cela [Aguierre Smith et Davenport, 1992, Aguiere-Smith, 1992], Davis propose non plus d'attacher des descripteurs de contenu à des segments dûments repérés mais à des strates simplement repérées par leur instants début et leur fin.

La stratification permet la cohabitation de considérations différentes sur des morceaux de vidéo identiques. Son indépendance par rapport à toute segmentation permet de définir des strates liées à tout niveau d'analyse du document que celle-ci soit consacrée aux textures ou aux personnages.

La stratification lie donc de façon indissoluble l'annotation et le segment annoté. Un document annoté est alors un ensemble de strates et leurs annotations. Une structuration des strates entre elles n'existe alors pas *a priori*.

Granularité de l'annotation

La *granularité* de l'annotation globale d'un document audiovisuel correspond au degré de précision de l'annotation, c'est à dire à la taille des segments considérés dans le processus d'annotation. Il est en effet possible d'annoter une simple image d'un document comme le document dans son ensemble.

La notion de granularité peut s'étendre aux considération spatio-temporelles sur un document. Ainsi on peut considérer différentes granularités d'analyse d'une image détemporalisée,

2. Aux problèmes liés aux transitions « exotiques » entre plans prêt : un fondu enchaîné ou un volet seront moins évidents à repérer automatiquement qu'un simple *cut*.

ou bien d'un flux, c'est à dire par exemple d'un objet filmé qui se maintient à l'image pendant quelques secondes.

Il va de soi que les caractéristiques d'annotation d'un document audiovisuel dépendront du niveau de granularité choisi. On annotera par exemple un document avec le nom de son réalisateur³, tandis qu'une image pourra être décrite par une composante globale de texture. Ces choix de granularité sont également guidés par les buts de l'application à mettre en place, et ce au niveau de l'indexation, mais aussi à celui de la présentation des résultats : dans un système documentaire audiovisuel basé sur des notices textuelles comme Questel⁴. par exemple, le grain choisi est gros, puisqu'une notice représente un document. Dans un système plus évolué, on pourra choisir d'indexer plan par plan, et d'offrir la possibilité à l'utilisateur de visionner les documents retrouvés plan par plan (mais on aurait aussi pu indexer le document globalement, tout en gardant la visualisation au niveau du plan).

Quelques auteurs proposent une classification des caractéristiques liées à l'audiovisuel, Jain [1996] considère trois classes de caractéristiques liées aux images :

- la classe F_u contient des *méta-caractéristiques*, qui peuvent être automatiquement extraites des informations associées aux documents (taille, auteur, format, *etc.*), ou bien fournies directement par l'utilisateur. Ces caractéristiques ne sont pas directement extraites des images.
- F_d contient des caractéristiques directement extraites (*dérivées*) des images au moment de leur insertion dans la base de documents, de façon automatique ou semi-automatique (taux de couleurs, *etc.*) ;
- F_c contient des caractéristiques également extraites des documents, mais qui ne le sont qu'au moment où le besoin s'en fait sentir.

Rowe *et al.* [Rowe *et al.*, 1995] après une étude des utilisateurs et des types de requêtes audiovisuelles considèrent trois types de caractéristiques :

- les *données bibliographiques* (titre, résumé, sujet, genre, producteur, acteurs, *etc.* ;
- les *données de structure* (hiérarchie de plans, scènes, séquences) ;
- les *données de contenu* (*content data*) qui peuvent être 1) un ensemble d'images-clé du document ; 2) des mots-clés extraits de la bande-son ou des sous-titres ; 3) des indices d'objets qui indiquent les images d'entrée et de sortie de chaque objet ou personnage significatif.

Dans [Jain et Hampapur, 1994], Jain *et al.* présentent le modèle de données ViMod de représentation de documents audiovisuels, et les caractéristiques utiles, qui sont réparties en cinq classes en fonction :

- de leur *dépendance au contenu* (extractibles des données vidéo brutes ou non) ;

3. Par exemple, dans le schéma d'indexation de l'INRIA pour les « audiovisuels » en 1982 [Calderan *et al.*, 1982], les données de description globales sont réparties en données de catalogage (Signalement : titre, durée, support, documents d'accompagnement ; Production-Réalisation : date, pays, producteur, auteur... ; Distribution) et données d'analyse (Contenu : explicitation du contenu, résumé, descripteurs ; Traitement audiovisuel : genre, analyse globale, son, image ; Utilisations : utilisations possibles, public, observations ; Appréciations et compléments : points forts, nature de l'analyse effectuée, *etc.*).

4. <http://www.questel.orbit.com/>

- de leur *extension temporelle* (spécifiables ou non à partir d'une image seule) ;
- de la possibilité de les *qualifier* (*Q-features*) selon un modèle du domaine — par exemple une passe dans le contexte d'un match de basket — ou non, — par exemple une donnée numérique exprimant qu'une forme ronde se déplace *R-features* (*raw features*).

Nous classifions, en accord avec ce que nous avons déjà pressenti, les caractéristiques d'annotation en deux classes : les caractéristiques *calculées* par la machine (par exemple une forme rectangulaire) et en caractéristiques *interprétées* par un individu (par exemple une fenêtre). Un deuxième niveau critère de comparaison concerne les caractéristiques temporelles (prenant en compte le flux temporel) ou non.

Nous suivons dans la suite un ordre de présentation se basant sur le degré d'intelligibilité des caractéristiques, depuis les primitives (*features* des publications anglaises) de bas-niveau jusqu'aux caractéristiques interprétées de haut-niveau (mises en place de façon semi-automatique ou manuelle). Un degré intermédiaire concerne les caractéristiques calculées ayant vocation à ressembler à des caractéristiques interprétées.

3.2.2 Primitives

Signatures globales images ou son. Les signatures globales correspondent à un calcul sur l'ensemble d'une image ou la longueur d'une « strate sonore », et concernent le plus souvent les textures ou les histogrammes couleurs, sur lesquels sont réalisés des similarités. Ces primitives correspondent au plus bas-niveau de l'analyse, et nombre de systèmes ne reposent que sur elles, par exemple [La Cascia et Ardizzone, 1996, Chiueh, 1994, Flickner *et al.*, 1995, Kelly *et al.*, 1995, Wang et Adelson, 1994]. Ces primitives correspondent à ce qui est parfois appelé *niveau physique* [Mechkour, 1995]. L'extraction d'un spectre sonore relève de ce niveau de signature.

Contours, formes, positions relatives. L'extraction de régions (ou « objets visuels ») se base sur la détection de contours et de formes en prenant en compte les couleurs, les textures, la luminance ou toute autre caractéristique des images [Zhong et Chang, 1997, Seyrat *et al.*, 1998]. Les positions relatives entre régions peuvent également être calculées, l'image étant considérée comme un graphique dont les éléments sont détectés. Ce niveau peut également être appelé *niveau image* [Mechkour, 1995]. Les similarités peuvent ici se baser sur des comparaisons de régions en termes de forme et de couleur, ou encore sur des relations respectives entre régions.

Points d'intérêt. Chercher des points d'intérêt dans une image consiste à essayer d'y trouver des points invariants qui en soient caractéristiques, le plus souvent placés sur de forts gradients de couleurs. Par exemple [Bres et Jolion, 1999] réduisent ainsi une image à quelques centaines de points d'intérêt contenant des informations de couleurs, mais également de forme, sur lesquels il est possible de calculer des similarités.

Mouvements de caméra et d'objets. Comparer deux images permet d'en trouver les différences, donc de détecter des mouvements globaux de formes, par exemple d'objets, qu'on peut alors coder avec leur couleur, leur texture, leur taille, leur trajectoire et leur vitesse, et rechercher avec ces critères [Chang *et al.*, 1997b, Yoshitaka *et al.*, 1996, Dimitrova et Golshani, 1995], . Il est également possible de repérer les mouvements de caméra, ce qui correspond à un mouvement

global du fond de l'image [Etievent *et al.*, 1999], ou bien d'extraire les arrières plan d'une scène [Wang et Adelson, 1994].

Sons. Les travaux sur les primitives sonores concernent principalement des caractéristiques perceptuelles et acoustiques permettant de retrouver par exemple des mélodies ou des sons [Wold *et al.*, 1996, Uitdenborgerd et Zobel, 1998], ou de détecter des occurrences de mots recherchés. Une étude des travaux sur les systèmes de recherche d'informations sonores se trouve dans [Foote, 1999], la navigation dans les bases sonores est également étudiée. A notre connaissance, ces travaux ne sont pas encore intégrés aux systèmes d'information audiovisuels, mais les travaux sur la norme MPEG-7⁵ prennent en compte cette dimension.

3.2.3 Caractéristiques calculées de haut-niveau d'abstraction

L'objectif principal du traitement automatique est de définir des caractéristiques qui aient une signification pour une majorité d'êtres humains, c'est à dire qui puissent être interprétées facilement. Tout est bien évidemment une question de degré : un taux de couleur aura sans doute du sens pour le scientifique qui l'a calculé, lequel fera le lien entre les images et les résultats de calcul, et sera donc capable d'interroger un système basé sur elles, tandis que pour le reste du monde la primitive sera incompréhensible.

L'enjeu principal est donc de passer à des caractéristiques de *haut-niveau*, c'est à dire ayant des résultats qui ne soient plus sous forme numérique, par exemple faire le lien entre un changement numérique d'histogramme de couleur et une rupture de plan. Parmi la multiplicité des traitements numériques possibles sur un flux audiovisuel, il s'agit donc de trouver ceux qui permettent d'atteindre au contenu de l'image tel que l'homme l'interprète. Mais il ne s'agit pas non plus de s'illusionner, nous avons en effet vu que l'image était un matériau qui nécessitait une interprétation, c'est à dire au minimum quelques prescriptions culturelles ou factuelles. C'est pourquoi un certain nombre de traitements font également appel à des connaissances *extérieures* au document, ou au moins des connaissances sur le type de document afin de pouvoir mettre en place des caractéristiques de haut-niveau.

Les caractéristiques calculées ont toutes des résultats sous forme numérique, et c'est l'interprétation que l'utilisateur fera de ces derniers qui permettra de les qualifier selon des caractéristiques interprétées. Par exemple, le repérage d'un ensemble de vecteurs mouvements sur les bords d'images successives ne représente que lui même. Si le dépassement d'un seuil mène à un résultat binaire 0 ou 1 pour une caractéristique, celle-ci n'en reste pas moins calculée. Alors soit le résultat du calcul est validé à la mise en place, donc interprété ; soit l'utilisateur est conscient d'avoir devant lui le résultat d'un calcul, et a accès à l'algorithme correspondant, il peut donc interpréter ce résultat ; soit l'origine de la caractéristique a disparue, et toutes les erreurs d'interprétation sont possibles.

Nous présentons dans la suite quelques classes de calcul correspondant à des caractéristiques calculées de haut-niveau.

Ruptures de plans. Comme nous l'avons vu, le plan est l'unité de montage vidéo, et est considéré comme l'unité minimale d'analyse. Comme de plus il est relativement aisé de calculer les ruptures de plans nettes à l'aide de caractéristiques globales, on concevra qu'une littérature pléthorique soit consacrée au sujet depuis de nombreuses années. Les difficultés principales consistent d'une part à éviter les fausses ruptures de plans (par exemple dues à des mouvements

5. Le format de compression de fichiers audio MPEG-3 n'a rien à voir avec une éventuelle indexation.

rapides de la caméra, ou bien à un objet qui occulte l'image), et à détecter convenablement les effets de montage différents du simple « cut » (par exemple les fondus ou les volets). On notera que le problème de la détection des plans ne se posera plus à moyen terme (sauf pour les documents anciens), puisque le montage numérique permet de récupérer exactement les plans tels qu'ils ont été découpés, qu'il suffira d'intégrer au format du document.

Détection de scènes. La problématique de la détection de scènes résulte de la volonté de dépasser le niveau du plan pour atteindre à des unités structurelles de plus haut-niveau sémantique. Il s'agit donc de regrouper des plans en des unités ayant une continuité sémantique sous le rapport d'une action dramatique déterminée.

Plusieurs méthodes ont été utilisées. Une méthode totalement automatique repose sur la comparaison de similarités d'images entre plans (*i.e.* entre images considérées comme représentatives de plans). Par exemple [Yeung *et al.*, 1996, Yeung et Yeo, 1996] mettent en place des graphes de similarités entre plans, et considèrent des groupes (clusters) de plans similaires temporellement liés, qui sont alors supposés représenter des « unités d'histoire », et donc des scènes du documents. L'analyse menée sur des séries télévisées donne des résultats intéressants, mais il convient d'en considérer les limites : si une des règles de mise en place de fiction implique une unité de lieu correspondant à une scène, c'est dans le cas des séries télévisées que ce principe est appliqué de façon caricaturale (nombre de décors limité, peu de mouvements de caméra), acteurs relativement statiques, *etc.* Des travaux du même ordre ont été menés par [Chen *et al.*, 1998b, Kobla *et al.*, 1997].

Il est également possible d'utiliser d'autres connaissances de règles de montage pour détecter des unités structurelles de plus haut-niveau que les plans. Par exemple [Aigrain *et al.*, 1995b] ont mené un travail d'expertise auprès de spécialistes de l'audiovisuel, et mis en place quelques règles permettant de décider de reconnaître des séquences. Ces règles portent par exemple sur les types d'effets de transitions, la répétition de plans, l'apparition de la musique, le rythme d'édition, ou les mouvements de caméra et permettent de mettre en place une macro-segmentation basée sur des opérations de fusion, de précédence, *etc.* [Bolle *et al.*, 1998] ont également travaillé sur ce sujet, et détectent des types d'événements (dialogue, narration), des nombres de personnages. Si par exemple des plans similaires deux à deux en classes A et B s'enchaînent sous la forme ABABA, on pourra alors décider de regrouper tous les plans en une scène de dialogue (champs/contre-champs). Dans le même ordre d'idée, [Carrive *et al.*, 1998] propose de regrouper les plans en scènes à l'aide de règles de subsomption portant sur des concepts de montage cinématographique décrits dans une logique de description étendue à la gestion de règles temporelles. [Li *et al.*, 1996] regroupe les plans suivant des critères issus de l'analyse de la bande-audio. Des plans contigus et ayant une bande sonore similaire (typiquement un commentaire dit par une même personne) sont fusionnés dans une même scène.

[Wachman, 1996] a pour objectif de relier de façon automatique le synopsis détaillé d'un épisode d'une série télévisée à la bande vidéo, afin de segmenter celle-ci. La reconnaissance des personnages (aidée par l'utilisateur) et l'utilisation des sous-titres permettent la décomposition du document en scènes et en unités structurelles telles qu'indiquées dans le script.

Travail sur les genres de documents. Les travaux que nous venons de citer s'appuient en fait sur des connaissances extérieures au document, qui portent sur la nature suffisamment stéréotypée de ceux-ci pour pouvoir en reconnaître la structure. Si certaines de règles sont de nature relativement générale par rapport aux documents audiovisuels (dialogue = ABAB), c'est souvent sur le genre du document que l'on s'appuie.

Par exemple, [Swanberg *et al.*, 1993] décomposent un document de genre extrêmement pré-

cis⁶ « Journal Télévisé CNN » en une structure prédéterminée, en se basant sur des diagrammes d'états et des modèles d'épisode à l'aide d'un automate. Les séquences plateau sont ainsi reconnues et séparées des séquences de reportage. [Faudemay *et al.*, 1997] tirent également partie du genre du document.

La détection de formes, et leur qualification peut être guidée par des connaissances de genre. Par exemple si on étudie des journaux télévisés d'une chaîne sur une période donnée, alors la position du présentateur ne change pas dans certains plans, et donc il est possible de détecter sa présence plus facilement [Merialdo, 1998].

Des modèles de genre de documents peuvent également être utilisés pour reconnaître et identifier des documents comme leur appartenant. Par exemple [Fischer *et al.*, 1995] séparent la publicité des actualités.

Travail sur le son. Nous mettons dans cette catégorie de travaux tout ce qui concerne la reconnaissance automatique de la parole, la transcription des dialogues par exemple, l'identification de la langue d'un locuteur, l'identification de locuteurs, *etc.* Si la majorité de ces travaux ne sont pas propres aux documents audiovisuels, d'autres sont plus spécialisés sur certains genres de documents (par exemple les reportages touristiques). [Taniguchi *et al.*, 1995] détecte ainsi l'apparition et la disparition de musique ou les applaudissements comme significatifs. La recherche de mots-clé appartient également à cette catégorie.

Textes. L'extraction des textes d'une image se révèle de grand intérêt par exemple dans le cas de sous-titres, de bandeaux à l'écran. Les méthodes employées ne sont en général pas propres à l'audiovisuel, mais se basent sur les résultats de reconnaissance automatique de l'écriture.

Autres travaux. Parmi les autres travaux portant sur le calcul de caractéristiques de haut-niveau, il convient de citer la reconnaissance du jour et de la nuit [Faudemay *et al.*, 1996], des intérieurs et des extérieurs, de la ville et de la nature.

D'autres travaux encore s'attachent à détecter des passages importants dans un documents. [Nakamura et Kanade, 1997] font de la corrélation entre les images (visages, personnes, extérieurs) et les passages textuels (discours, réunion, visite, lieu, foule) « importantes » afin de distinguer quels sont les passages « importants » d'un journal télévisé. Le système Informedia [Christel, 1995] se base sur une coopération entre une analyse textuelle du synopsis, fréquentielle de la parole et de la vidéo (introduction de plans, plans similaires, mouvements d'objets, reconnaissance de personnages) afin de segmenter celle-ci en strates d'intérêt censées être les plus représentatives du document. [Taniguchi *et al.*, 1995] utilise des connaissances du domaine (en fonction des types d'émission considérés) pour repérer les *événements* signifiants dans le flux audiovisuel (changements de plan, opérations de caméra, mouvement d'objets, apparition/disparition de musique, applaudissements, ...).

En fait, plusieurs types d'approches s'opposent quand il s'agit de calculer des caractéristiques de haut-niveau à partir du signal brut.

Dans un premier type d'approche, il s'agit d'identifier les connaissances du domaine et de les lier à des « indices » calculables par la machine. C'est par exemple le cas de [Cavazza *et al.*, 1998] qui montre que pour reconnaître des types de costumes dans une base d'images de vêtements, la position et le nombre de boutons sont des critères suffisamment discriminatoires. C'est encore le cas de la décomposition d'un journal télévisé en fonction d'un modèle.

6. Véritable « patron » de production.

Un deuxième type d'approche considère qu'il y a lieu d'*apprendre* de l'humain des indices repérant telle ou telle caractéristique de haut-niveau. Il s'agit alors de faire un lien entre des primitives de bas-niveau supposées fournir des briques de base de la représentation et des caractéristiques de haut-niveau sémantique. [Minka et Picard, 1997] apprennent par bouclage de pertinence (exemples positifs et négatifs fournis par l'utilisateur) des combinaisons de mesures de similarités décrivant des « concepts visuels ». C'est également un des objectifs des travaux de Chang [Chang *et al.*, 1997a] que de lier bas-niveau et haut-niveau en apprenant de la manière dont l'utilisateur cherche. Plus récemment, [Buijs et Lew, 1999] font de l'« apprentissage visuel de sémantique simple » toujours à partir d'exemples positifs et négatifs. [Wachman, 1996] enfin repère des apparitions de personnages de séries télévisées en opérant un calage du script sur l'image, et apprend à reconnaître les personnages sur confirmation de l'utilisateur. [Düffing, 1999] organise des descripteurs image de haut-niveau dans un hiérarchie, les termes les plus « concrets » (par exemple ciel) étant illustrés par des *réalisations* en terme de primitives image (majorité de bleu par exemple). Enfin, des approches de types « fouille de données » se mettent en place afin d'apprendre des descriptions des concepts par classification d'ensemble de pondération de caractéristiques dans les requêtes [Bouet et Djeraba, 1998].

3.2.4 Caractéristiques interprétées

Les caractéristiques interprétées sont en fait tout ce qui est mis en place en présence de l'humain, que ce soit de façon semi-automatique ou manuelle. Dans le premier cas, le système automatique fournit alors un *assistant à l'annotation* qui fait une partie du travail de l'utilisateur, et propose à la validation les caractéristiques qu'il a extraites.

Davis [Davis, 1993] considère que la première tâche de représentation du contenu d'un document audiovisuel consiste à se donner un ensemble d'unités à même de représenter les flux temporels d'images et de sons, donc de parler d'événements temporels, d'actions humaines et d'objets dans l'espace, de personnages, d'objets impliqués dans les actions, de mise en scène, des propriétés du médium d'enregistrement, des positions relatives des objets dans les deux dimensions de l'écran, de montage (transitions cinématiques), de pensées subjectives à propos du document, *etc.*

Davis utilise alors de icônes comme caractéristiques, organisées dans un réseau de catégories⁷ permettant de les retrouver facilement, tandis que l'annotation proprement dite consiste à associer une ou plusieurs icônes à une strate.

Cette approche illustre plusieurs remarques concernant les caractéristiques de haut-niveau.

En premier lieu, celles-ci sont le plus souvent des termes, des mots-clé, associés à une partie de document audiovisuel. Cette partie peut être un morceau temporel, mais aussi une zone image. Par exemple, les systèmes FourEyes [Picard, 1996] ou QBIC [Flickner *et al.*, 1995] permettent à l'utilisateur de spécifier une zone à l'écran, que le système étend automatiquement à un repérage d'objet, il suffirait alors de nommer l'objet image repéré pour l'annoter

En deuxième lieu, les termes sont le plus souvent organisés en catégories correspondant à des niveaux de descriptions. Par exemple, [Domeshek *et al.*, 1996] organisent les descripteurs en objets, lieux, temps, activités et personnes. L'activité est considérée comme centrale pour la recherche, et des liens dans la base de connaissance permettent de mettre en relation des concepts proches, par exemple l'activité *naviguer* sera connectée au lieu *océan* et à la chose *bateau*. Dans le

7. Catégories non limitatives: space, time, wheather, characters, objects, characters actions, objects actions, relative position, screen position, recording medium, cinematography, shot transitions, subjective thoughts about the material.

thésaurus de l'INA, les termes sont organisés en quatre grandes catégories : personne physique⁸, personne morale, noms propres géographique, nom commun. Les relations du thésaurus sont les suivantes *scope note* (note d'application du terme), *synonymie* (utilisé pour), relation *associative* ou *de voisinage* (voir aussi), relations *hiérarchiques* (terme spécifique ou terme générique). Les relations hiérarchiques ne sont utilisées que pour les nom propres géographiques et les noms communs.

En troisième lieu, différents niveaux de description par des caractéristiques existent : du simple nom d'un personnage à des remarques sur l'atmosphère d'un film, toutes est possible. Les termes autorisés et utilisés dépendent bien entendu des visées de l'indexation. [Adali *et al.*, 1996] décident par exemple que ne sont utiles que les objets vidéo, qui sont les entités présentes à l'écran, et les activités, qui sont des classes d'action : ils construisent leur schéma de base de données suivant ce principe.

Le cas particulier des images représentatives. Un cas particulier est le suivant : certains systèmes utilisent des *images-clé* (*K-frames*) extraites d'un plan. Celles-ci peuvent être comptées comme caractéristiques de ce plan, ainsi que les primitives qui en seront extraites (grâce à une transitivité de bon aloi). Lorsque l'on utilise une image clé comme annotation, on peut parler d'*indexation iconique*. [Haddad *et al.*, 1996] y voit un certain nombre d'avantages : l'expression du contenu est plus explicite et liée au média, et permet de faire (voire est) une synthèse de différents niveaux d'analyse ; il est possible d'en extraire des caractéristiques automatiquement (comme déjà remarqué plus haut) et donc de définir des mesures de similarité ou d'envisager des requêtes visuelles. Restent cependant un certain nombre de problèmes, liés au non captage de la dynamique du média, aux problèmes de choix de l'image représentative dans le plan, et à la limite du plan elle-même, qui n'autorise des recherches qu'à ce niveau.

Considérer une image clé comme caractéristique d'un plan relève donc du choix de détemporaliser le contenu image du plan, en suspendant éventuellement l'analyse automatique de celle-ci, ou en réservant l'image clé à l'interprétation : l'utilisateur navigant dans les images peut choisir celles qui conviennent à ses désirs et ainsi atteindre le contenu audiovisuel.

Conclusion. Nous avons distingué trois types parmi les caractéristiques qui sont les briques de base de l'annotation. D'abord, les primitives concernent principalement les données numériques non interprétables. Ensuite, les caractéristiques calculées de haut-niveau résultent d'une extraction plus ou moins automatique de caractéristiques interprétables, à partir du document audiovisuel, mais aussi de ses documents d'accompagnement et de connaissances sur les genres. Enfin, les caractéristiques interprétées sont mises en place par un utilisateur. Nous nous intéressons dans la partie suivante aux structurations possibles des caractéristiques.

3.3 Modèles de structuration

Considérant qu'il existe deux types d'informations dans les documents audiovisuels, à savoir celles qui ne dépendent pas du contexte — car attachées à un morceau de vidéo —, et celles qui en dépendent — car ce morceau est lui même immergé dans un document, Davis [Davis, 1993] considère que l'objectif de la représentation audiovisuelle en général est de fournir un cadre à même de déterminer, de représenter et de rendre compte de ces deux types d'informations.

8. Un mot-clé pour un acteur ou un personnage n'est créé qu'après au moins cinq apparitions : il ne suffit pas d'accéder à l'écran une fois pour passer à la postérité...

La représentation d'un document audiovisuel dans un système de recherche d'information contient alors des informations tout à la fois *locales*, c'est-à-dire attachées à des parties du document, et des informations *globales*, concernant le document dans son ensemble.

Les informations locales peuvent être intrinsèques au morceau considéré, elles sont alors organisées structurellement *en tant* qu'annotation (3.3.1). Elles peuvent également découler de considérations plus contextuelles, auquel cas la médiation de l'action du contexte se fait par le biais de la description structurelle du document, et ce de façon implicite ou explicite.

La notion d'« informations globales » recouvre toutes les informations ayant trait à la structure du document, structure *implicite* quand l'axe temporel est le seul lien entre parties (3.3.2), structure *explicite* quand cet axe temporel se redouble d'une structure logique hiérarchique de document, (3.3.3) ou d'une structure de graphe d'annotation (3.3.4).

Une recherche d'ordre temporel dans le document peut donc se faire dans tous les cas en utilisant l'axe temporel, mais si une structure explicite est disponible, il est possible d'utiliser celle-ci et les parties qu'elle organise.

3.3.1 Structuration des annotations

Suivant l'approche d'analyse que nous choisissons, une partie de document audiovisuel est décrite par une annotation construite à partir de caractéristiques audiovisuelles. L'annotation peut être simplement composée d'une caractéristique, par exemple on associe un histogramme de couleur à un plan ; ou d'un ensemble de caractéristiques.

Organisation d'un ensemble de caractéristiques

Cet ensemble peut être organisé (structuré) ou non. Par exemple, une suite de mots-clé annotant un document entier est un ensemble de caractéristiques non structuré⁹. Au contraire, si l'ordre des mots-clé à de l'importance¹⁰ alors on peut parler d'annotation structurée.

Ainsi, [Davis, 1993] considère la possibilité de construire des « phrases » d'icônes (*cf.* figure 3.1), par exemple du type **sujet - action - [objet|direction]** ou **sujet - position-Relative - objet** ou encore **camera - mouvement - objet** (*ex.* : « la caméra suit Steve »).



FIG. 3.1 – « Arnold enfle son manteau » et « la scène se passe au Texas, en haut d'une rue » dans *MediaStream*. L'icône Texas nous semble un argument contre l'universalité des icônes que suppose [Davis, 1993].

Similairement, organiser les caractéristiques en ensemble de couples d'attributs-valeurs participe d'une structuration. Par exemple, dans VideoStar [Hjesvold et Midtstraum, 1994], l'annotation attachée à un segment de document audiovisuel est divisée en *lieux*, *personnes*, et *événements* (dans le cadre des journaux télévisés), les attributs d'objets d'annotation dans le système OVID [Oomoto et Tanaka, 1993] sont équivalents.

Dans [Hibino et Rundensteiner, 1996], les objets descriptifs contenus dans les annotations de strates se limitent au champs suivants : nom (de la personne qui fait l'action), action, receveur

9. Nonobstant l'organisation *hors-documents* des descripteurs dans le thésaurus.

10. Comme par exemple pour les descripteurs d'ouvrages de la base Rameau de la Bibliothèque Nationale, où le deuxième terme de description, s'il est présent, est un qualificateur du premier [Genest, 1999].

(objet de l'action), et à deux champs libres (l'un à mots-clé, l'autre à texte libre). Enfin, pour Adali *et al.* [Adali *et al.*, 1996], l'annotation se limite à des descripteurs classifiés en *objets vidéo* (personnages, objets), en *types d'activités* (aller à la pêche), et en *événements*, qui sont des instanciations de types d'activité.

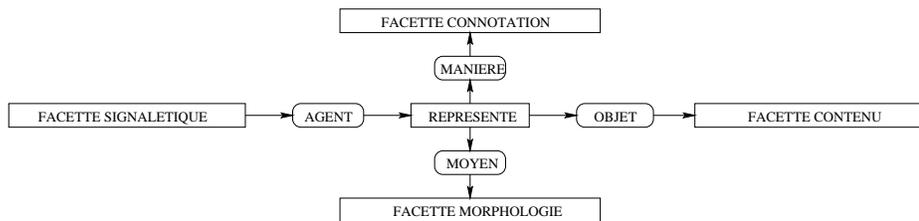


FIG. 3.2 – Modélisation générale d'un segment de vidéo sous la forme d'un graphe conceptuel

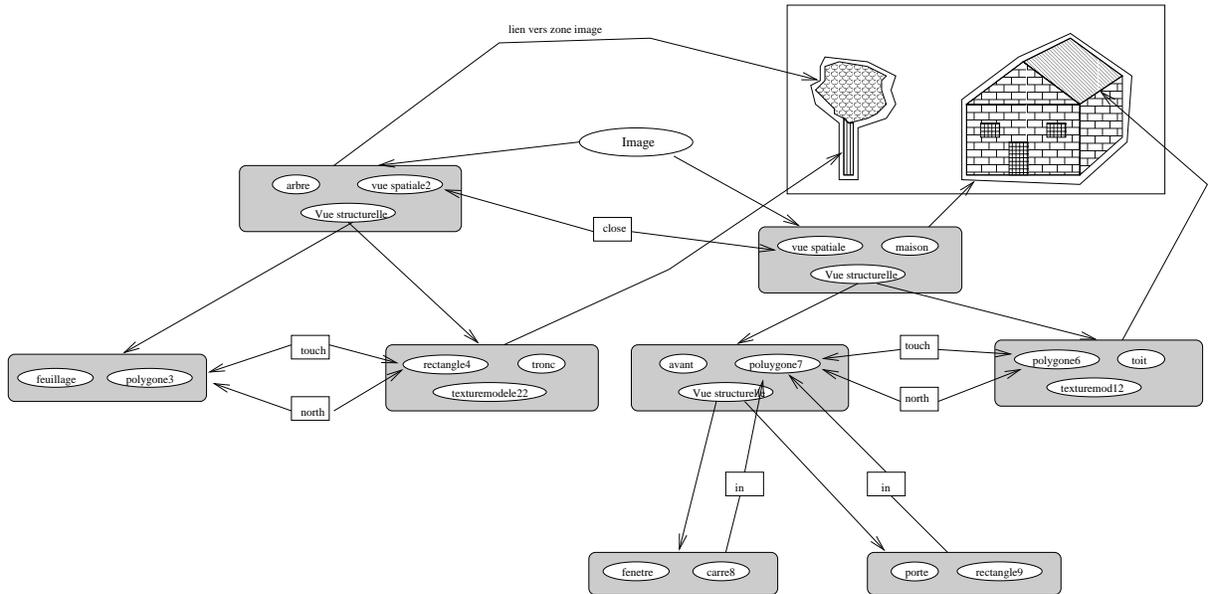
Il est possible de mêler caractéristiques de haut et de bas-niveau dans l'annotation. Par exemple [Simonnot, 1995] adapte l'annotation image basée sur les graphes conceptuels aux segments vidéo (figure 3.2). Un morceau de flux est alors décrit par des facettes, elles aussi organisées en graphes conceptuels. La facette signalétique contient les méta-caractéristiques classiques (auteur, date), et représente une facette contenu qui fournit des renseignements tels que la description d'actions, avec des possibilités de préciser les lieux, intervalles de temps, instruments. La facette morphologie (le « moyen de la description ») contient quant à elle des primitives image (couleur, type de plan, mouvement de caméra) et la facette connotation se limite à des descripteurs textuels. [Haddad, 1996] propose également de décrire des plan de façon multifacette pour réaliser une recherche d'information vectorielle.

Organisation du contenu image

Comme on peut considérer qu'une image représente un plan, l'annotation d'une image peut valoir pour celle du plan. Mechkour [Mechkour, 1995] propose avec le modèle *EMIR*² un modèle très riche de représentation d'image basé sur des objets-image à vues symboliques, spatiales, structurelles et perceptives intégrées avec les relations idoines dans une vue logique de l'image. La vue structurelle permet de définir dans l'image des objets, lesquels peuvent être décomposés à volonté en d'autres objets structurels. La vue spatiale et les relations spatiales permettent de spécifier les positions relatives des objets dans l'image. La vue perceptive concerne les primitives de couleur, luminosité et texture des objets, tandis que la vue symbolique permet de nommer et classer dans une hiérarchie de types les concepts de description de haut-niveau d'abstraction (*tronc* par exemple)¹¹. On définit également une relation *composé-de* permettant d'inclure les objets les uns dans les autres. Un exemple est donné figure 3.3. On extrait de la vue logique (*i.e.* de l'ensemble des vues et leurs relations) un graphe conceptuel de description de l'image, et les requêtes dans la base image se font à partir de graphes conceptuels de requête. [Agius et Angelides, 1999] proposent de représenter une image par les objets d'intérêt que l'on y trouve (personnages, objets), et les relations entre ceux-ci à l'image (par exemple *touches*, *beneath* ou *right*). La description se fait donc purement au niveau conceptuel.

[Hibino et Rundensteiner, 1996], qui s'intéressent en priorité aux requêtes spatio-temporelles sur les mouvements dans la vidéo organisent les annotations de segments avec la position dans l'image de l'objet en mouvement et un descriptif de cet objet (couleur, forme...).

11. Les liens entre objets-image (régions) et types sémantiques sont mis en place par un humain.

FIG. 3.3 – Exemple de description d'image dans EMIR²

[Corridoni *et al.*, 1996] considère des annotations structurées de plans par des descriptions spatio-temporelles d'objets (c'est-à-dire un ensemble de types instanciés et de relations spatiales et temporelles entre ces instances) qui reproduisent le contenu visuel en trois dimensions¹².

3.3.2 Structuration implicite de documents audiovisuels

La représentation implicite de la structure du document découle de l'utilisation des parties choisies qui y sont temporellement situées. La représentation du document audiovisuel est donc la somme des annotations temporellement situées qui le composent. Cette organisation implicite est commune à tous les systèmes d'annotation annotant un document en parties temporellement comparables (*i.e.* situées dans une même base temporelle).

Si certains auteurs négligent au niveau de la recherche la possibilité d'exploiter les relations temporelles implicites mises en place ([Davis, 1993] ne gère que des réponses comme intersections de strates annotées par les termes-icônes utilisés dans une requête), d'autres au contraire se basent sur elles comme fondamentales. Par exemple [Hibino et Rundensteiner, 1996] ont pour objectif avoué de permettre aux chercheurs de découvrir et de capter de façon assistée les relations entre objets et événements à l'intérieur d'une vidéo, ceci passant par la mise en place d'un outil graphique puissant d'expression de requêtes à caractère temporel.

Dans le système OVID, [Oomoto et Tanaka, 1993], dans l'objectif de tester la validité d'une base de données vidéo orientée objet, mettent en place un réseau sémantique associant annotation de strates et gestion du caractère temporel de la vidéo par le biais d'*objets vidéo* (*video-objects*)

12. Cette annotation est donc très riche, et difficile à mettre en place par l'utilisateur, c'est pourquoi un outil graphique d'aide à l'annotation de plan a été mis au point : il s'agit de spécifier dans un monde en trois dimensions vu du point de vue de la caméra les différents objets, leurs mouvements et interrelations, l'outil se chargeant de transcrire cette vue en description basée sur la *logique spatio-temporelle* STL [Del Bimbo *et al.*, 1995], qui est un langage de représentation de relations spatio-temporelles. Une projection de la vue en trois dimensions sur un plan en deux dimensions d'après la position de la caméra permet d'obtenir le plan tel que représenté sur l'écran. Les requêtes sur les plans peuvent alors se faire par l'exemple, avec un langage iconique par exemple.

évolués. Le paradigme objet (attribut/valeur classique) est étendu afin de prendre en compte les intervalles de temps, attributs des objets vidéos, à l'aide de la notion d'*héritage d'inclusion d'intervalles*, tandis que des opérations temporelles de composition telles que le recouvrement ou la projection sont définies dans une théorie formelle permettant l'édition de documents. Une hiérarchie spécialisation (*is_a*) pour les attributs textuels autorise l'héritage d'attribut.

La recherche dans la base objet se fait à la fois sur des requêtes textuelles de valeurs d'attributs, ou bien en naviguant dans la base d'objets, avec notamment la possibilité de désagrégation d'objets vidéo, basée sur l'héritage d'inclusion d'intervalles. Le modèle objet étendu permet de prendre en compte de façon simultanée le contenu des strates, mais aussi de leurs caractéristiques temporelles, avec les multiples possibilités d'héritage que cela permet.

3.3.3 Structuration hiérarchique

La structuration hiérarchique de documents audiovisuels se base sur le plan comme unité minimale (on est donc dans l'approche de la segmentation *a priori*), et sur l'organisation des plans en unités de plus haut niveau sémantique, telles les scènes. Celles-ci peuvent alors être regroupées au sein du document, ou dans d'autres unités encore. La structure du document est celle d'un arbre, qui peut s'apparenter à une structure documentaire classique telle qu'on la trouve dans les documents décrits dans des langages de balises (SGML, XML [XML, 1998, Michard, 1999]).

L'arbre peut être construit automatiquement si on fait de la détection de scène telle que déjà présentée en 3.2.3. Une approche plus générale, telle celle de [Chen *et al.*, 1998a] consiste à construire une « pyramide de similarité » par regroupement successif d'images de plus en plus représentatives du document (la pyramide est ensuite utilisée pour la recherche et la navigation).

L'arbre est cependant le plus souvent construit manuellement en suivant par exemple une hiérarchie document/scène/plan. A chacune des parties représentées par les nœuds de l'arbre sont associées des annotations. L'exemple suivant est représentatif de ce schéma général d'annotation, qui s'est mis en place quand le statut de *document* audiovisuel a été perçu en tant que tel¹³.

[Corridoni *et al.*, 1996] attachent aux différents segments de leur structure hiérarchique filmique un certain nombre de caractéristiques : les caractéristiques générales (auteur, date, titre) sont associées au document dans son ensemble, des descriptions textuelles aux scènes. Les plans sont quant à eux caractérisés par des primitives techniques d'une part (mouvement de caméra, angle, profondeur de champ), ainsi que par des objets. Les auteurs se basent sur les études sémiologiques de Metz [1977] et en insistant sur la nécessité de prendre en compte les unités structurelles du cinéma, considèrent un film suivant une métaphore linguistique, opposant l'axe syntaxique (les éléments structurels du film) et l'axe sémantique (contenu des unités syntaxiques précédemment définies) qui lui est orthogonal. Les transitions sont considérées comme « signes de ponctuation » de la syntaxe du cinéma. Un film est alors représenté comme illustré figure 3.4. La recherche dans un tel système peut s'assimiler à un appariement de sous-graphes dans le graphe global du film¹⁴.

La structure logique hiérarchique d'un document semble maintenant communément admise [Ponceleon *et al.*, 1998], et certains auteurs ont tenté de se baser sur elle pour mettre en place des DTD SGML de description de documents audiovisuels. [Carrer *et al.*, 1997] sont à notre

13. Par exemple [Chua et Ruan, 1995] annotent des plans avec des mots-clé, et regroupe ceux-ci en scènes, également annotées elles-mêmes, sans pour autant mettre en place de hiérarchie arborescente pour le document (on en reste au niveau de scènes définies à volonté).

14. A titre d'indication, l'annotation complète (après passage des outils automatiques) d'un film comme *Le guépard* par les techniciens et les experts de la Cinémathèque Toscane a nécessité une journée, ce qui est peu en regard de l'investissement en temps nécessaire à la réalisation d'un tel film.

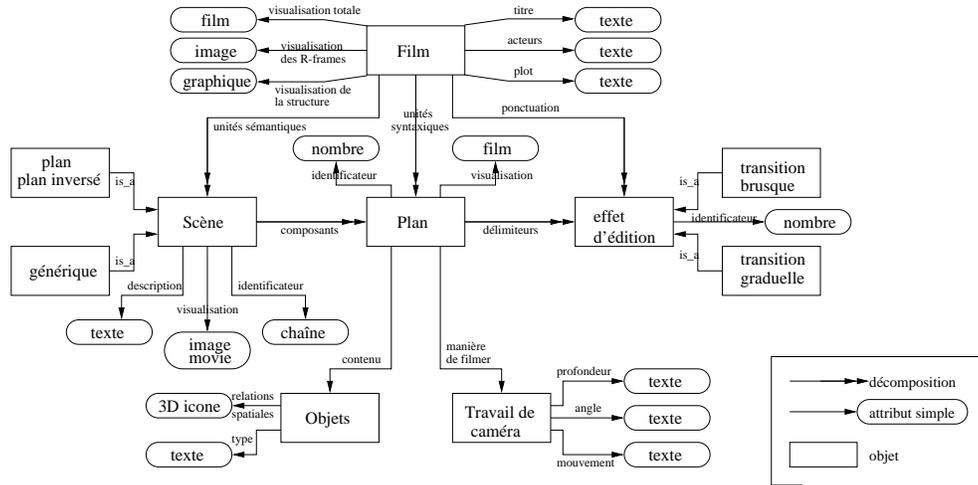


FIG. 3.4 – Schéma de graphe modélisant la structure d'un film.

connaissance les premiers à avoir stocké des annotations de documents audiovisuels sous la forme de documents structurés. Considérant en effet que la structure d'un document est un arbre à trois niveaux, ils décident de décrire les éléments de ces niveaux avec des éléments SGML dont les noms sont issus du domaine (par exemple un cours se divisera en sujets et sous-sujets). Les parties de documents sont alors annotées par des mots-clé à l'aide d'une interface d'annotation, qui offre la possibilité de modifier la DTD du document en fonction de l'annotation que l'on veut réaliser, et stocke les documents dans une base de données relationnelle pour les requêtes.

Un autre type de structuration hiérarchique « libre » est défini dans le système Algebraic Video [Weiss *et al.*, 1995]. Les annotations de parties de documents sont quelconques, et forment les feuilles d'un arbre qu'il est possible de construire en les connectant à des nœuds de niveau supérieur, eux-mêmes annotés. La structure globale est alors un modèle stratifié hiérarchique emboîté. Les auteurs proposent des primitives de manipulation des segments décrits par les feuilles et les nœuds supérieurs, notamment en vue de leur recombinaison, et l'héritage d'attributs entre nœuds est possible.

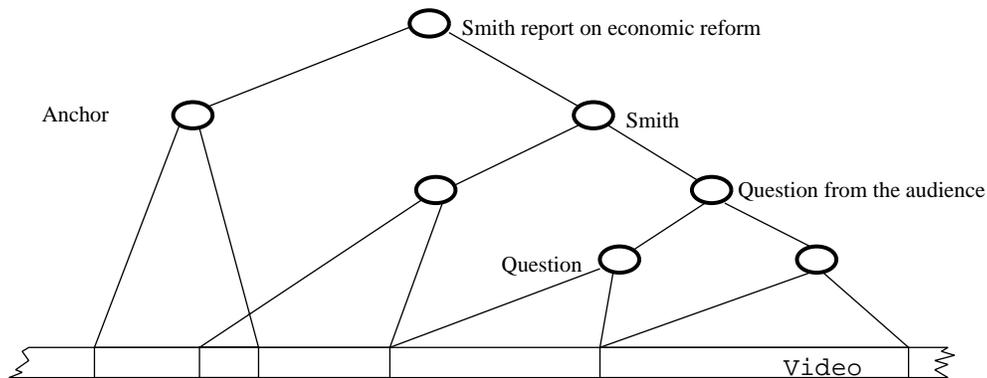


FIG. 3.5 – Schéma de graphe modélisant la structure d'un film.

3.3.4 Structuration par graphes

L'approche de [Zettsu *et al.*, 1997]¹⁵ est la suivante : certains « points » temporels du document audiovisuel sont annotés par un terme (timestamp) (un personnage, une action par exemple) dès que l'annotateur le désire. Les annotations sont ensuite mises en relations de trois manières : par des relations de généralisation (par exemple entre *avion* et *Chasseur F15*) ; de « sens commun » (par exemple entre *Aladin* et *Contes de mille et une nuits*) ; et des relations « normales » (notion assez floue¹⁶). La recherche se fait alors par mots-clés, les différentes relations entre ceux-ci servant à déterminer quelle partie du document il convient de rendre comme réponse.

L'approche de [Uehara *et al.*, 1996] se décompose en deux niveaux de structuration. D'un part des événements temporels sont représentés à l'aide de graphes de dépendance conceptuelle (dûs à Schank), dans lesquels des concepts *hors* du flux (le bandit, la santé, etc.), sont connectés à l'aide d'un nombre limité de marqueurs de primitives d'action (par exemple MOVE ou PTRANS pour un transfert physique), lesquels peuvent être reliées à des parties temporelles du document. Cela revient donc à avoir une description d'événement sous la forme d'un graphe dont certains nœuds sont connectés au flux. Un deuxième niveau de structuration — hiérarchique — permet de structurer les événements selon un arbre d'histoire généré grâce à une « grammaire d'histoire » due à Rummelhart (par exemple *story* se réécrit en *Setting + episodes*, et *episode* en *Event_Description*, etc.). Une fois la structure d'arbre organisant les graphes mise en place, la recherche se fait par recherche d'isomorphismes de sous-graphe. Les auteurs organisent également les graphes d'événement en les classifiant afin de permettre une recherche plus aisée, et proposent d'acquérir des concepts de l'annotation en repérant des sous-graphes qui se répètent. Ils considèrent également de la recherche « fondée sur l'inférence », laquelle consiste à traverser un graphe de description décrivant une action.

3.4 Fonctionnalités de systèmes d'information audiovisuelle

Après avoir présenté la manière dont il était possible de structurer les caractéristiques d'annotation pour la description de documents audiovisuels, nous décrivons dans cette partie quelques fonctionnalités de systèmes — ou systèmes complets — d'information audiovisuelle de la littérature.

3.4.1 Indexation textuelle

Les approches textuelles concernent en fait l'annotation de documents audiovisuels dans leur ensemble par un certain nombre de caractéristiques de *cataloguage* (auteur, titre, format) et de mots-clé, voire de caractéristiques sous la forme de textes libres.

Un exemple prototypique est le système utilisé à l'INA, du type Questel ou Mistral. Les notices sont stockées dans un système de gestion de documents, sur lequel la recherche se fait par mots-clé ou par données de cataloguage. Une interface graphique permet de construire les requêtes de façon simple, de les stocker, présente le thésaurus à l'utilisateur, etc. Les notices de l'Inathèque¹⁷ diffèrent légèrement par les champs et la considération du découpage d'un journal

15. Le lecteur notera que les deux travaux d'origine japonaise présentés ici ne nous étaient pas connus au moment de l'étude bibliographique fondatrice de notre travail.

16. Ces relations sont mises en place implicitement quand il y a co-occurrence entre descripteurs ou explicitement quand l'utilisateur le spécifie en mettant deux nœuds en relation.

17. <http://www.ina.fr/Inatheque/index.fr.html>.

télévisé en séquences. La Vidéothèque de Paris¹⁸ indexe ses documents en signalement, résumé, générique (personnes ayant participé à la création du document).

Un autre exemple concerne le moteur de recherche Altavista, qui reconnaît les documents audiovisuels indexés à leur extension, et collecte des mots-clé de description dans les pages Web qui entourent le fichier-document. Une image clé est également extraite de celui-ci. Il est alors possible de spécifier dans la requête qu'on cherche explicitement des documents audiovisuels.

[Merlino *et al.*, 1997] présentent un système utilisé par des chaînes de télévision, dans lequel les journaux télévisés sont découpés en reportages, et annotés par des champs Lieu, Personne, Date... L'annotation est mise en place automatiquement à partir des sous-titres (dans lesquels on détecte les mots-importants), et d'un modèle de journal télévisé.

Le système Informedia [Smith et Kanade, 1997], passé dans le domaine industriel¹⁹, annote les documents audiovisuel en repérant les passages importants par une analyse de la bande-son et de l'image. Un ensemble de mots-clé est alors extrait pour un passage important, et cette annotation est calée sur le flux temporel. La recherche se fait alors textuellement, à un tel point que [Christel et Martin, 1998] présentent une interface intelligente de présentation multidimensionnelle de mots-clé, dans laquelle l'utilisateur navigue, sans lien²⁰ avec le documents audiovisuel !

3.4.2 Recherche par similarités et évolutions

A l'opposé de la recherche textuelle pure, la recherche par similarité a la faveur d'une communauté issue du traitement du signal. Les systèmes utilisés se limitent le plus souvent à des prototypes.

On notera la possibilité de *recherche par similarité pure*, en utilisant une ou plusieurs primitives [Ardizzone et La Cascia, 1997, Zhang *et al.*, 1995, La Cascia et Ardizzone, 1996] dans des prototypes ayant pour granularité le plan.

La *navigation par similarités* est possible dans une base d'images (donc une base de documents audiovisuels si celles-ci représentent des plans) [Aigrain *et al.*, 1995a, Gupta *et al.*, 1997, Foote *et al.*, 1998], mais aussi dans des bases de documents sonores où il s'agit de passer d'une image (d'un son) à une image (à un son) qui lui ressemble dans un espace en général multidimensionnel [Blackburn et DeRoure, 1998].

Les *approches mixtes* associent plus raisonnablement une recherche de haut-niveau et de bas-niveau, c'est à dire des mots-clé et des primitives image, par exemple [Kuo et Chen, 1996, Foutain et Tan, 1998]. [Bolle *et al.*, 1998] affirme qu'une recherche dans une base de documents audiovisuels commence par une phase de navigation et de recherche dans des descripteurs interprétés, avant éventuellement de mener une recherche plus fine faisant appel à des primitives image, une fois que l'espace de recherche est déjà relativement réduit. Une recherche dans la base WebSEEK [Smith et Chang, 1997] passe d'abord par une réduction de l'espace de recherche à l'aide d'un terme choisi dans une hiérarchie (par exemple *transport*, *avion*), avant de chercher une image par similarité, avec la possibilité d'utiliser du bouclage de pertinence. L'apprentissage de critères de similarité adaptés aux utilisateurs permet dans certains cas d'améliorer les performances des systèmes [Nastar *et al.*, 1998, Heinrichs *et al.*, 1999].

18. <http://www.vdp.fr/>.

19. Plus généralement, plusieurs produits destinés aux professionnels existent sur le marché, on en trouvera un état de l'offre sur <http://www.footage.net/directory/dbsystems/>. Le site propose également de rechercher des documents dans un réseau de plusieurs archives de documents audiovisuels.

20. Cela pourrait accréditer la thèse de la nécessité d'une phase de pré-recherche textuelle dans la base avant de s'intéresser à des composantes images du document.

3.4.3 Présentation de documents

La présentation de documents audiovisuels à l'utilisateur d'un système est intégrée au système audiovisuel et à l'interface elle-même. Nous nous interrogerons tout d'abord sur les liens entre présentation et recherche, avant de présenter quelques méthodes de présentation actuelles.

La présentation devenant partie intégrante du SIAV, celle-ci est fortement liée à la représentation des données et la recherche, que celle-ci se fasse par requêtes directes, par l'exemple ou par navigation.

Recherche et présentation

La recherche par l'exemple peut se faire sur un certain nombre de documents tels que : images, dessins ou schémas, parties de vidéo, morceaux de bande-audio, morceaux de documents audiovisuels.

La recherche par navigation consiste à passer d'un document à un autre dans la base de documents, mais aussi à l'intérieur d'un même document audiovisuel. Les types de lien considérés peuvent alors varier dans les limites imposées par la représentation choisie. Les liens peuvent être mis en place manuellement avant la présentation, ou bien calculés. Il est encore possible de naviguer en calculant au besoin le ou les documents cibles (ou le morceau de document cible si l'on navigue dans le document), ce qui peut revenir à la mise en place transparente ou non d'une nouvelle requête, *etc.*

La granularité joue également un rôle : quel est le niveau de granularité récupéré en retour de requête ? Cela peut aller d'une image au document entier, en passant par toutes les variations possibles (plans, scènes, strates quelconques). Il va de soi que la représentation choisie joue encore à ce niveau. Une requête à composante structurelle peut permettre de spécifier ce que l'on attend, tandis qu'une simple requête par termes se limitera aux parties de document qui y répondent. Simonnot [1996] met en place un système qui choisit en fonction du type de requête (exploratoire, précis, *etc.*) le type de granularité de retour qui y correspond.

Une fois fixé le niveau de granularité, encore faut-il décider de la manière dont le résultat de la requête va être présenté. Le plus simple consiste bien entendu à jouer le morceau de vidéo récupéré, mais il est possible et même souhaitable d'utiliser une présentation plus évoluée prenant en compte le média temporel qu'est l'audiovisuel, ainsi que les limitations inhérentes à un écran d'ordinateur. On peut donc par exemple souhaiter présenter l'ensemble d'une séquence audiovisuelle en un écran, avec une appréhension immédiate de son contenu, ce qui ne peut que passer par l'utilisation de méthodes permettant de projeter la composante temporelle autrement que sur l'axe temporel « réel ».

En résumé, une requête rend un certain nombre de documents ou de morceaux de documents à un certain niveau de granularité. Ces documents sont présentés à l'utilisateur selon un niveau de granularité qui peut être différent, et avec certaines méthodes de présentation : toutes les possibilités sont loin d'avoir été explorées.

Temporalité et présentation

Le caractère temporel de la vidéo pose évidemment problème quand on veut dépasser la visualisation simple de documents audiovisuels. Si une requête retourne une heure de documents audiovisuels, il faudra une heure pour les visualiser afin de découvrir les quelques secondes ou minutes d'intérêt, ce qu'on ne peut se permettre *a priori* pas plus qu'on ne visualisera tout un film à cet effet.

Nous avons vu qu'il était souhaitable de disposer de possibilités plus étendues de présentation, permettant l'appréhension du contenu d'un morceau de document de façon plus rapide que le simple défilement²¹. Il s'agit en conséquence de réduire artificiellement la dimension temporelle du document, mais aussi, et afin de ne pas perdre ce caractère fondamental de présenter explicitement celle-ci. Pour dépasser cette temporalité sans la gommer, l'approche structurelle explicite est nécessaire.

Considérant qu'il y a lieu de séparer document audiovisuel dans sa globalité, et parties de celui-ci, deux grandes catégories de problèmes se posent généralement :

- comment présenter de façon temporellement réduite, voire immédiate un morceau de document audiovisuel ?
- comment donner accès à la structure du document dans sa globalité, et/ou — en lien avec ce qui précède — organiser les présentations de parties afin de présenter l'ensemble du document ?

Dans les parties suivantes, nous essayerons de voir quelles réponses à ces questions ont pu être proposées.

Présenter une partie de document audiovisuel

La partie dont il est question ici se rapporte le plus souvent à un plan dont les frontières ont été détectées automatiquement, l'enjeu est alors de présenter de la façon la plus synthétique possible le contenu du plan, notamment avec les objets les plus marquants. On peut cependant décider de présenter par exemple une minute de vidéo, sans s'attacher à la succession des plans.

La présentation d'un plan de façon synthétique doit souvent passer par des connaissances du domaine à même de favoriser la construction d'*icônes vidéo*, qui se définissent pour comme des images physiques associées à des informations complémentaires (durée, titre, annotation, indices visuels) résumant le contenu d'une suite d'images. L'objectif est de permettre de se rendre compte d'un simple coup d'œil du contenu du plan considéré.

La méthode la plus simple consiste bien entendu à extraire une image clé de l'ensemble des images dont la partie considérée est composée²². A ce niveau se posent déjà un grand nombre de questions, concernant l'image particulière à extraire. Plusieurs propositions ont été faites : certains auteurs prennent systématiquement la première (ou une des premières) image(s) du plan, d'autres préfèrent extraire une image du milieu. D'autres encore, se basant sur les mouvements de caméra détectés préconisent d'en tenir compte : par exemple si le mouvement est un zoom avant, on prendra la première image. Mais en cas de travelling horizontal, une seule image peut ne pas être suffisante : il y aurait alors lieu d'en prendre plusieurs... Ajoutons qu'une détection automatique de certaines caractéristiques peut permettre de choisir la ou les bonnes images, par exemple si un gros objet y bouge (caractéristiques de couleur) et qu'il y a possibilité de définir des densités d'images-clé par plan, par minutes, *etc.*

Afin de rendre compte des divers mouvements (caméra et objets), l'ajout de renseignements à l'image représentative choisie se révèle utile. On ajoutera ainsi des éléments indiquant la durée du plan présenté, ou bien les mouvements de caméra.

21. Même accéléré ou ralenti, le défilement devient rapidement incompréhensible, au niveau de l'image, mais aussi, et de façon encore plus critique, au niveau du son.

22. Méthode qui peut d'ailleurs servir également à l'indexation, comme nous l'avons vu plus haut.

Dans le projet IMPACT, [Ueda *et al.*, 1993] définissent des *Micons* (*moving icons*) qui sont des présentations de plan composées d'une image clé dont la profondeur (en 3D) indique la durée, et sur laquelle des flèches (traces) indiquent les mouvements. Les *images représentatives* [Arman *et al.*, 1994] mêlent également image clé, indications de mouvement et de durée. [Cherfaoui et Bertin, 1994] proposent de construire une représentation synthétique d'une portion de vidéo (une portion est un morceau de vidéo où n'apparaît qu'un seul mouvement de caméra), qui comprendra notamment les objets contenus dans la portion, mais aussi les mouvements. Pour cela, il s'agit de choisir en fonction des mouvements de une à trois images représentatives (par exemple, pour un travelling horizontal, on choisira les première et dernière images, ainsi qu'une image intermédiaire), tandis que les mouvements d'objets seront décrits par des traces. Ceci reste de l'ordre de la proposition cependant, mais procède d'un domaine encore peu exploré qui est celui de la *schématisation* d'images ou de séquences vidéo.

La simple extraction d'une image-clé et l'ajout de traces ne suffisant généralement pas, il est possible de *synthétiser* une image décrivant le contenu de façon globale. Ainsi les *salient stills* (« instantanés pertinents ») développés au MIT [Massey et Bender, 1996] se donnent-ils pour objectif de représenter en une image le contenu d'un morceau de vidéo de façon à préserver le contexte spatial et temporel du plan tout en en représentant les objets les plus marquants. L'objectif avoué des auteurs est de faire des *salient stills* de véritables images ayant un contenu cinématique marqué. Le processus de synthèse passe par une phase de détection du flot optique (mouvement), suivi d'une opération de superposition d'images extraites du plan, avant un filtrage des objets non marquants en fonction de leur durée d'apparition à l'écran. Le résultat permet d'associer sur une image (dont le cadre n'existe souvent plus du fait des mouvements de caméra) un fond mouvant et des objets mouvants²³. A noter que les *VideoSpaceIcons* [Tonomura *et al.*, 1993] sont des images du même type. [Irani *et al.*, 1996] améliorent les *salient-stills* en *mosaic images*, et les *panorama excerpts* de [Taniguchi *et al.*, 1997] relèvent de principes similaires.

Un autre exemple de construction d'image représentative d'un morceau de vidéo est possible. [Yeung et Yeo, 1997] mettent en place des *posters* qui sont des résumés en images. Les auteurs partent du principe que la présentation vidéo doit être compacte, intuitive et facile à comprendre et que les éléments les plus importants (personnages principaux, actions) doivent être représentés. Ils réservent alors dans un écran un ensemble de zones, et leur attribuent des images extraites des plans de la séquence avec une taille dépendante de leur importance dans celle-ci. Un plan considéré comme très important fournira par exemple une image du poster deux fois plus grosse que les images d'autres plans, jugés moins représentatifs de la séquence.

Alors que toutes les méthodes jusqu'ici présentées négligent la composante sonore des documents audiovisuels (au moins au niveau de la présentation), [Taniguchi *et al.*, 1995], décrivent un plan à l'aide de *mini-videos*, composées d'une image-clé extraite et de la bande son associée au plan en question, considérant que dans un objectif de navigation, il n'est pas essentiel de visionner toutes les images d'un plan.

Nous terminerons cette partie par quelques remarques : une première question concerne le moment du calcul des descripteurs visuels, qui peut se passer pendant la phase d'indexation, auquel cas ceux-ci devront être stockés avec les documents audiovisuels dans la base de documents, ou bien à la volée alors qu'un document ou morceau de document a été sélectionné. [Tonomura, 1997] montre qu'il est — au moins actuellement — nécessaire que les descripteurs soient calculés *avant* d'être utilisés dans l'interface pour des contraintes de puissance de calculs.

23. Exemples disponibles sur <http://nif.www.media.mit.edu/stills> ou encore <http://mmassey.www.media.mit.edu/people/mmassey/SSindex.html>.

Une deuxième remarque concerne la différence entre l'image et le son. Alors qu'un ensemble d'images animées peut être présenté de façon non linéaire (par exemple par juxtaposition, ou par la synthèse de salient stills), et surtout perçu comme tel par l'observateur (comme présentation non linéaire d'un contenu qui l'était à la base), il n'est pas, ou très peu possible d'exécuter une même opération de délinéarisation avec le son, peut-être tout simplement parce que celui-ci ne se perçoit pas de manière spatiale *et* quasi-immédiate. Un changement de modalité est cependant possible, par exemple pour la parole, qui permet de passer au texte.

Présenter un document audiovisuel

Après avoir parlé de quelques méthodes de représentation du contenu d'un plan de façon imagée à l'aide d'icônes plus ou moins construites, nous devons nous intéresser à la présentation du contenu d'un document audiovisuel dans son ensemble, c'est-à-dire à la prise en compte de données de structuration.

La présentation de la structure d'un document est destinée à en donner une vue synthétique, mais est également liée à la *navigation* au moins intra-document qui peut s'ensuivre. L'objectif d'un tel système est de permettre d'accéder de façon rapide au contenu intéressant (en supposant que c'est tout le document, ou au moins une partie signifiante de celui-ci qui est retournée). La structuration choisie y joue donc un rôle important.

[Aigrain *et al.*, 1996] considèrent que l'on peut distinguer trois modèles principaux pour la navigation interactive :

- la navigation *hiérarchique* dans un document segmenté, et dont chacun des segments est représenté par au moins une image représentative ;
- la navigation le long de *lignes de temps* ou de *strates* ;
- la navigation dans des *graphes* orientés dont les nœuds sont des ensembles de plans (navigation par séquences donc).

Nous constatons pour notre part que les différentes structurations de documents, c'est-à-dire *implicites*, *hiérarchique* ou *en graphes* que nous avons présentées plus haut nous permettent de classer tout à la fois les méthodes de navigation et les modèles de présentation globale en deux catégories, les documents à structure implicite ou explicite.

Présentation de documents à structure implicite

Cette catégorie de présentation s'adresse aux documents dont la représentation a été basée sur une structuration implicite (modèles de stratification notamment), qui rappelons-le découle simplement de l'inscription temporelle d'un morceau de vidéo dans le document lui-même.

Cette présentation s'effectue le long d'une ligne de temps horizontale, orientée de gauche à droite, et au niveau de laquelle on peut trouver la présentation iconique et les annotations attachées à une strate (c'est-à-dire entre deux instants matérialisés sur la ligne de temps). On notera tout de suite que la limitation de la taille de l'écran nécessitera le plus souvent qu'on considère plusieurs lignes de temps et qu'on adopte une approche soit multi-perspectives (plusieurs types d'indication sur plusieurs lignes), soit multi-échelles (plusieurs lignes de temps d'échelles variables), soit les deux ensemble (voir figure 3.6).

Certains systèmes se contentent de présenter un ensemble de descripteurs classés par ordre chronologique. Par exemple un ensemble de descripteurs visuels de plans (R-frames, images-clé,

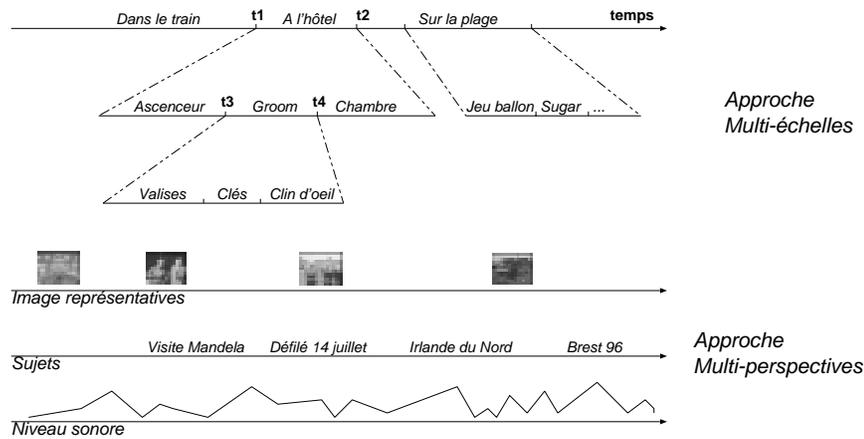


FIG. 3.6 – Approches multi-échelles et multi-perspectives

mini-videos), classés dans un tableau se lisant de gauche à droite et de haut en bas, avec la possibilité de « zoomer » sur une partie représentée par un descripteur, qui donne un autre tableau dont la couverture temporelle est moindre, *etc.* [Yeung et Yeo, 1997] proposent de présenter un *résumé pictural* d'un document comme une collection de posters arrangés chronologiquement. [Brown *et al.*, 1995] range les « événements » du document sur une ligne et leur attribue une importance (en taille) en fonction de leur degré de pertinence en regard de la requête posée. [Davis, 1993] présentent plusieurs lignes de temps simultanément, annotées par les icônes de MediaStream, mais aussi des *videograms* (sortes de salient-stills), ou l'amplitude du son (des barres verticales indiquant une pause)²⁴. L'interface Mediascope de l'Inathèque permet l'annotation et la visualisation d'annotation de documents le long d'une ligne de temps.

Le principe des lignes de temps est dynamique (on peut faire défiler le temps), et aussi bien multi-échelles que multi-perspectives. Le contenu du document est visible facilement avec la précision voulue. Cependant, la structure logique du document ne reste qu'implicite, c'est-à-dire centrée sur les relations temporelles basiques, à charge pour l'utilisateur de la reconstruire.

Présentation de document à structure explicite

Quand la représentation du document est explicitement structurée, il est intéressant d'utiliser les propriétés de cette description.

Le système VideoScribe d'annotations de plans de l'INA gère la structure par images-clé, structure qui a été explicitée par l'utilisateur sous la forme de regroupements de plans et de choix de nouvelles images-clé pour les regroupements effectués, jusqu'à l'image — toute théorique — présentant tout le document. Il y a alors possibilité de se déplacer dans l'arborescence mise en place.

[Yeung et Yeo, 1996] proposent de naviguer dans le graphe « historique » des scènes qui a été mis en place de façon automatique, et dont on a une illustration figure 3.7. Il est cependant remarqué dans [Aigrain *et al.*, 1996] que cette représentation du document en graphe amène l'utilisateur à perdre une linéarité temporelle qui était par exemple bien représentée dans les systèmes à ligne de temps.

²⁴. Pour une actualisation de ce prototype et quelques copies d'écran, voir <http://www.interval.com/projects/mediast/index.html>.

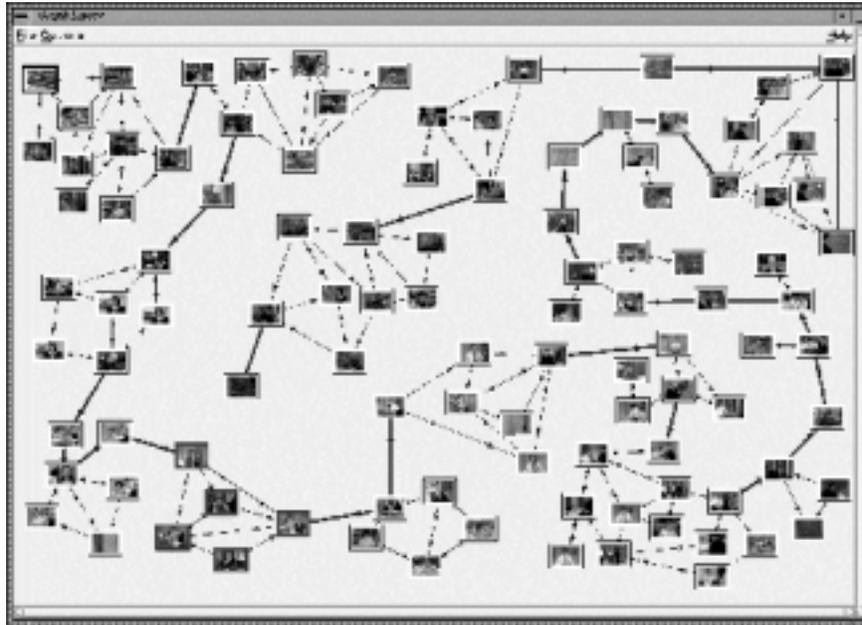


FIG. 3.7 – Graphe de transition de scènes

Dans les modèles de représentation à la fois structurels et conceptuels, les possibilités de navigation et de présentation peuvent être fournies.

Par exemple dans le système VideoStar [Hjelsvold *et al.*, 1995b, Hjelsvold *et al.*, 1995a] l'utilisateur a devant lui un véritable sommaire (au sens textuel du terme) du document, découpé en parties et sous-parties (séquences et plans vidéos), avec pour chaque entrée un descripteur textuel issu de l'annotation et une indication de durée, tandis qu'une fenêtre rappelle le contenu du document en terme d'événements marquants. On a alors accès à des possibilités de navigation aussi bien conceptuelle (par le contenu d'annotation) que structurelle.

Alors que le système VideoStar présente une abstraction du modèle de données de l'application, l'approche adoptée par [Corridoni *et al.*, 1996] est autre : il s'agit ici de donner un accès multi-caractéristique au contenu du film, tout en rendant le modèle de composition explicite. Le schéma global de représentation du document sous forme de graphe (*cf.* figure 3.4) est alors présenté directement à l'utilisateur, qui procède à la recherche par une méthode de *filtrage de données* en spécifiant une partie de graphe correspondant à une *perspective* sur le film et en spécifiant quels sont les descripteurs qui l'intéressent suivant cette perspective. Le système mène alors une recherche sur les caractéristiques liées aux éléments structurels considérés comme pertinents²⁵. [Zettsu *et al.*, 1997] choisissent également de présenter directement le graphe d'annotation à l'utilisateur.

25. Plus précisément on définit un schéma conceptuel, avec des classes, des attributs, des relations de propriété et d'héritage, puis de ce modèle abstrait, on procède à un appariement sur un graphe orienté aux relations nommées, qui est la structure de données. On exploite ensuite les propriétés des graphes pour définir des outils de manipulation, qui permettront les opérations de filtrage de données. La perspective est un sous-graphe de mêmes propriétés, le filtre est un ensemble de conditions sur les attributs d'objets, la sélection étant l'application de ce filtre par le biais de la perspective.

Conclusion sur la présentation

Tous les systèmes permettent de visionner de la vidéo de façon classique, séquentielle et linéaire, mais également d'avoir une vue la plus globale possible du document afin qu'il soit possible de se faire une idée de son contenu d'un coup d'œil. Cela passe par la mise en place d'une part de méthodes plus ou moins élaborées²⁶ de présentation du contenu de morceaux de document, ainsi que de présentation du document dans son ensemble en organisant les descripteurs audiovisuels précédents. Les possibilités de navigation offertes permettent de se déplacer dans le document (navigation *intra-document*), et la possibilité de navigation *inter-document* (*i.e* entre parties de documents) n'a pas ou peu été étudiée. La structuration des représentations de documents audiovisuels permet d'organiser une navigation autre que temporelle ; on retiendra en fait que tous les niveaux déjà vus d'un SIAV sont liés : requêtes, présentation, navigation dépendent essentiellement du choix du modèle de représentation.

3.4.4 Génération de documents

La génération de documents audiovisuels dans des SIAV peut avoir plusieurs objectifs, selon que l'on veut créer un nouveau document pour répondre à une requête (génération de réponse) ou pour le considérer comme document diffusable et valable *per se*.

Dans tout les cas, il s'agit de mettre en place :

- des connaissances sur les parties de documents qui concernent à la fois leur contenu de filmage (mouvements de caméra, types de document) et leur contenu « factuel » (que se passe-t-il, où cela se passe-t-il, *etc.*), donc des annotations ;
- des connaissances sur les contraintes et les règles de filmage, sur la structure d'un document, bref sur la manière d'arranger des parties de documents ensemble pour former un tout cohérent.

[Sack et Davis, 1994] proposent de générer des séquences vidéo à partir de strates annotées. Les connaissances de structuration sont des « plans d'histoires », et une version de GPS (General Problem Solver) est utilisée pour la planification. Cela représente une première tentative de représenter ce que les auteurs appellent « sens commun » télévisuel, c'est à dire les contraintes dûes au média. [Carlos et Uehara, 1998] considèrent une même histoire²⁷ vue suivant plusieurs lignes de temps selon les protagonistes, un service de résumé automatique grossier est alors mis en place à l'aide de règles (unité partagées, causes-effets, scènes dominantes). [Lienhart *et al.*, 1997] cherchent quant à eux à mettre en place des extraits de films (d'action) permettant de servir de représentants à ceux-ci. Ils repèrent les parties de dialogues, d'explosions et les textes, et éditent les morceaux en utilisant des règles basiques calquées sur les bandes-annonces de films d'action américain.

De façon plus systématique, [Nack et Parkes, 1997] étudient soigneusement la manière dont les documents audiovisuels sont produits et réalisés et explore un certain nombre de théories filmiques. Les parties de documents disponibles sont alors décrites suivant des critères liés à la description du contenu image, tandis que des règles de haut-niveau et des modèles d'histoire recherchent les parties de documents nécessaire à la génération de scénettes drôles (*c'est en effet*

26. Il faut également tenir compte des capacités humaines d'attention, [Tonomura, 1997] décrit des interfaces de présentation vidéo en prenant en compte cet aspect.

27. Les histoires enfantines japonaises — en dessins animés — inspirent les chercheurs, serait-ce parce que leur trame est facilement représentable et étudiable? Ceci peut être mis en lien avec les analyses de Propp sur les contes russes.

au thème de l'humour que s'attaquent les auteurs). [Ahanger et Little, 1998] étudient et annotent à l'aide de mots-clé les journaux télévisés des chaînes CNN et ABS, et en déduisent des règles de composition de reportages. Ils utilisent alors ces règles pour réutiliser les parties de reportages qu'ils ont annotées dans des nouveaux documents. Les documents générés sont alors comparés aux anciens à l'aide de métriques (par exemple continuité thématique ou structurelle, progression du contenu) portant sur les annotations, et les résultats mettent en évidence une qualité comparable entre reportages, validant ainsi les règles de composition mises en place. Curieusement, les auteurs ne valident pas leurs résultats par des tests devant de simples téléspectateurs, et il est difficile de juger de la pertinence des séquences générées.

Dans une étude plus récente, [Nack et Steinmetz, 1998] proposent non plus de générer automatiquement des documents, mais bien de fournir un *environnement* de production numériques adapté aux contraintes télévisuelles, afin de *documenter* la création d'un document, de sa pré-production à sa diffusion, en passant par son tournage numérique. Le système gère alors sous un format²⁸ commun l'ensemble des connaissances utilisées par les professionnels du montage.

[Nack et Steinmetz, 1998] montrent également qu'alors que les progrès technologiques devraient conduire à une production numérique de DAV la plus automatisée possible, la situation n'est pas telle. Il constate alors un manque certain dans le domaine. Nous ne partageons pas son inquiétude, au sens où par exemple dans le domaine textuel, les systèmes de génération de textes ne sont pas exactement au point, alors que les textes et l'informatique cohabitent depuis des années. La génération automatique de documents audiovisuels utiles passe par une formalisation d'un système fonctionnel minimal de l'audiovisuel, qui permette de décrire les règles non écrites dûes au contraintes du média, ainsi que les structures audiovisuelles en découlant. Tant que des langages de description de ces contraintes n'existeront pas, et que les études du média et des règles de production n'auront pas été menées, la génération automatique n'a que peu de chances de fonctionner. Il est également vraisemblable que c'est du domaine des jeux vidéos qu'un ensemble de contraintes devraient provenir et permettre de générer des documents audiovisuels synthétiques.

3.5 Conclusion

Nous avons dans ce chapitre présenté ce qu'était la représentation audiovisuelle dans la littérature, en l'étudiant sous l'angle de l'annotation. Après une rapide présentation des diverses communautés de recherche impliquées dans la mise en place de systèmes d'information audiovisuelle, nous avons tout d'abord étudié quelles étaient les caractéristiques de description disponibles, à savoir les primitives images ou son, les caractéristiques calculées et interprétables, et enfin les caractéristiques interprétées. Nous nous sommes ensuite intéressés à la structuration de ces caractéristiques d'une part dans des annotations complexes, d'autres part en étudiant la structuration des documents. Nous avons ensuite présenté rapidement quelques fonctionnalités de SIAV existant, de la recherche textuelle à la génération de documents audiovisuels, en passant par leur présentation à l'utilisateur.

Nous ne concluons pas dans ce chapitre sur la représentation audiovisuelle, car une discussion y est consacrée dans le chapitre suivant. L'objectif du prochain chapitre est de définir au regard

28. Ce format devrait avoir à faire à MPEG7 et à XML (voir section 9.2.4). On notera que [Lindley et Vercoustre, 1998] propose de décrire des prescriptions de mise en place de documents audiovisuels dans un format lié à SGML, tandis que [Auffret et Bachimont, 1999] compte utiliser XSL pour générer de nouveaux documents à partir de descriptions de documents en XML : la prise en compte de description de documents audiovisuels comme des documents structurés décrits dans des langages de balises progresse donc.

des enseignements des trois premiers chapitres quelles sont les nécessités de modélisation de documents audiovisuels pour un système d'information audiovisuelle idéal.

Chapitre 4

Réflexions sur la modélisation audiovisuelle

Sommaire

4.1 Représenter pour exploiter	91
4.2 Sur les connaissances d'exploitation	93
4.2.1 L'image et le symbole	93
4.2.2 L'organisation des connaissances décrites: la structure	95
4.2.3 Organiser les connaissances de description	98
4.3 La notion de contexte dans l'audiovisuel et sa nécessaire prise en compte	99
4.3.1 Contexte et systèmes d'information	99
4.3.2 Contexte et audiovisuel	101
4.4 Description et exploitation en contexte	103
4.4.1 Description de documents	103
4.4.2 Ecriture et lecture en contexte	104
4.5 Nécessités pour la modélisation audiovisuelle	105

L'objectif de cette partie est de définir quelques nécessités concernant la modélisation audiovisuelle, nécessités auxquelles le modèle de représentation audiovisuelle que nous définissons dans la deuxième partie de ce mémoire devra répondre. Pour cela, nous revenons tout d'abord sur la notion de représentation, que nous considérons comme centrale pour toute exploitation numérique de documents audiovisuels. Nous étudions ensuite les connaissances d'exploitation d'un système d'information audiovisuelle, en revenant sur les caractéristiques de modélisation, et sur la structuration de ces caractéristiques. La section suivante est consacrée à la notion de contexte, dans les systèmes d'information mais aussi dans les documents audiovisuels. Nous proposons enfin une approche de l'indexation et de l'exploitation de documents comme *écriture* pour la description. Pour conclure, nous présentons les points importants qu'un modèle de représentation doit prendre en compte.

4.1 Représenter pour exploiter

Comme nous l'avons déjà argumenté plus tôt, c'est à partir de sa représentation en machine qu'il est possible d'exploiter, c'est à dire d'utiliser un document audiovisuel. La représentation

minimale du document est une suite d'images, une bande son, munies des connaissances permettant de le « jouer ». Le codage du document est donc déjà un modèle minimal de représentation (par exemple MPEG2).

Avec la mise en place d'une indexation, le document audiovisuel est décrit par un ou plusieurs autres documents, composés de l'ensemble de ses index et de la structuration de ceux-ci.

La recherche d'un document entier ou d'une partie de celui-ci repose sur une recherche dans les documents d'indexation. Une recherche directe dans un grand fond supposerait en effet un traitement automatique de tous les documents à chaque requête pour vérifier si ceux-ci y répondent en terme d'arrangements de caractéristiques calculables, ce qui est à écarter d'emblée.

La navigation autre que temporelle repose également sur les index du document de description. Une interface de navigation dépassant les simples capacités de lecture d'un magnétoscope numérique (lecture, stop, pause, retour et avance rapide, accès direct) permettra par exemple une navigation plan par plan, ou bien une navigation « conceptuelle » (aller directement à la prochaine apparition du personnage principal, ou bien à la résolution de l'énigme).

L'édition d'un nouveau document, son montage, prennent appui sur l'indexation de documents sources en décrivant les parties intéressantes pour la tâche en cours. Par exemple, un système peut rechercher un plan avec Jacques Chirac et Nelson Mandela, vus de dos, qu'il enchaînera avec un plan quelconque de ce dernier en noir et blanc.

Une lecture « active » de document doit permettre à un utilisateur de l'annoter lui-même, c'est à dire d'ajouter librement ses propres données d'indexation au document. L'indexation du document par un documentaliste professionnel relève de la même démarche, à la différence que la politique de gestion du fond documentaire impose une certaine manière de décrire.

Le modèle de représentation de document choisi peut également être d'importance au niveau de l'organisation du stockage des documents numériques. Par exemple une émission de divertissement décrite séquence par séquence peut tirer avantage d'une organisation physique des données sur les mémoires basée sur ce découpage, s'il s'avère que le spectateur a tendance à sauter une chanson qui ne lui plaît pas.

Enfin, il est évident que toute aide à l'utilisateur dans ses tâches d'exploitation d'un système d'information audiovisuelle passera par une modélisation des possibilités de manipulation de documents offertes par l'application, lesquelles s'appuieront sur le modèle de représentation de l'indexation des documents.

L'enjeu premier de la mise en place de formats de représentation et d'indexation de document audiovisuel est donc d'importance, car si dans un premier temps la documentation (l'indexation) des documents audiovisuels sera totalement extérieure à ceux-ci, à plus long terme, il n'est pas exclu d'imaginer que certaines parties de celle-ci seront *intégrées* aux futurs formats audiovisuels. Le document contiendra alors les données « brutes », mais aussi des connaissances permettant une exploitation autre que la navigation basée sur la temporalité du médium.

Selon l'angle d'analyse que nous avons choisi, c'est l'annotation, c'est à dire l'alignement de connaissances sur des parties de documents audiovisuels que nous devons étudier. Il s'agit donc en premier lieu de s'interroger à nouveau sur les *caractéristiques* d'annotation, sur la *structuration* de celles-ci dans les documents et sur leur *organisation* dans des bases de descripteurs.

4.2 Sur les connaissances d'exploitation

Nous nous intéressons dans cette partie aux connaissances d'exploitation, c'est à dire au *résultat* de l'annotation, mais aussi aux connaissances *permettant* la mise en place des connaissances d'exploitation.

4.2.1 L'image et le symbole

Nous avons distingué, suivant leur mode de mise en place, trois types de caractéristiques d'annotation.

Primitives. Les *primitives* sont des résultats de calculs menés sur l'image, le son ou la suite d'images et fournissent des résultats *non interprétables*¹ si ce n'est en connaissance de la sémantique du calcul. Un histogramme de couleur ou une texture en sont des exemples canoniques. La vocation des primitives est d'être des représentants de parties du flux *pré-calculées*, sur lesquelles il sera possible de mener des comparaisons de similarité sans faire à nouveau appel aux données brutes.

Une conséquence est qu'il est quasi-impossible de mener une recherche sur des primitives en en spécifiant un arrangement *a priori*. Cela reviendrait en effet pour l'utilisateur à extraire/calculer lui-même des primitives équivalentes d'une « image mentale » représentant ce qu'il cherche (image, son, mouvement), avec les approximations qu'on imagine. Mais ce problème peut être résolu avec l'utilisation d'un *exemple* fourni par l'utilisateur et censé exprimer ce qu'il cherche, duquel la machine extrait elle-même les primitives². Il n'y a alors plus qu'à espérer que la signification de la mesure de similarité-machine (qui dérive, mais n'est pas seulement fonction de la ou des primitives extraites) corresponde à quelque chose qui ait du sens pour l'utilisateur, et — mieux — que ce sens soit celui qui était attendu.

L'extraction de primitives pour la similarité se révèle donc un thème extrêmement délicat, car basée sur la supposition qu'une image signifie par elle-même, et qu'il est possible d'extraire le même sens qu'un humain d'un tableau de couleurs de pixels en général. Nous avons déjà vu que pour l'être humain le sens n'était pas intrinsèque à l'image : une image doit venir avec des interprétants pour être interprétable *justement*³. De la même manière, une image a un genre, est prise selon des canons esthétiques et est lue comme telle (en d'autres termes, le fait de partager une représentation sous la forme d'un tableau de valeurs de couleurs de pixels ne signifie pas que toutes les images sont perçues de la même manière, alors que pour la machine, si). Au niveau de l'audiovisuel, ce phénomène existe également : une émission de variétés actuelle n'est pas une série policière des années soixante-dix, et ne devrait pas toujours être traitée automatiquement

1. Et surtout non interprétés.

2. Passons sur deux problèmes importants. D'une part, exhiber une image comme requête signifie simplement l'espérance que la machine va elle-même y reconnaître les traits pertinents qu'on veut y trouver : on ne spécifie pas ce que l'on veut. Pour pallier ce problème, certains auteurs proposent que l'utilisateur *dessine* lui-même ce qu'il cherche, c'est à dire « écrive sa vision ». Deux difficultés se posent encore à ce niveau. En premier lieu, tout le monde n'est pas capable de s'exprimer par le dessin. En deuxième lieu, la mise en place d'une *icône* obéit à des règles fortement culturalisées, ne participe pas de la même modalité sémiotique que l'image filmée [Vaillant, 1997]. Il n'y a donc *a priori* aucune raison pour qu'un traitement sur une icône donne des résultats exploitables par exemple sur une photographie. Pour le son, la question est sans doute plus nuancée, au sens où l'expression est « plus directe » dans le matériau sonore, est plus phénoménologiquement située. Les erreurs de rythme ou les fausses notes n'en existent pas moins.

3. Ces interprétants peuvent également être partagés culturellement : une photographie noir et blanc jaunie ne sera pas *a priori* considérée comme ayant été prise la veille. La perception de son contenu sera alors médiée par la supposition que les objets et les personnages représentés sont anciens.

de la même manière. Les calculs universels n'existent donc pas, pas plus que les vues d'une image ne sont universelles, mais *partagées*.

Est-ce à dire que tous les travaux sur les primitives sont voués à l'échec en termes de similarité, donc de recherche? Auquel cas faudrait-il renoncer à les utiliser comme connaissances d'indexation et de description?

Sans doute si l'on considère une similarité uniquement liée à la machine, et d'une part indépendante de la tâche de l'utilisateur qui l'utilise, d'autre part indépendante de cet utilisateur.

Au contraire, il nous paraît en premier lieu nécessaire que les primitives (et la similarité adjointe) soient *expliquées* à l'utilisateur, c'est à dire que celui-ci soit en mesure de comprendre le mode de fonctionnement de la similarité pour pouvoir l'utiliser. Ce mode de fonctionnement peut bien sûr être appris au fil des utilisations, par expérimentation, mais on gagnera du temps et de l'exactitude en l'expliquant⁴. La nécessité de telles explications risque de faire disparaître les primitives qui n'ont d'intérêt que pour le scientifique au profit de primitives signifiant quelque-chose pour l'utilisateur. Dans le même temps, il va de soi que ce sont de primitives basées sur des éléments ayant une pertinence pour l'homme qui seront utiles. L'extraction de celles-ci se basera alors par exemple sur des « bonnes formes » au sens de la Gestalttheorie⁵.

En deuxième lieu, il est nécessaire de prendre en compte des connaissances *extérieures* au document. Connaissances sur le genre de document tout d'abord: par exemple, il peut être très facile de distinguer des plans de plateau dans un journal télévisé, et de les rechercher par similarité, car les arrangements de couleurs sont similaires. Connaissances sur l'utilisateur et sa tâche ensuite: reconnaître une zone potentiellement cancéreuse sur une image issue d'un scanner est une tâche qui est fortement contrainte par les connaissances médicales, le type d'image, *etc.* Les systèmes mis en place en informatique médicale se basent alors sur des modèles extrêmement sophistiqués et complexes. On conçoit que pour l'audiovisuel, la question ne puisse être simple, et doive disposer de réponses certes les plus génériques possibles, mais sans se faire d'illusion sur *la* réponse, la primitive et la similarité ultimes. Notons enfin qu'une tendance intéressante dans les systèmes utilisant des calculs de similarité pour retrouver des documents est de compléter les descriptions par des primitives à l'aide de concepts de beaucoup plus haut niveau comme des mots-clé, par exemple [Kuo et Chen, 1996].

Caractéristiques calculées interprétables. En fait, dès que se pose la question de la tâche (et non plus du calcul d'une primitive résolvant tous les problèmes) se pose la question de l'interprétabilité par l'utilisateur de la caractéristique: dans le cadre d'un tâche donnée, l'utilisateur va avoir des besoins précis, qui correspondent à la manipulation de caractéristiques de description précises.

Il s'agit bien alors, guidé par les besoins des utilisateurs, de mettre en place des caractéristiques calculées interprétables par l'être humain, et plus, verbalisables, descriptibles en tant que concepts. La similarité dans ces conditions devient similarité conceptuelle. Par exemple, si la machine calcule qu'une forme ronde se déplace de gauche à droite dans un plan, il est possible d'exprimer ceci par des symboles, par exemple `Forme; Type=rond; Déplacement: mode=linéaire, direction=horizontal, sens=gauche/droite`. De la même manière, la détection d'un plan conduit à une annotation `Plan`.

4. D'autre part, il ne nous semble pas pertinent de supposer qu'il est possible d'apprendre une combinaison (par exemple pondérée) de mesures de similarités d'un utilisateur, c'est à dire de supposer *a priori* que les primitives ininterprétables extraites pourraient former par combinaison des primitives interprétables.

5. Théorie de la perception développée au début du XXème siècle en Allemagne, qui a par exemple mis en évidence le fait que l'humain repère plus spontanément certaines formes géométriques simples.

Comme nous l'avons déjà souligné, il convient là encore que l'utilisateur qui utilise ce genre de caractéristiques soit conscient de leur caractère calculé, et surtout des limites de validité du calcul mis en place. Cependant, le fait que le résultat soit du niveau interprétable est crucial, puisqu'il permet de jeter un pont entre la machine et l'être humain, et ouvre la voie à une collaboration fructueuse. L'utilisation intensive de connaissances extérieures au signal, de modèles de documents audiovisuels notamment fait que le calcul va plutôt s'adapter à des modèles de description humains, que l'humain s'adapter à des modèles de description calculatoires (quand cela est possible⁶).

Bien évidemment, tout est encore ici question de degré. Une variation d'histogramme de couleur *peut* être interprétée comme un changement de plan par certaines personnes. Mais on a ici un changement qualitatif ; *plan* désignera plus en tant que terme qu'une simple variation d'histogramme, par exemple que la partie ainsi annotée est une unité de montage, choisie à un moment donné pour s'insérer entre deux autres plans, et est donc un plan « humain » et non calculatoire.

Le problème du calcul de caractéristiques de haut-niveau est donc bien celui de l'articulation entre des résultats de calcul signifiants en tant que résultats, et leur signification dans le cadre de la tâche humaine (en collaboration avec la machine) d'exploitation de ces résultats⁷.

Par exemple, une similarité d'image peut ne pas être intéressante du tout pour qu'un utilisateur fasse de la recherche, mais peut l'être par exemple pour mettre en place des correspondances entre plans signifiant leur appartenance à une même *scène* laquelle sera utilisée comme unité de découpage du document, y compris en oubliant tous les calculs intermédiaires.

Caractéristiques interprétées. Dès qu'une caractéristique calculée de haut-niveau est validée par un utilisateur, elle prend le statut de caractéristique interprétée, c'est à dire qu'elle n'est plus redevable au calcul dans ses utilisations, mais bien au modèle de description de haut-niveau mis en place par l'homme. L'articulation définie plus haut disparaît donc dans la mise en place avec validation de caractéristiques de description, par exemple *Plan*, *Chirac* ou *Journal télévisé*.

Les mots-clé, les textes, les images ou tous documents utilisés pour décrire des documents audiovisuels résultent d'une même approche, au sens où l'être humain, appréhendant un document en réalise directement une description dans un format adapté à la machine, sans autre médiation calculatoire que la conformance, justement, au modèle de description.

Nous étudierons par la suite les modèles de description, essentiellement en terme de structuration des documents, et de structuration et d'organisation des connaissances de description.

4.2.2 L'organisation des connaissances décrites : la structure

Structuration interne d'une annotation de partie

Nous avons vu au chapitre précédent que les annotations d'une partie de document pouvaient être structurées, par exemple pour décrire le contenu d'une image en termes de relations entre

6. Citons cette éclairante déclaration de R. Jain [1996] : « Keywords often provide a better description of the person who assigns them than they do the image » ce qui signifie pour les auteurs que la diversité de l'appréhension humaine d'une image ne doit non pas être résolue en agissant sur les possibilités de description par l'homme des images (problème étudié depuis fort longtemps dans les bibliothèques), mais bien en allant chercher dans celles-ci un contenu objectif, supposé partagé, et — cela va sans dire — calculable.

7. Cela montre également qu'il n'est pas utile de traiter du signal pour traiter du signal, mais bien dans le cadre d'une aide à l'exploitation des documents en tant que signal. Le traitement du « signal documentaire » doit donc être guidé par les besoins réels des utilisateurs de documents, qui n'ont pas à s'adapter aux similarités qu'on leur propose.

formes simples (*cf.* la figure 3.3 page 77), ou bien pour organiser des descripteurs en fonction de leurs catégories de description (par exemple une description des objets et une description des sensations ressenties sur ces objets (*cf.* la figure 3.2 page 76)). Il s’agit alors de mettre ensemble des caractéristiques de façon organisée (et non plus simplement les unes à côté des autres).

Structuration hiérarchique documentaire

Le niveau suivant d’organisation des caractéristiques consiste à inscrire celles-ci dans une temporalité explicite (donc allant au delà de la simple structuration implicite qui découle, rappelons-le, des relations temporelles entre les parties de documents qu’elles annotent). Il s’agit donc de mettre en place une structuration des parties de document que l’on a découpées.

Nous avons également vu que nombre d’auteurs — commençant paradoxalement à prendre la mesure du caractère *documentaire* du document audiovisuel — en proposant désormais des organisations hiérarchiques de parties. Ces hiérarchies sont le plus souvent à trois niveaux, à savoir documents, scènes et plans, et autorisent une navigation le long de la hiérarchie, mais aussi la mise en place d’annotations à tous les niveaux, c’est à dire des description multi-niveaux. Par exemple, on annotera le document dans son ensemble à l’aide d’une structure de cataloguage comportant les noms des réalisateurs, acteurs, l’année de tournage et le titre d’un film, et on annotera des plans en décrivant ce qu’il se passe à l’image. Un modèle comme celui de [Corridoni *et al.*, 1996] est à cet égard un exemple paradigmatique.

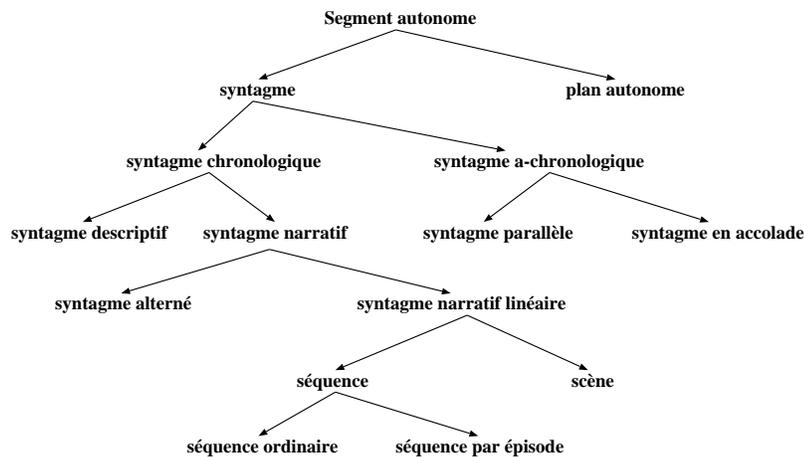


FIG. 4.1 – *La Grande Syntagmatique*, d’après [Colin, 1992], page 49.

Il nous faut cependant interroger cette structuration hiérarchique, héritée d’une tradition structuraliste française organisée autour de C. Metz. En effet, celui-ci a popularisé dans sa théorie de la Grande Syntagmatique l’idée de *grammaire filmique*, c’est à dire d’analyse du film comme composé d’une ensemble d’unités *syntaxiques*, que sont les plans (équivalent à des phrases) (une décomposition due à Colin est présentée figure 4.1). Les plans sont à leur tour regroupés en *segments autonomes* (généralement appelé *scènes* jusqu’ici, mais plutôt *séquences* dans le cadre de la sémiologie du cinéma). Le segment autonome permet « un certain niveau de segmentation de la surface discursive du filmé » [Colin, 1992] : un film est découpé en segments autonomes, puis en plans. Pour Metz [1968], « en distinguant le «plan» et la «séquence», le langage courant marque bien qu’il y a au cinéma deux choses différentes (sans préjudice d’éventuels niveaux intermédiaires) : d’une part le segment minimum, qui est le plan, d’autre part le segment autonome ». Colin, dans sa revisitiation de la GS précise : « au départ, donc, la notion de segment autonome

cherche à rendre compte de ce fait intuitif, qui est que le film en tant qu'il est discours en images, ne peut pas être simplement défini comme succession de plans, un peu de la même façon qu'un discours verbal ne peut être simplement défini comme succession de phrases » et ajoute « il y a aussi bien entendu la différence déjà faite par les techniciens du film entre séquence et plan, pour rendre compte de la pratique du découpage et du montage, qui implique des « règles » d'agencement entre les plans afin, notamment, de rendre intelligible les relations spatio-temporelles qu'ils entretiennent » [Colin, 1992].

Il s'agit donc de découper un document en plans, et de regrouper ceux-ci en scènes, afin d'affirmer dans la documentation l'expression d'un contenu dépassant le niveau du simple plan, mais restant en deça de l'annotation du document entier.

Il convient cependant de remettre en cause ce type de hiérarchie, pour un certain nombre de raisons :

- il n'est pas certain que la description d'un plan soit utile dans toute indexation d'un document audiovisuel. S'il est évident que le plan est l'unité de montage vidéo, et est à la base de « l'écriture cinématographique » du document, il n'en est pas forcément l'unité d'exploitation ultime. Par exemple, une retransmission de match de tennis pourra se structurer en fonction des points du jeu et être utilisée comme telle ;
- tout document audiovisuel n'est pas un film, mais participe d'un genre, qui peut être autre. La partition en trois niveaux se retrouve par exemple en difficulté pour décrire des documents qui en contiennent d'autres comme les journaux télévisés.
- un document audiovisuel superpose plusieurs flux, y compris des flux audio. Il peut être utile de décrire un concert filmé de musique classique par une hiérarchie basée sur les mouvements d'une symphonie *et* sur une hiérarchie de plans et de séquences.

Revenons un instant sur la notion de segment autonome. Colin [1992] finit par considérer qu'il y a en fait deux approches différentes dans la Grande Syntagmatique : la segmentation, et la catégorisation. La première vise au repérage des segments autonomes, la seconde se donne pour objectif de les catégoriser⁸. Ceci donne lieu à un certain nombre de difficultés. Ainsi, les segments autonomes tels qu'ils sont repérés peuvent ne pas avoir de pertinence dans le cadre de la catégorisation : il est possible de considérer des syntagmes trans-segmentation, de même que des syntagmes enchâssés⁹.

En fait, il nous semble que le syntagme metzien est une unité d'un document audiovisuel qui est définie suivant une analyse donnée, dans un contexte donné. Plus généralement, décréter un segment revient à lui attribuer une qualité sémantique au regard de l'analyse, et en vue d'une exploitation future (y compris ensuite, par exemple pour mettre en rapports deux segments). On conçoit alors que de multiples analyses pouvant coexister, de multiples types de segments le puissent aussi, tandis qu'il faut à cela ajouter la multiplicité due à la superposition de flux. Enfin, c'est seulement dans le cadre d'une analyse vidéo du flux que l'on considérera le plan comme unité minimale de structuration.

Il apparaît donc qu'un document peut être décrit par autant de hiérarchies d'analyse que désiré, les éléments minimaux des hiérarchies n'ayant pas tous la simplicité du plan et le nombre de niveaux étudiés n'est évidemment pas limité.

8. On a ainsi les syntagmes parallèles, les syntagmes en accolade, les syntagmes descriptifs, les syntagmes alternés, les scènes et les séquences.

9. Ce qui prouve que Metz n'est pas si radical qu'on a pu l'interpréter, en figeant sa pensée dans un langage du cinéma strict, se rapprochant d'une grammaire telle qu'envisagée en informatique.

Cela ne signifie bien évidemment pas qu'il puisse exister des analyses valables pour tous les documents, et que des hiérarchies « génériques » ne puissent être mises en place, mais celles-ci restent à notre sens à découvrir¹⁰.

Structuration « complexe »

Nous avons vu dans le chapitre précédent que d'autres structurations que la structuration hiérarchique de segments pouvaient être mise en place

On peut par exemple structurer les descripteurs eux-mêmes, par exemple une structuration par abstraction entre un descripteur *Victor Hugo* et un descripteur *Poète* [Zettsu *et al.*, 1997], ce qui relève alors *a priori* d'une organisation de connaissances *permettant* de décrire des connaissances *utilisées* pour décrire, cas que nous évoquerons à la fin de cette partie.

Il est également possible de structurer une annotation pour exprimer par exemple une action avec un graphe de concepts liés par des primitives d'action telles que celles proposées par Schank [Uehara *et al.*, 1996], l'important étant ici que la mise en place de relations entre caractéristiques annotant des parties différentes du document soit possible. Par exemple, si *Gaston aime Jeanne* correspond à ce qui est annoncé au début d'un film par la présentation de Gaston puis celle de Jeanne, alors il est possible de mettre en place une relation d'amour entre la caractéristique représentant Gaston (plan 1) et celle représentant *Jeanne* (plan 2) liant ainsi les parties de document plan1 et plan2 par une relation factuelle n'ayant *a priori* rien à voir avec une relation hiérarchique ni une relation temporelle entre les deux plans¹¹. Au niveau suivant c'est l'histoire que l'on peut décrire, par exemple un film tel que la *Guerre des Etoiles* (1978) peut être décrit par une structure simple tirée d'une étude narratologique semblable à la morphologie du conte russe étudiée par Propp [1970] : genèse du héros, problèmes du héros (les méchants), résolution du problème (victoire sur les méchants), apothéose (joie)¹².

On voit donc qu'il peut être nécessaire pour les besoins d'une application de structurer des annotations situées dans des parties indépendantes d'un document audiovisuel, sans que cette structure soit hiérarchique, puisqu'elle ne fait que lier les parties *par l'intermédiaire* des caractéristiques (alors que dans une hiérarchie d'un match de tennis, c'est le segment désigné par *jeu* qui est en relation hiérarchique avec le segment désigné par *set*).

4.2.3 Organiser les connaissances de description

Les caractéristiques d'annotation peuvent être structurées *hors-documents*, c'est à dire organisées en tant que connaissances d'annotation. En effet, à partir du moment où les caractéristiques d'annotation ne sont pas libres, il est nécessaire de les organiser dans une « base de connaissances ». Par exemple, un thésaurus permettra de hiérarchiser des mots-clé dans un arbre (relation de spécialisation donc), tandis que des relations annexes permettront d'en faciliter l'utilisation (par exemple *voir aussi* ou *utilisé pour*).

D'autres connaissances, par exemple les connaissances sur la structure d'un journal télévisé qui permettent de reconnaître les différentes parties, ou encore certaines règles d'édition cinématographique telles que celles reprises dans [Chua et Ruan, 1995] sur la théorie du montage

10. On peut observer le même phénomène pour les documents écrits, si par exemple un rapport de recherche est décrit suivant une hiérarchie document/partie/chapitre/section/etc., il pourra également l'être suivant les segments bibliographiques, originaux, de résultats, etc. sans que ce découpage reflète forcément le premier.

11. Sans vouloir anticiper trop, c'est ce genre de relations que nous autoriserons de façon naturelle dans les Strates Interconnectées par les Annotations.

12. Lucas avouait lui-même volontier s'être inspiré des théories structuralistes de Propp pour construire le scénario de son film.

(changements de temps, d'espace, de rythme ou idéologiques) peuvent être stockées et organisées. Ce seront alors des règles de repérages de structure, par exemple, si on a repéré un plan de présentateur statique dans un document, alors ce document a des chances d'être un journal télévisé. Ces connaissances sont en fait des connaissances du domaine de l'application visée.

Enfin, terminons avec les règles de description de document, ou schémas de description, qui explicitent *comment* les descripteurs doivent être utilisés pour décrire les documents audiovisuels. Par exemple, une DTD en SGML ou en XML explicite comment il convient de décomposer un document pour le documenter.

4.3 La notion de contexte dans l'audiovisuel et sa nécessaire prise en compte

L'annotation d'un document audiovisuel consiste à décrire certaines de ses parties à l'aide d'annotations, elles-mêmes structurées entre elles soit directement, soit par la médiation de la structuration des segments de description. Il va de soi que toute mise en place de structure permet d'en envisager l'utilisation comme connaissances permettant des mises en contexte (un plan peut être considéré comme étant dans le contexte d'une scène par exemple).

Cette section est consacrée à l'étude du contexte et des relations contextuelles dans les systèmes d'information et dans les documents audiovisuels.

4.3.1 Contexte et systèmes d'information

Les systèmes d'information ont été, et sont de plus en plus étudiés sous l'angle du contexte. Il s'agit alors de définir comment la notion de contexte peut y être utilement définie, et peut permettre, de façon opératoire, d'aider à la résolution de problèmes d'exploitation d'information.

Il convient d'abord de remarquer que les termes *contexte* ou *contextuel* sont souvent utilisés *incidemment* parce que des phénomènes d'ordre contextuel ont été identifiés qui doivent être analysés. Par exemple, on peut calculer dynamiquement une page d'un système d'information hypertexte en fonction du contexte (*i.e.* du lien) qui a permis d'y accéder [Staff, 1997] ; le contexte peut également être, dans une recherche d'information documentaire, le genre d'un document (ce qui correspondrait à un rayonnement dans une bibliothèque) [Smoliar et Wilcox, 1997], *etc.*

C'est en fait seulement depuis peu que le contexte est étudié *pour lui même*¹³ et que des tentatives sont faites en vue de définir plus précisément le concept en tant que tel.

La tâche n'en est pas moins ardue, et il est admis que la notion a des multiples définitions, suivant l'angle d'analyse et les visées des auteurs qui l'étudient [Brézillon et Saker, 1999]. Le contexte a en effet été au départ étudié en linguistique (par exemple en sémantique : comment l'entour d'un mot influe-t-il sur sa signification?), mais aussi en sciences cognitives (*e.g.* comment le contexte et la situation influencent-ils la cognition?), ou encore dans les systèmes à base de connaissances et les systèmes logiques.

[Brézillon et Cavalcanti, 1997, Brézillon et Saker, 1999] tirant les leçons d'une conférence interdisciplinaire sur le contexte essaient d'en dégager quelques caractéristiques essentielles, que nous présentons dans la suite.

13. Des conférences et ateliers lui sont ainsi consacrés régulièrement ces dernières années, par exemple, les *International and Interdisciplinary Conference on Modeling and Using Context* (Rio de Janeiro, 1997 — Trente, 1999), les conférences ISIC *Information Seeking in Context* tous les deux ans depuis 1996 ou l'atelier *AAAI Workshop on Reasoning in Context for AI Applications* de la conférence AAAI 1999.

Ainsi, le contexte est une notion utile pour les systèmes dans lesquels une part de raisonnement, de compréhension, d'interprétation ou de diagnostic sont nécessaires, car ces activités reposent entre autres sur des connaissances d'expérience en général non complètement explicitées. La notion de contexte permet alors d'introduire une part de flexibilité dans l'utilisation de la connaissance, laquelle est liée aux réactions de l'utilisateur.

Le contexte est quelque-chose *autour* d'un objet, qui lui donne sens, et agit alors plus sur les *relations entre objets* que sur les objets eux-mêmes. De même, le contexte ne peut pas être considéré indépendamment de son utilisation, c'est une notion obligatoirement opératoire.

Il y a différents types de contextes suivant les objets considérés et le domaine (si certaines définitions précises existent, elles sont trop limitatives pour pouvoir définir de façon générale le concept). Les différents contextes que l'on peut définir dans un même système sont cependant interdépendants.

Il existe de multiples représentations possibles du contexte.

Si on le considère par exemple comme *ensemble de connaissances*, on distinguera entre la connaissance contextualisée (effectivement utilisée) et la connaissance contextuelle (contraignant la connaissance contextualisée). Par exemple, dans beaucoup de cas, le contexte est considéré comme un ensemble de *restrictions* limitant l'accès à des données. Ce point de vue se rencontre plus du côté de l'ingénierie, de la représentation des connaissances où il s'agit de représenter et de raisonner sur un nombre limité de connaissances symboliques au sein duquel un problème doit être résolu. Les contextes peuvent alors être formalisés, manipulés, ordonnés, abstraits (décontextualisés), *etc.* et sont considérés comme *discrets*¹⁴. Ils représentent un ensemble de propriétés associées à une entité suivant l'environnement dans lequel celle-ci baigne, c'est à dire qu'ils permettent une variation dans la définition des connaissances en fonction de leur utilisation.

On peut aussi voir le contexte, dans le cadre de l'interaction entre agents, comme *processus de contextualisation*. Ce genre de considérations se rencontre plutôt au niveau des Sciences Cognitives et de l'étude et de la modélisation des interactions et des situations, des échanges d'information, par exemple entre un humain et un système. Le contexte découle alors de l'interaction comme situation partagée, et permet également de contraindre des possibilités d'interprétation. [Brézillon et Saker, 1999] distinguent connaissances, informations et données, et considère que les données deviennent informations au travers du processus de contextualisation fondé sur les connaissances disponibles au moment de l'observation. C'est alors l'étude des humains qui régit la construction des modèles de contextes, et principalement le *contexte d'interaction* où l'environnement de la communication sert au filtrage d'information et à la co-construction de sens (par exemple dans les modèles pragmatiques d'étude de la communication entre agents).

Les auteurs considèrent que le manque de représentation du contexte en intelligence artificielle est responsable d'échecs dans les systèmes à base de connaissances, de l'acquisition de connaissance, de l'apprentissage et de la génération d'explications. Il est donc nécessaire de rendre *explicite* le contexte afin de développer des systèmes dans lesquels l'utilisateur joue un rôle important de décision. On parlera alors généralement d'explications contextuelles, d'apprentissage sensible au contexte, et d'acquisition incrémentale de connaissances pour les systèmes d'assistance intelligente à l'utilisateur.

Ainsi, se limitant aux systèmes de recherche d'information, [Brézillon et Saker, 1999] proposent une adaptation d'un modèle de Newell et Simon comme cadre permettant de prendre en

14. Dans une approche logique par exemple, [McCarthy, 1993] définit un contexte comme une collection d'hypothèses. La relation $ist(c, p)$ signifie que la proposition p est vraie dans le contexte c , une proposition ist étant elle-même toujours vraie dans un certain contexte. Une grande partie des travaux en logique se base sur l'étude des interactions entre contextes, la mise en place de contextes élémentaires qui sont ensuite couplés en fonction de la tâche courante, *etc.*

compte les différentes vues sur le contexte. Nous avons dans [Prié *et al.*, 1999a] étendu ce schéma (*cf.* figure 4.2) pour le spécialiser dans un système d'information documentaire.

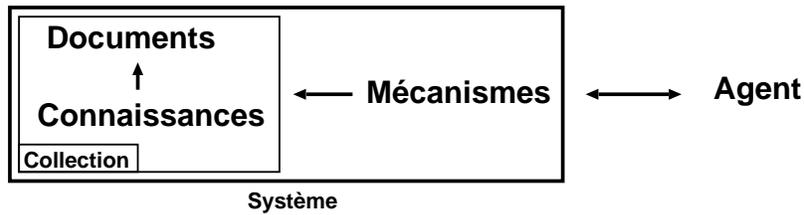


FIG. 4.2 – Un cadre permettant d'intégrer les différents types de contextes

Le système interagit avec un *Agent* (humain ou machine), et se compose de *Mécanismes* de raisonnement et d'une *Collection* d'objets sur lesquels interviennent les mécanismes, que nous subdivisons en base de *Connaissances* et en base de *Documents*. *Connaissances* contient par exemple des connaissances du domaine, des modèles de tâches ou des ontologies, tandis que *Documents* contient les documents du système d'information documentaire.

Il est alors possible de considérer des contextes à tous les niveaux du schéma : on peut définir des contextes liés aux agents (par exemple un contexte d'organisation, ou d'énonciation, ou alors liés aux objectifs des utilisateurs) ; le contexte d'interaction (lié à la pragmatique et à l'analyse du discours) se situe au niveau de la double flèche (par exemple un historique de l'ensemble des transactions) ; tandis que le contexte de la recherche d'information est lié à la flèche entre *Mécanismes* et *Collection*. Le contexte de représentation de connaissances est lié à *Connaissances*, et le contexte lié à la flèche entre *Connaissances* et *Documents* permet d'« éclairer » les documents. Au niveau des *Documents*, on retrouve les contextes linguistiques internes aux documents [Desclés *et al.*, 1997], résultant directement d'une interaction entre le lecteur et le document, ou bien bénéficiant d'un éclairage basé sur les *Connaissances*.

L'extension du schéma que nous proposons nous permet de prendre en compte les documents en tant que tels dans un système d'information documentaire, et de les séparer des données du système, donc de leur donner une indépendance. Cela nous permet de raffiner les définitions des contextes au niveau des données de la collection d'objets *Collection*, et d'intégrer par exemple le contexte linguistique documentaire dans le schéma.

Il nous semble en effet que la principale dichotomie dans la définition du contexte dans les systèmes d'information documentaire oppose l'approche textuelle/linguistique du contexte *interne* aux *Documents* (et liée directement à l'utilisateur), et une approche plus cognitive liées au point de vue de l'utilisateur de documents, et à la situation de sa pratique (*Agent*). Dans ce grossier schéma, l'intelligence artificielle se situe entre ces deux extrêmes, participant d'une part à éclairer les documents par des connaissances, et, d'autre part, au niveau de l'utilisateur, à gérer la tâche de celui-ci (en fait la représentation en machine de cette tâche).

L'ordinateur agit alors comme médiation entre le document et l'utilisateur/utilisation du document, prenant en compte aussi bien les contextes documentaires liés aux connaissances de documentation des documents que les contextes d'utilisation liés aux utilisateurs, et à leur tâches d'utilisation des documents. Nous nous intéressons dans la partie suivante au contexte documentaire dans les documents audiovisuels.

4.3.2 Contexte et audiovisuel

C'est une trivialité de dire qu'un document audiovisuel est éminemment contextuel, c'est à dire que la perception d'une quelconque de ses parties subit l'influence de son entour pour sa

compréhension (souvenons-nous des études des formalistes russes sur le montage et la compréhension des plans en contexte). En retour, toute partie d'un document contribue par sa présence à la compréhension des autres et du document dans sa globalité. Remarquons par exemple que présenter un plan seul comme extrait d'un document à l'utilisateur dans l'interface graphique d'un système d'information audiovisuel n'est pas forcément pertinent, puisqu'il y aura alors perte de sens du fait de la perte d'enchaînement avec les plans précédents et suivant. Le montage d'un document est la mise ensemble d'éléments suivant des règles mises en place au fil des années, dans l'objectif de faire passer un contenu documentaire au spectateur, en faisant une utilisation la plus rationalisée possible des règles d'appréhension du contexte dans l'audiovisuel. Par exemple, un bruit sourd accompagnant l'image d'un coup de poing impose au spectateur de ressentir de façon bien plus importante le coup en tant que tel. Une musique pourra accompagner un personnage au début d'un film, et être ensuite utilisée pour évoquer (consciemment ou non) la présence du personnage dans la suite. Les exemples sont aussi nombreux qu'il y a de manière de filmer, de monter et bien entendu d'appréhender les documents audiovisuels.

Nous ne nous intéressons pas ici à la manière dont la vision d'un document audiovisuel et son interprétation sont contraintes par le contexte « cognitif » et les liens réalisés dans le cerveau de l'utilisateur. Nous nous limitons au contexte tel qu'il peut être étudié au niveau de documents audiovisuels annotés, c'est à dire tels que les connaissances utilisées pour étudier le contexte soient opérationnalisées en machine¹⁵.

Nous considérons alors qu'il existe dans l'audiovisuel deux types de contextes liés à leur annotations.

Contexte temporel

Le *contexte temporel* est le contexte d'appréhension le plus simple, lié à la temporalité du médium. Deux plans qui se suivent seront appréhendés en tant que tels, c'est à dire que le deuxième plan sera regardé en connaissant déjà le premier. Dans le cadre de la stratification, les annotations de deux strates qui se recouvrent partagent un même contexte temporel. Davis [1993], à la suite de [Aguierre Smith, 1991] considère en effet que deux types d'informations se trouvent dans les documents audiovisuels : les informations *locales*, qui décrivent ce qu'il se passe à l'image¹⁶ et doivent être le plus possible décontextualisées ; et les informations *globales*, qui sont traitées contextuellement et mettent en relation les informations locales. Le courant de recherche autour de la stratification s'est donc dès le départ opposé à l'approche de segmentation, prônant une annotation libre du flux (*i.e.* autorisant autant de couches de description que nécessaire), et a étudié le contexte, reconnaissant qu'il était nécessaire de contextualiser toute partie de document afin de lui donner sens. [Aguierre Smith, 1991] : « The task that lies ahead is to recognize context as an essential element of the computer representation of the moving image ».

Le vecteur du contexte temporel est simplement la temporalité du flux, qui permet de mettre en rapport les parties de documents, et dans le même temps les annotations.

Contexte conceptuel

Le *contexte conceptuel* ou *sémantique* concerne toutes les autres relations contextuelles, par exemple le fait qu'un ensemble de plans soient regroupés dans une scène, ou que la voix d'un

15. Ces dernières peuvent bien entendu expliciter des phénomènes cognitifs, mais il y a lieu que ce soit en machine.

16. Rappelons que le système de [Aguierre Smith et Davenport, 1992] a été mis en place pour gérer les notes de travail vidéo des anthropologistes. Le « contenu de génération » correspond alors à la réponse aux questions où, qui, quand, pourquoi, comment, et doit être documenté au moment du filmage.

personnage en accompagne un autre. En fait, le contexte sémantique fait appel à n'importe quel lien de *co-référence explicite* (explicitée) entre objets temporels du flux audiovisuel. Par exemple, le personnage habillé en noir qui est sorti de la voiture est bien le même que celui qui ouvre la porte au plan suivant, ce qui va nous permettre de comprendre ce qu'il fait ensuite.

Le contexte *structurel* provient de la structuration hiérarchique du document : les parties de document sont mises en contexte par le biais des liens hiérarchiques. Par transitivité, les annotations annotant des parties de documents sont en relation par ce même biais, c'est à dire que certaines annotations peuvent être expliquées et enrichies par les annotations avec lesquelles elles sont en relation. Par exemple, les annotations d'un plan peuvent annoter contextuellement la séquence qui le contient [Chua et Ruan, 1995], 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 structurelles [Chiaramella, 1997], le problème étant alors de décider quels attributs vont être propagés et quels autres ne le seront pas.

Les contextes sémantiques autres que le contexte structurel prennent avantage de toutes les relations possibles de structuration du document. Par exemple un graphe d'annotation basé sur une description narrative telle que proposée dans [Uehara *et al.*, 1996] permet de mettre en relation contextuelle deux parties de document annotées par des actions participant d'une même histoire. D'une manière légèrement différente, les liens de connaissances entre annotations — par exemple le fait d'utiliser deux fois la même annotation pour annoter deux parties différentes d'un même document — permettent de mettre en relation contextuelle ces deux parties. [Hjelsvold *et al.*, 1995c] met en place des contextes basés sur la réutilisation de parties de documents dans d'autres, ce qui est fréquent dans les journaux télévisés. La structure ici utilisée concerne donc des « liens de réutilisation » de parties de documents, très utilisée dans le domaine des journaux télévisés.

En conclusion de cette partie, retenons que le contexte audiovisuel peut et doit être pris en compte dans l'exploitation des documents, par le biais des annotations. Deux vecteurs de contextualisation existent, d'une part la temporalité inhérente aux flux audiovisuels permet de toujours situer temporellement l'une par rapport à l'autre deux annotations d'un même flux ; d'autre part la structuration de l'annotation permet de mettre en place et d'utiliser des relation contextuelles conceptuelles explicites.

4.4 Description et exploitation en contexte

Nous sommes maintenant à même de présenter ce que sera notre approche de représentation audiovisuelle. Nous considérons en effet qu'il faut mettre en place des annotations et de la même façon le contexte qui les explique au moment de l'écriture, permettant leur compréhension par contextualisation au moment de la lecture.

4.4.1 Description de documents

Nous considérons alors l'indexation comme une description en contexte, une *écriture* sur le document à l'aide des caractéristiques de description disponibles qu'il s'agira de structurer.

Plus, nous posons que les principales tâches exploitant un système d'information audiovisuelle peuvent se décrire en terme de *description*. Ainsi :

- l'*indexation* est bien évidemment une description par un documentaliste suivant un ou plusieurs schémas de description qu'il y a lieu de suivre ;

- la *recherche* se ramène à la description d'un document (ou d'une partie de document) telle que la personne qui recherche aimerait la trouver. On pose ainsi une requête dans les termes du langage de description ;
- la *navigation* dans le document (ou d'une partie de document à une partie d'un autre document) correspond au calcul (éventuellement trivial car un lien peut être explicite) d'un point d'arrivée qu'il s'agit de décrire ;
- l'*édition* d'un nouveau document audiovisuel correspond à sa description à l'aide des éléments de caractérisation disponibles, à charge pour le système de trouver les parties de documents convenables ;
- l'*analyse* d'un document correspond à sa sur-description par exemple par un chercheur. Le processus de lecture/écriture mis en place correspond à celui qui est utilisé en phase d'indexation, mais l'annotation est moins contrainte. Par exemple Hibino et Rundensteiner [1997, 1998] mettent en place une interface d'annotation de documents audiovisuels basée sur l'exploration temporelle du médium, et s'en sert pour analyser des compétitions sportives ou bien des réunions de travail coopératif¹⁷. Notons aussi que [Bachimont, 1999] reconnaît que les outils du documentaliste et de l'analyste n'ont pas de différence de nature.

4.4.2 Ecriture et lecture en contexte

Nous avons vu qu'il était nécessaire de prendre en compte le contexte dans les annotations de documents audiovisuels, et que ce contexte prenait appui soit sur la temporalité inhérente au médium, soit sur la structure construite. En d'autres termes, l'annotation pourra toujours être interprétée dans le contexte temporel des autres annotations (et de la visualisation des parties de document qui y correspondent), mais la mise en place d'une structuration de celle-ci lui permettra d'être elle-même son propre substrat d'interprétation contextuelle.

C'est à dire que si le contexte temporel est partagé dans toute utilisation de document audiovisuel, principalement pour la visualisation (du document, mais aussi des annotations), nous considérons que le contexte sémantique est *lié aux circonstances de la contextualisation* par l'utilisateur, donc à ses différentes tâches d'exploitation.

Comme nous avons vu que ces tâches impliquaient toujours une description de parties de documents audiovisuels, cette description n'est rien d'autre que l'explicitation d'annotations symboliques et leur placement dans des contextes où elle prennent sens. Nous considérons que la mise en contexte, la contextualisation est une opération de base de toute tâche de description.

Ainsi,

- dans l'indexation, chaque annotation symbolique choisie pour la représentation d'un événement d'intérêt dans le document est expliquée et décrite par son contexte ;
- dans la recherche, une partie de document cherchée est décrite par un ensemble d'annotation et le/les contextes dans lesquels celles-ci prennent sens ;
- la navigation au travers d'une document se fait en passant d'une partie à l'autre en suivant des liens « faisant sens » entre parties définies par les contextes dans lesquels elles prennent leur signification ;

17. L'outil est explicitement un outil d'analyse : « This case study illustrated how our approach *can* be used to examine and identify temporal trends and how *different* types of relationships can be easily explored within MultiMedia Visual Information Seeking (MMVIS) ».

- l’analyse d’un document passe par sa description en une sur-indexation, et l’exploitation de cette dernière, par exemple en y cherchant des régularités de mises en contextes ;
- l’édition consiste à chercher des parties de documents décrites avec un certains nombre de contextes, et à en décrire les nouvelles utilisations et relations contextuelles explicites présidant au montage.

Nous posons donc qu’il existe un lien entre les contextes orientés sur les tâches (au niveau de l’utilisateur, par exemple contexte de production ou de réception du document) et les contextes intra-documentaires (niveau du document). En effet, prendre conscience d’une relation contextuelle dans un document audiovisuel est une opération réalisée dans le cadre d’un processus de description.

Finalement, nous ramenons toute tâche d’exploitation d’un système d’information audiovisuelle à une tâche de lecture et/ou d’écriture en contexte sur le document. Il s’agit alors d’être à même de décrire des annotations et des contextes, pour mettre en place et exploiter un *texte* d’annotations (le terme « texte » est ici pris dans le sens d’ensemble structuré de symboles manipulables en machine et interprétables par l’être humain). En d’autres termes, il s’agit non seulement d’annoter un flux audiovisuel, mais aussi de documenter l’action d’annotation, c’est à dire d’explicitier contextuellement pourquoi cette annotation a été mise en place et ce qu’elle signifie. Il va de soi que c’est la structuration, autorisant les relations contextuelles, qui permettra de mettre en place cette possibilité de contextualisation.

Quelques questions qui se posent alors sont : comment décrire une image, un son, un flux audiovisuel avec des termes-symboles ? Comment structurer les connaissances de description ? Comment écrire avec des symboles sur des documents non interprétés pour les documenter en vue de leur exploitation ? Comment exploiter et lire en contexte de tels « textes » ? Enfin comment lier une exploitation à la tâche d’un utilisateur ?

Autant de questions auxquelles nous tenterons de fournir un début de réponse dans la deuxième partie de ce mémoire.

4.5 Nécessités pour la modélisation audiovisuelle

Nous concluons ce chapitre en évoquant quelles sont à notre sens les nécessités à prendre en compte lorsqu’il s’agit de mettre en place des modèles de représentation de documents audiovisuels pour des systèmes d’information audiovisuelle modernes.

Tout d’abord, il nous semble nécessaire de pouvoir prendre en compte *n’importe quelle caractéristique de description*, quel que soit son niveau (primitive ou caractéristique) et le flux auquel elle s’adresse (un flux particulier ou l’ensemble).

Cependant il nous paraît nécessaire de nous concentrer sur les caractéristiques interprétables mises en place manuellement. Il y a plusieurs raisons à celà. La première est que les caractéristiques calculées interprétables souffrent le plus souvent de n’être calculées que parce qu’elles *pourraient* être utiles (par exemple un mouvement). Définir des caractéristiques de haut-niveau utiles peut alors permettre de fixer des objectifs aux calculs de caractéristiques, guidant par les besoins la recherche en traitement automatique des flux audiovisuels.

La seconde raison concerne le lien avec les primitives. Nous avons en effet vu que les calculs de similarités étaient le plus souvent considérés pour eux-mêmes, également hors-besoins¹⁸,

18. Ou plus précisément pour des besoins « standards » tellement généraux qu’ils en deviennent inutiles, par exemple « trouver une image qui ressemble à ».

et ne signifiaient la plupart du temps pas grand chose. D'autre part, les calculs de primitives nécessitent beaucoup d'espace de stockage et surtout de calculs de similarité, ils gagnent donc à être utilisés *associés* à des caractéristiques de haut-niveau, une première recherche sur celles-ci permettant de réduire l'espace de recherche dédié à la mesure de similarité. Mieux, nous pensons que c'est la mise en contexte par les annotations de haut-niveau qui permet d'expliquer pourquoi une primitive et une similarité sont utilisées¹⁹. Il conviendra donc de permettre l'étude des relations entre caractéristiques de haut-niveau et primitives dans des modèles mêlant le plus harmonieusement possible les deux niveaux.

Enfin, il nous paraît que la modélisation de documents audiovisuels devra permettre d'intégrer toute caractéristique non encore connue ou étudiée, ainsi que toute (future) documentation livrée avec le document électroniquement créé.

Il est également nécessaire d'avoir une description non figée d'un document audiovisuel, qui permette tout ajout. L'annotation doit être adaptée au caractère temporel du flux, et le niveau de granularité ne peut être fixé *a priori*. La description doit servir de base à toute utilisation des documents audiovisuels (recherche, navigation, édition, lecture active, présentation,...), c'est à dire qu'elle servira de documentation permettant toute exploitation des documents. L'entour d'un document (articles, scripts,...) doit pouvoir être considéré par la médiation du modèle de représentation.

Le modèle devra aussi permettre l'annotation comme écriture sur les flux audiovisuels. Cette annotation pourra se faire en contexte, temporel ou conceptuel, c'est à dire que l'on pourra expliciter toutes les structures nécessaires. L'annotation pourra se faire de façon totalement libre.

Il y a également lieu de fournir le moyen de spécifier des schémas de description, c'est à dire des règles d'indexation des documents en fonction de leurs genres (comment décrire, avec quoi, avec quelle structure). Une approche documentaire de l'audiovisuel devient alors possible²⁰, et les schémas de description devront pouvoir être étudiés pour eux-mêmes, en rapport avec les annotations qu'ils permettent de mettre en place.

La description de contextes, aussi bien intra- qu'inter-documentaire devra être possible, afin de pouvoir gérer les contextualisation de lecture ou d'écriture. En effet l'utilisateur (comme la machine) doivent pouvoir s'intéresser à la description d'une quelconque unité documentaire considérée comme immergée dans le document et la collection de documents auxquels elle appartient, contexte qui devra être contrôlé suivant la tâche courante. Le contexte en tant que tel (temporel²¹ ou conceptuel) doit être partie intégrante du système dès la mise en place de celui-ci.

19. Par exemple, si je cherche des objets archéologiques, alors il peut être utile d'utiliser telle ou telle similarité, parce que les images considérées seront connues comme images de résultats de fouilles archéologiques (données techniques, de forme, connaissance de la photographie comme prise devant un fond uni, *etc.*).

20. Nous avons tenté, au début de notre travail de définir des invariants de description, notamment structurels (en se basant sur la narratologie ou l'analyse filmique) qui conviennent aux documents audiovisuels en général. L'échec de cette approche et notre volonté de ne pas définir *ex-nihilo* des schémas de description font que nous nous sommes plutôt concentrés sur des modèles abstraits de description adaptés à l'audiovisuel.

21. En se limitant à des parties comparables temporellement, c'est à dire issues d'un même flux.

En plus et avec les schémas de description, la modélisation de documents audiovisuels devra permettre un contrôle des connaissances de descriptions en termes de descripteurs. Ceux-ci devront être organisés en fonction des tâches de l'utilisateur et les visées du système.

Le système sera en fait un système d'information, qui permettra de gérer l'ensemble de ses informations comme des connaissances, idéalement de la même manière, car toute connaissance sera connaissance d'utilisation. Ainsi, il sera par exemple utile de pouvoir raisonner sur cette connaissance, l'indexation étant alors réellement intelligente²².

La modélisation de documents audiovisuels devra permettre de tendre vers une intégration des modélisations, c'est à dire permettre de décrire au maximum toutes les utilisations d'un document en tant que description. Par exemple, la recherche comme la navigation doivent au maximum être décrites par des descriptions dans le même modèle que les documents. Ajoutons que la recherche par requêtes et la navigation devraient idéalement converger vers des descriptions uniques permettant de fondre ces deux modes de recherche.

La modélisation devra permettre la prise en compte aisée, et dès la conception, de l'expérience d'exploitation du système. Deux objectifs sont liés à cette capitalisation des traces d'expérience. D'une part pouvoir analyser les exploitations diverses qui seront faites des descriptions, la pratique de l'utilisateur, ainsi par exemple pouvoir les améliorer. D'autre part assister l'utilisateur dans ses tâches d'exploitation. Ce dernier point nécessite d'être à même de décrire les tâches de l'utilisateur comme les services rendus par le système, afin de pouvoir conserver l'expérience, la documenter, la réutiliser. l'assister.

Enfin, la modélisation choisie doit permettre l'expérimentation. En effet, à l'inverse des textes pour lesquels des modèles de description existent déjà, et ont été largement étudiés, utilisés, *etc.* bref sont raisonnablement fondés, les documents audiovisuels indexés n'existent que depuis peu, et leur exploitation numérique se cherche encore. L'écriture sur les documents audiovisuels doit donc *a priori* être très libre, simplement dans l'objectif de pouvoir étudier celle-ci et de mettre petit à petit en place des canons de descriptions pour les tâches actuellement imaginables, mais aussi à découvrir.

Nous nous intéresserons donc à un modèle de représentation qui autorise toutes descriptions et toutes utilisations. Il s'agit alors de pouvoir exprimer le maximum de choses, de pouvoir interroger de façon très variée, tout en offrant la possibilité d'apprendre du système naturellement les descriptions et les modes de description utiles.

22. Cf. le premier chapitre, p. 44 : « une indexation offrant *dans son mode de fonctionnement/représentation* même la possibilité de l'interroger sur elle-même, en tant qu'indexation et métadonnée. Tout document indexé ainsi offre par la description de son contenu structuré et indexé la possibilité d'inférences, de calcul sur ses index, que ceux-ci soient décrits en dehors de lui ou bien en son sein. Toute utilisation du document passe par conséquent par l'un de ses index, et peut être considérée comme ensemble de connaissances — comme le sont également les index utilisés comme tels en situation d'action —. Le système est alors entièrement géré par les connaissances, puisque pour toute utilisation, tout est connaissance. »

Deuxième partie

Strates Interconnectées par les Annotations pour l'exploitation contextuelle de documents audiovisuels

Chapitre 5

Modélisation de documents audiovisuels en Strates-Interconnectées par les annotations

Sommaire

5.1 Annoter un flux audiovisuel	111
5.1.1 Appréhension de flux et objets d'intérêt	112
5.1.2 Dimensions d'analyse	112
5.1.3 Relations	113
5.1.4 Structuration des connaissances d'annotation	113
5.2 Présentation générale des Strates-IA	113
5.2.1 Graphes d'annotation	113
5.2.2 Connaissances d'annotation	120
5.2.3 Discussion	123
5.3 Une présentation plus formelle des Strates-IA	124
5.3.1 Objets des Strates-IA	125
5.3.2 Graphe orienté étiqueté d'objets	126
5.4 Conclusion de cette partie	128

Dans ce chapitre nous présentons notre approche de description de documents audiovisuels en Strates Interconnectées par les Annotations (Strates-IA), ce qui correspond à la mise en place de graphes de description éclairés par une base de connaissances.

5.1 Annoter un flux audiovisuel

Nous questionnons à nouveau dans cette section la notion d'annotation de flux audiovisuels, cette fois-ci du point de vue de l'appréhension du flux, et justifions intuitivement notre approche de représentation.

5.1.1 Appréhension de flux et objets d'intérêt

Lorsqu'un utilisateur appréhende un flux en vue d'une utilisation autre que la simple visualisation, il en mène une analyse en vue d'objectifs liés à sa tâche courante. Par exemple un documentaliste des Archives de l'INA cherchera dans un journal télévisé principalement les séquences réutilisables facilement dans d'autres plans, les personnages principaux à l'écran, et le sujet d'actualité qui est illustré, en négligeant complètement par exemple les mouvements de caméra ou le rythme du reportage.

A partir du moment où est repéré quelque-chose répondant aux *attentes* (conscientes ou inconscientes) de l'utilisateur, ce quelque-chose est descripteur du flux, et peut être utilisé en tant que tel.

Nous appelons *objets d'intérêt audiovisuels* les objets — au sens le plus général du terme — qui peuvent être repérés lors de l'appréhension d'un flux audiovisuel. On peut considérer la notion d'objet d'intérêt comme le pendant conceptuel humain des caractéristiques audiovisuelles, qui en sont alors les « réifications » sous forme symbolique.

Ainsi, le repérage de l'objet d'intérêt « Chirac » (nous écrirons désormais par des phrases entre guillemets l'expression de ce qui est repéré) peut permettre de réifier celui-ci en la caractéristique symbolique $\langle \textit{Chirac} \rangle$.

Les caractéristiques-réifications d'objets d'intérêt sont par définition des caractéristiques interprétées de haut-niveau, puisqu'elles sont mises en place par l'homme. Exprimant le repérage dans le flux d'objets d'intérêt, il convient de leur associer la partie du flux qui y correspond. Par exemple, la caractéristique exprimant le repérage de « Chirac » sera associée à une partie d'un flux audiovisuel allant de l'instant t_1 à l'instant t_2 sur un flux commençant par définition à l'instant 0. Alors la caractéristique $\langle \textit{Chirac} \rangle$ annotera la partie de flux délimitée et définie par (t_1, t_2) . De la même manière, le repérage d'un « plan » conduira à la mise en place d'une caractéristique d'annotation $\langle \textit{Plan} \rangle$.

Nous avons vu que deux approches d'annotation existaient : l'approche de segmentation *a priori*, dans laquelle la partie de flux est définie avant que l'utilisateur l'appréhende et l'annote ; et l'approche de stratification dans laquelle c'est le repérage de l'objet d'intérêt qui est premier, et conduit à la mise en place de la partie de flux qui lui correspond.

Notre approche sera basée sur la stratification pour la mise en place des parties de flux et des annotations qui les définissent. Les parties de flux ainsi définies (le découpage obtenu), seront alors complétées à la manière de l'approche de segmentation *a priori*.

5.1.2 Dimensions d'analyse

Il apparaît tout de suite que n'importe quel objet, à *partir du moment où il a été repéré dans le flux* est objet d'intérêt audiovisuel, donc qu'il y a autant d'objets d'intérêt que l'on peut mener d'analyses particulières du flux.

Nous regroupons les analyses qui permettent de repérer des objets d'intérêt de même type en *dimensions d'analyse*. Une dimension d'analyse au sens général est alors un regroupement d'objets d'intérêt audiovisuels.

On pourra par exemple considérer la dimension d'analyse liée au repérage des « plans » $\langle \textit{Plan} \rangle$ ou la dimension d'analyse liée à celui des « visages ». Mais aussi des dimensions d'analyse liées à la recherche d'« unités structurelles de tout niveau » ($\langle \textit{Plan} \rangle$, $\langle \textit{Scène} \rangle$, $\langle \textit{Document} \rangle$), ou encore à l'analyse de la présence d'un « personnage public » sont également possibles : les dimensions ne sont pas figées *a priori* et permettent de regrouper des types de caractéristiques de façon pertinente par rapport aux objectifs de l'analyse en cours.

5.1.3 Relations

Nous avons jusqu'ici évoqué l'annotation de flux comme la recherche d'objets d'intérêt, leur réification informatique en caractéristiques d'annotation. Une strate est alors annotée par un certain nombre de caractéristiques. Un premier ensemble de relations entre caractéristiques d'annotations existe et découle de la manière dont elles sont temporellement situées : il est possible de calculer à partir des relations temporelles entre parties du flux des relations temporelles entre caractéristiques. Ces relations sont alors implicites. Par exemple si $\langle \textit{Chirac} \rangle$ et $\langle \textit{Mandela} \rangle$ annotent une même partie de flux, alors leur co-occurrence temporelle permet par exemple d'inférer une relation entre ces deux personnalités, par exemple « Chirac et Mandela se rencontrent ».

D'autres relations implicites peuvent exister. Par exemple si deux parties du flux différentes sont annotées par $\langle \textit{Chirac} \rangle$ alors il est possible d'en inférer que ces $\langle \textit{Chirac} \rangle$ correspondent bien à la présence du même « Chirac » dans le flux.

D'autres relations — explicites celles-là — peuvent également être détectées entre des objets d'intérêt préalablement repérés. Ces relations entre *objets d'intérêt* sont alors considérées elles-mêmes comme objets d'intérêt. Par exemple, si on a repéré les deux objets d'intérêt « Chirac » et « Mandela », alors on peut repérer un objet d'intérêt exprimant le fait qu'il est utile d'annoter « Chirac et Mandela se serrent la main ». De la même manière, si on désire explicitement affirmer que « le plan est contenu dans la scène », alors peut mettre en place un objet d'intérêt exprimant une relation d'inclusion structurelle entre $\langle \textit{Plan} \rangle$ et $\langle \textit{Scène} \rangle$.

Ce dernier exemple est important au sens où les relations structurelles sont dans beaucoup de travaux exprimées comme relations entre *parties nommées* (partie de type *plan*, partie de type *scène*) et relèvent des relations entre parties, et non entre caractéristiques d'annotation.

Nous choisirons pour notre part de n'accorder aucun primat à une éventuelle décomposition structurelle d'un document. Nous exprimerons alors tout ce qui peut être dit/pensé d'un flux comme objet d'intérêt réifié en caractéristique d'annotation, et toutes les relations d'intérêt explicites comme relations entre caractéristiques.

5.1.4 Structuration des connaissances d'annotation

Comme nous l'avons souligné dans le chapitre précédent, il est nécessaire de contrôler les connaissances de description. Il conviendra donc de décrire les caractéristiques d'annotation au moyen de caractéristiques d'annotation « abstraites » ou prototypiques. Celles-ci devront de plus être organisées dans une base de connaissances, de la même manière que les termes de description d'un système de recherche d'information textuel sont organisés dans un thésaurus.

La section 5.2 est consacrée à une description plus précise du modèle de représentation que nous mettons en place, tandis que la section 5.3 le présentera de manière plus formelle.

5.2 Présentation générale des Strates-IA

5.2.1 Graphes d'annotation

Notre modèle de représentation sera basé sur des *graphes de description* de documents audiovisuels. Nous en présentons d'abord les *unités audiovisuelles*, qui seront les parties de documents annotées, puis les *éléments d'annotation* qui permettront d'exprimer les caractéristiques d'annotation. La dernière partie est consacrée à la structuration de l'annotation.

Unités audiovisuelles

Un flux audiovisuel est un objet informatique contenant des données audio et vidéo, débutant à un instant t_0 , et se terminant à un instant final t_f .

On définit une (par la suite écrite UAV) comme une entité abstraite représentant un morceau quelconque du flux audiovisuel. Une unité audiovisuelle est donc au moins caractérisée par un identificateur de flux audiovisuel, ainsi que deux instants t_1 et t_2 du flux, et permettant de situer sans ambiguïté la partie de document audiovisuel qu'elle représente.

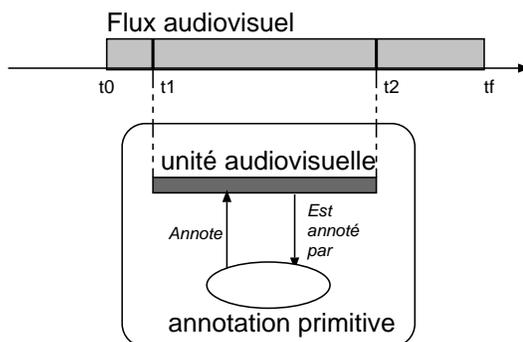


FIG. 5.1 – Une unité audiovisuelle et la caractéristique d’annotation qui la fonde

Une UAV prend naissance à partir du moment où son existence devient d’intérêt dans l’analyse du flux, c’est-à-dire dès qu’elle a été définie comme liée à un objet d’intérêt audiovisuel. Toute UAV est par définition annotée, c’est-à-dire associée à une caractéristique correspondant à l’objet d’intérêt détecté. On appelle *annotation primitive* cette annotation première et définitoire de l’UAV (voir figure 5.1).

Plusieurs remarques sont ici nécessaires.

En premier lieu, une unité audiovisuelle ne véhicule pas de contenu sémantique, si ce n’est qu’elle a été jugée d’intérêt au moment de sa création. Elle n’est donc ni nommée ni typée. C’est son annotation qui va lui donner sens, c’est à dire qu’elle n’est qu’un support d’*ancrage* de l’annotation dans le document.

En second lieu, nous ne nous intéresserons dans ce travail qu’à des documents « entiers », c’est à dire parfaitement identifiés, et dont l’instant de départ est t_0 . De tels documents seront définis par des identifiants, lesquels pourront correspondre par exemple à des fichiers MPEG, pour peu que ceux-ci les recouvrent exactement ; ou bien à des documents virtuels tels que ceux définis dans [Auffret *et al.*, 1999]¹.

Enfin, nous ne nous intéresserons dans ce document qu’à des parties de documents audiovisuels définies suivant la dimension temporelle, c’est à dire que nous négligeons la dimension spatiale des images du flux vidéo. Cependant, on peut aisément imaginer annoter une partie d’une image située à un instant t (par exemple une zone ronde située dans le coin supérieur gauche), voire un objet repérable tout au long d’une partie de document (par exemple la zone déformable correspondant au déplacement d’un personnage dans un plan). On conçoit donc l’intérêt de prendre en compte cette dimension spatio-temporelle.

1. Les auteurs proposent ainsi qu’un document virtuel soit composé de parties de documents éventuellement définis dans plusieurs fichiers différents. Par exemple, un document peut être considéré comme entier même s’il se trouve sur plusieurs bandes, ou alors il est possible de couper des parties d’un enregistrement (comme des publicités) pour les « projeter » sur un document entier virtuel.

Nous considérons que la notion d'unité audiovisuelle peut facilement s'étendre, sans changement conceptuel, à celle d'objet audiovisuel tel que défini par MPEG4, pourvu que ces objets soient considérés comme faisant partie explicitement d'un flux, *i.e.* qu'ils soient situés temporellement dans celui-ci. Par exemple des objets MPEG4 pourraient représenter ces parties d'images animées. Il conviendrait alors d'annoter — de documenter — des objets MPEG4, par exemple le fond d'une scène MPEG4 et les objets de représentation construits au premier plan.

Eléments d'annotation

Un (EA) est l'expression d'une caractéristique d'annotation, il possédera donc toujours un *nom* qui sera le terme exprimant la caractéristique interprétée. Par exemple $\langle \textit{Chirac} \rangle$, $\langle \textit{Objet Rond} \rangle$ ou $\langle \textit{Glaucque} \rangle$ seront des noms d'éléments d'annotation.

Un élément d'annotation est en *relation d'annotation* (R_a) avec l'unité audiovisuelle qu'il annote, laquelle est en relation *annotée par* (R_a^{-1}) avec l'EA.

Un élément d'annotation possède autant d'attributs que nécessaire *en plus* de l'attribut principal *nom*. Ces attributs supplémentaires représentent un second niveau de connaissances internes à l'EA qui viennent préciser celui-ci. Ils peuvent être de tous types, notamment des types correspondant à des primitives.

Donnons-en quelques exemples, qui sont illustrés sur la figure 5.2 (les éléments d'annotation sont représentés par leur *nom* dans un ovale, et leurs autres attributs leur sont associés) :

- un EA de nom *Discours* peut contenir un attribut de type *Texte* ayant pour valeur le texte du discours disponible sur la bande audio, si l'on annote par exemple une unité audiovisuelle correspondant à un discours politique à l'occasion d'un salon, mais aussi un attribut *Type de discours* et un attribut *Date* ;
- un EA de nom *Plan* peut avoir un attribut de type *image*, correspondant par exemple à une image représentative du plan. Par exemple une image extraite contenant principalement l'homme politique s'exprimant.
- un EA de nom *Éléments Description Visage* aura un attribut de type *Primitive_image_visage* avec pour valeur un ensemble de données sur le visage repéré (par exemple les rapports relatifs yeux-nez-bouche, et la couleur des cheveux...);
- un EA de nom *Jacques Chirac* pourra avoir un attribut de type *Courriel*.

A partir de maintenant, nous noterons tout élément d'annotation sous la forme suivante : $\langle \textit{Nom}; \textit{attribut1:valeur1}; \textit{attribut2:valeur2}; \dots \rangle$. Un EA sera obligatoirement désigné par son nom et par ses attributs utiles pour la tâche en cours. Ainsi $\langle \textit{Chirac}; \textit{Courriel:jchirac@elysee.fr} \rangle$ pourra également être écrit $\langle \textit{Chirac} \rangle$ si l'attribut *Courriel* n'a pas de pertinence dans le contexte de la phrase.

Une unité audiovisuelle est créée en même temps que son annotation définitive. Nous appellerons *élément d'annotation primitif* d'une unité audiovisuelle l'élément d'annotation qui la définit et la fonde. Sur la figure 5.2, les trois unités audiovisuelles *UAV1*, *UAV2* et *UAV3* ont respectivement pour annotations primitives $\langle \textit{Document} \rangle$, $\langle \textit{Discours} \rangle$ et $\langle \textit{Plan} \rangle$ (indiqués en gras).

Une unité audiovisuelle, mise en place comme strate, peut être annotée par autant d'éléments d'annotation que nécessaire. Figure 5.2, les UAV *UAV1* et *UAV3* sont ainsi annotées par des EA supplémentaires.

Si nous revenons un instant sur les attributs des éléments d'annotation, nous pouvons remarquer qu'ils permettent d'exprimer n'importe quelle caractéristique, y compris des primitives

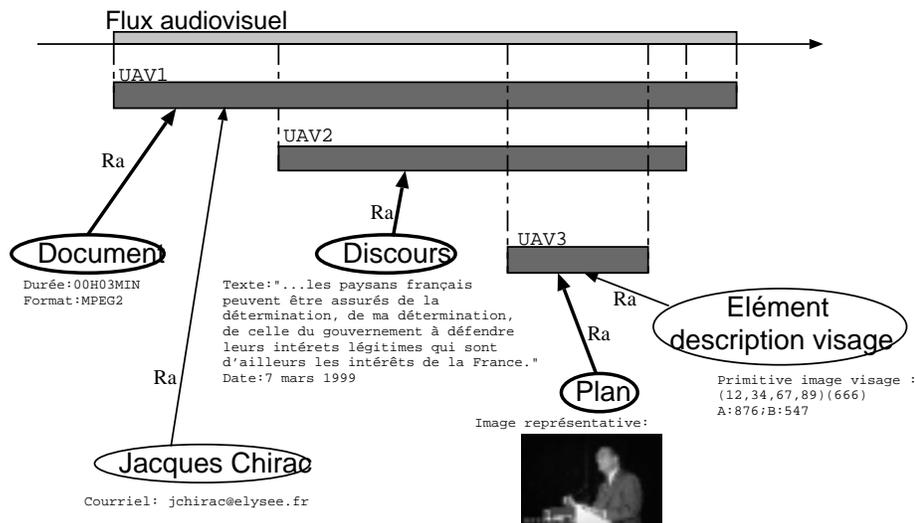


FIG. 5.2 – Quelques unités audiovisuelles et éléments d’annotation avec leurs attributs

(cf. 3.2.4). Qu’on ne s’y trompe cependant pas : le *nom* de l’EA est toujours une caractéristique de haut-niveau interprétable (par exemple *Elément description visage* ou *Discours*), et dans le cas des primitives, ce sont bien les attributs qui en portent le contenu. Les attributs supplémentaires (tel que *Texte*) permettent d’apporter des précisions utiles sur l’objet d’intérêt évoqué.

Il nous est par conséquent possible de représenter de façon homogène l’ensemble des caractéristiques que nous avons évoquées dans la première partie, quelle que soit la manière dont elles sont mises en place : analyse de couleurs, de formes, de mouvements, de ruptures de plan, des mouvements de caméra, des types de vues, de la bande-audio, d’objets, d’actions, du type de document, des auteurs, des sensations produites, *etc.*

Structuration des annotations

Nous avons vu dans la section introductive de ce chapitre qu’il y avait lieu d’explicitier les relations entre objets d’intérêt (qu’on peut également considérer comme des objets d’intérêt) en mettant ces derniers en relation. Cela se fera naturellement par la mise en relation des éléments d’annotation.

Cette mise en relation peut être envisagée de plusieurs façons :

- par la mise en place d’un ensemble de types de relations canoniques et acceptées, supposées suffisantes pour exprimer tout ce que l’application vise. L’objectif serait alors de définir toutes ces relations à l’image des travaux en intelligence artificielle et en traitement automatique de la langue visant à trouver par exemple des primitives casuelles [Fillmore, 1968] ultimes (par exemple Agent, Instrument, Localisation, Source, But, *etc.*).
- par la mise en place d’un ensemble de types de relations canoniques contrôlées, c’est à dire permettant de spécifier leur domaine d’application. Ceci suppose alors d’organiser les connaissances d’annotation et d’explicitier d’une manière ou d’une autre quels éléments d’annotation une relation spécifique peut lier. Par exemple, une relation *a pour objet* permettra de lier les EA $\langle \text{Interview} \rangle$ et $\langle \text{Personnalité} \rangle$.
- plus originalement en ne mettant en place qu’un seul type de relation — élémentaire — et

en autorisant un domaine d'application total. Alors tout éléments d'annotation peut être mis en relation avec tout autre.

C'est cette dernière solution que nous choisissons. Nous définissons la relation élémentaire R_e comme l'unique relation permettant de lier deux éléments d'annotation. Nous décrivons dans le prochain chapitre comment il est possible de spécifier quels EA peuvent être mis en relation élémentaire avec quels autres EA, c'est à dire de gagner du contrôle sur la structuration des annotations.

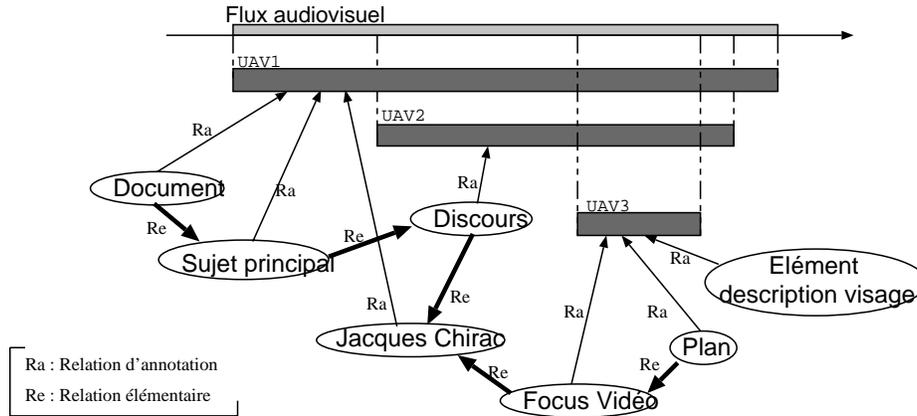


FIG. 5.3 – Mettre des éléments d'annotation en relation

Si nous continuons à nous intéresser au *processus* d'annotation, alors considérons la situation dans laquelle un utilisateur annotateur désire mettre en relation deux EA. Deux possibilités s'offrent alors à lui.

Mettre en relation directement les deux EA, c'est à dire exprimer sans plus de précision qu'il existe une relation entre les deux caractéristiques que ceux-ci expriment. On peut par exemple décider de lier $\langle \text{Discours} \rangle$ à $\langle \text{Jacques Chirac} \rangle$ (voir figure 5.3), sans expliciter outre mesure ce qu'on entend par là, l'important étant simplement la présence, à un moment donné dans l'esprit de l'annotateur d'un lien évident entre les deux EA. Cette mise en relation est donc libre, ou non précisée.

La deuxième possibilité consiste à expliciter le lien, en utilisant un EA intermédiaire qui servira de médiation permettant de préciser la sémantique de la relation. Cet EA intermédiaire joue le rôle de typage de la relation, et annote l'UAV déjà annotée par l'EA point de départ de la relation. Il exprime l'objet d'intérêt qu'est la relation repérée. Par exemple, pour exprimer que le focus vidéo (sujet principal image) d'un plan est une personnalité bien connue, on peut mettre en place un EA $\langle \text{Focus vidéo} \rangle$ annotant la même UAV que celle annotée par $\langle \text{Plan} \rangle$, et mettre en relation élémentaire d'une part $\langle \text{Plan} \rangle$ avec $\langle \text{Focus vidéo} \rangle$, d'autre part $\langle \text{Focus vidéo} \rangle$ avec $\langle \text{Jacques Chirac} \rangle$. On a alors $\langle \text{Plan} \rangle R_e \langle \text{Focus vidéo} \rangle R_e \langle \text{Jacques Chirac} \rangle$ (cf. figure 5.3).

Notre manière de considérer les relations appelle plusieurs commentaires.

En premier lieu, la relation élémentaire n'a pas sens par elle-même hormis le fait qu'elle met en relation deux éléments d'annotation. Le sens d'une relation élémentaire ne peut donc venir que de la connaissance de ce qu'elle met en relation. Ceci pourrait être rapproché des hyperliens tels qu'ils sont mis en place sur le Web: un lien n'a pas de sens pour lui-même, et ce sont ses extrémités qui en fournissent le sens.

Ensuite, les relations exprimées peuvent se situer à plusieurs niveaux : on peut avoir le niveau des relations directes ($\langle \text{Discours} \rangle \dots \langle \text{Jacques Chirac} \rangle$), celui des relations explicitées par un EA ($\langle \text{Document} \rangle \dots \langle \text{Discours} \rangle$), et puis les niveaux suivants, dépendant du nombre d'EA à traverser dans la relation par exemple une relation de « niveau 2 » entre $\langle \text{Document} \rangle$ et $\langle \text{Jacques Chirac} \rangle$, explicitée par les EA $\langle \text{Sujet principal} \rangle$ et $\langle \text{Discours} \rangle$). Il apparaît donc qu'un EA peut tout à la fois participer à l'annotation en tant que *concept* et en tant que *relation*, ceci dépendant de la manière dont il va être lu par un utilisateur. Nous évoquerons dans la suite des *éléments d'annotation de relation* (EAR) et d'autres *EA de concepts* (EAC) suivant la façon dont nous les considérerons². Il s'avère donc que le fait qu'il n'y ait pas de séparation entre concepts et relations au niveau de l'annotation autorise une grande liberté pour celle-ci. De plus, cela signifie qu'un EA utilisé pour expliciter une mise en relation fait de droit partie de l'annotation de l'UAV au même titre que les autres EA, et qu'on peut par exemple chercher les unités audiovisuelles possédant cet EA, donc cette relation.

Une mise en relation par un utilisateur peut conduire à plusieurs mises en relation élémentaires avec éventuellement création d'EA. Nous ne considérons pas *a priori* de relations inverses explicites. Toute relation élémentaire est orientée, et donc sa relation inverse existe, et est notée R_e^{-1} . Ceci signifie qu'il est possible de parcourir les relations dans n'importe quel sens, et donc qu'une relation entre deux EA est simplement un chemin non orienté qui permet de les joindre.

La mise en relation de deux EA quelconques peut conduire à plusieurs cas de figure. Si les deux EA annotent la même UAV, alors cela revient à une structuration de l'annotation interne à l'UAV, donc à mettre en place une sorte de réseau sémantique d'annotation. Si au contraire les EA annotent deux UAV différentes, alors cette structuration *intra-documentaire* permet également de mettre en place une relation entre les deux UAV, donc entre deux parties de document audiovisuel. Le nom de notre approche découle directement de cette propriété, puisqu'on parle alors de *Strates-Interconnectées par les Annotations* ou *Strates-IA* (en anglais : *annotation-interconnected strata* ou *AI-Strata*). Remarquons également que comme les deux EA sont quelconques, ils peuvent appartenir à deux flux différents, il est alors possible de mettre en place une structuration *inter-documentaire*, par exemple pour exprimer une relation de réutilisation d'une partie de document dans une autre (*e.g.* $\langle \text{Plan} \rangle R_e \langle \text{Réutilisation} \rangle R_e \langle \text{Document} \rangle$). Cette grande liberté dans la mise en relation se rapproche de celle qui est utilisée par [Davenport et Murtaugh, 1995] pour permettre une présentation originale d'un document audiovisuel composé de parties de documents à lire dans un ordre dépendant du lecteur.

Revenons un instant sur trois caractéristiques importantes du modèle d'annotation.

En premier lieu, il apparaît bien que toute unité audiovisuelle annotée peut être considérée comme unité structurelle d'un document, pourvu que son annotation se retrouve au niveau d'autres UAV et qu'il y ait un ensemble de relations entre ces annotations. Par exemple, si des UAV sont annotées par $\langle \text{Voix} \rangle$ et une autre par $\langle \text{Dialogue} \rangle$, et que $\langle \text{Dialogue} \rangle$ est dans les mêmes termes de relations avec tous les EA $\langle \text{Voix} \rangle$, alors il est possible d'extraire plusieurs structures d'arbre de la partie de document, aussi valables les unes que les autres (figure 5.4). Le modèle

2. Si les EA peuvent tous servir de relation, certains n'en sont pas moins plus « typés » comme lien, par exemple un EA $\langle \text{CStruct} \rangle$ (contient structurellement) servant à lier deux EA structurels ($\langle \text{Scène} \rangle$ et $\langle \text{Plan} \rangle$) ne sera vraisemblablement utilisé que comme EA de relation. Il en ira de même pour des EA des cas-relations tels que $\langle \text{Agt} \rangle$ (Agent) ou $\langle \text{Rcpt} \rangle$ (Récepteur) [Sowa, 1984], ou encore d'autres relations telles que $\langle \text{Meth} \rangle$ (Méthode), $\langle \text{Qty} \rangle$ (Quantité) [Biébow et Szulman, 1998]. On peut donc établir un continuum entre EA toujours utilisés comme relations et EA toujours utilisés comme concepts. Remarquons cependant qu'un EA de relation annotant une UAV n'est pas neutre : sa présence signifie qu'un autre EA de l'UAV est en relation par son intermédiaire, il y a donc bien information d'annotation.

est donc à même de prendre en compte la nécessité de multi-structuration que nous nous étions fixés³.

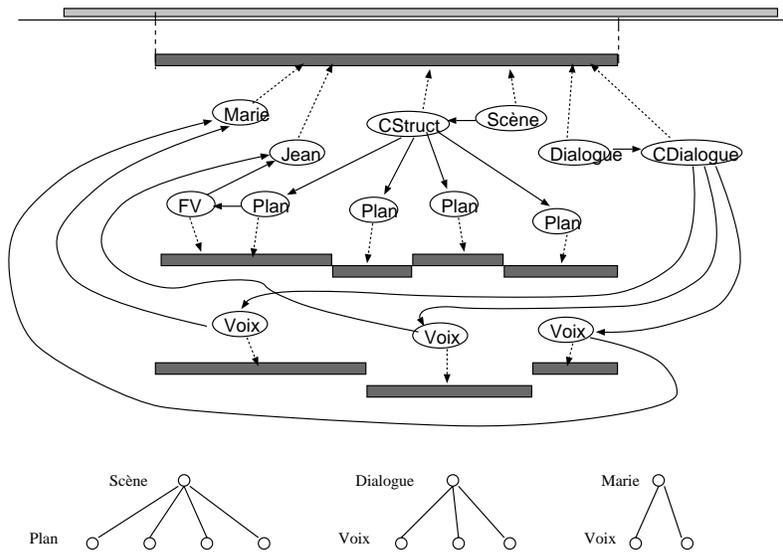


FIG. 5.4 – *Un exemple d'annotation: dans une scène Jean et Marie discutent sur quatre plans, le premier étant fixé sur Jean (par l'intermédiaire de l'EA Focus Vidéo (FV)). La bande son a trois composantes de voix. D'une telle structure on peut par exemple extraire trois structures hiérarchiques, suivant la tâche de lecture en cours.*

Une deuxième caractéristique intéressante du modèle concerne les rapports entre attributs d'éléments d'annotation et éléments d'annotation eux-mêmes. Le schéma 5.5 illustre deux cas de figure permettant de représenter la même information. Dans le premier cas, le nom du réalisateur du document est un attribut de $\langle \text{Document} \rangle$, dans le deuxième cas, on met en place un EA d'annotation exprimant que l'objet d'intérêt « Jean Dupond » a été repéré pendant l'appréhension du flux (par exemple dans le générique), tandis que $\langle \text{Réalisateur} \rangle$ permet d'expliciter le lien entre $\langle \text{Document} \rangle$ et $\langle \text{Jean Dupond} \rangle$. Le choix entre ces deux méthodes dépendra bien entendu des visées de l'application, et on remarquera que la question peut également se poser pour les autres attributs tels que *Format* ou *Durée*. En fait, ce problème est exactement le même que celui qui se pose à la création d'une DTD SGML ou XML, pour laquelle il faut choisir quelles informations seront des attributs d'éléments, et quelles autres auront le statut d'éléments, par exemple choisir entre

```
<Document Realisateur=JeanDupond>...<\Document>
```

et

```
<Document><Realisateur>Jean Dupond<\Realisateur>...<\Document>
```

Notre politique consistera à privilégier *a priori* l'annotation à l'aide des EA et de leurs relations, c'est à dire au premier niveau d'annotation. Le second niveau, celui des attributs internes

3. Nous n'en prétendons pas pour autant qu'il n'existe pas de structure « canonique » dans un document audiovisuel, il nous paraît simplement qu'il est nécessaire de pouvoir exprimer toute autre structure de la même manière. On remarquera qu'en fonction de l'utilisation certains éléments peuvent permettre de donner plus ou moins un statut d'« unités syntaxique » aux UAV qu'ils annotent. Par exemple $\langle \text{Plan} \rangle$ si celui-ci est systématiquement utilisé comme tel, ou alors $\langle \text{Voix} \rangle$, $\langle \text{Dialogue} \rangle$, $\langle \text{Musique} \rangle$ si l'intérêt principal d'une annotation du document est l'étude de la bande-son.

aux EA, est alors à réserver à la prise en compte des primitives et à d'autres renseignements supplémentaires dépendant de l'application, qui seront décrits dans les connaissances d'annotation.

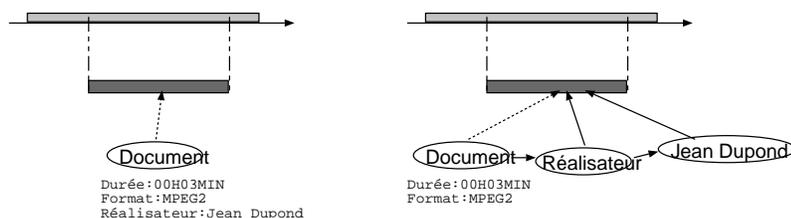


FIG. 5.5 – Deux manières de décrire le réalisateur d'un document, avec attribut interne et externe

Enfin, remarquons que l'annotation est toujours *incrémentale* : dès qu'un flux audiovisuel est inséré dans la base — par exemple en créant une unité audiovisuelle lui servant de représentant annotée par $\langle Flux \rangle$ — toute annotation est analyse d'une partie de document correspondant à une unité audiovisuelle, qui vise à la compléter. L'évolutivité de l'annotation est naturelle à l'approche, (sur-)annoter un document déjà annoté se fait exactement de la même manière. Quelle que soit la volonté de description, que ce soit par un documentaliste ou par exemple un chercheur analysant pour ses propres besoins un flux, l'annotation découle de principes identiques.

5.2.2 Connaissances d'annotation

L'approche que nous avons adoptée jusqu'ici est à la base fondée sur une forte extension des concepts de la description en Recherche d'Information textuelle classique :

- Nous étendons la notion de mot-clé à celle d'éléments d'annotations nommés et possédant des attributs secondaires⁴.
- Nous autorisons l'annotation de n'importe quelle partie de documents audiovisuels, l'annotation elle-même étant définitoire des parties considérées (par les éléments d'annotation primitifs). Cela revient à ce que les éléments d'une description tels que les UAV n'aient pas de sens par eux-mêmes, mais d'abord par leur annotation primitive, ensuite par l'ensemble des annotations auxquels ils sont liés.
- Nous autorisons toute structuration de l'annotation d'une manière homogène à l'aide de la relation élémentaire, la sémantique des relations considérées par l'utilisateur expliquées découlant d'éléments d'annotation intermédiaires. Ceci permet :
 - D'une part de structurer l'annotation interne d'une UAV, donc d'expliquer par exemple pourquoi tel EA est utilisé en compagnie de tel autre (ce qui se rapproche de la description de documents entiers par graphes conceptuels de [Genest, 1999]). Ceci permet également de spécifier le sens d'un EA par l'ajout de précision du fait des relations qu'il va entretenir *en contexte* avec d'autres termes.
 - D'autre part de structurer le document par la mise en place d'un véritable *réseau (ou graphe) d'annotation* dont tout ou partie vient finalement annoter une UAV, mais également le document. L'annotation dans son ensemble devient alors annotation structurée et complexe. Celà correspond par exemple à une généralisation des intuitions de [Uehara *et al.*, 1996].

4. Cela rejoint la critique des mots-clé et la solution proposée dans [Nack et Steinmetz, 1998], « gaining the temporality of the multilayered approach without the disadvantages of using keywords, as keywords have been replaced by a structured content representation »

- Enfin de structurer la base de documents dans son ensemble, laquelle se résume à un ensemble d'unités audiovisuelles⁵ et d'éléments d'annotation. Une base de documents est alors un graphe composé d'UAV et d'EA, le graphe d'annotation d'un flux particulier étant en fait un sous-graphe particulier du graphe global.

De la description par mots-clé de documents, nous sommes donc passés à une description par mots-clé « étendus » (ayant des attributs), qui portent toute la sémantique de l'annotation, laquelle est également connaissance de structuration du document et de la base de documents. En d'autres termes, l'utilisateur a toute latitude d'écriture d'un texte (comme ensemble organisé de symboles manipulables en machine et interprétables par l'homme) *sur* les documents (au moyen de l'ancrage fourni par les UAV) afin d'exprimer toute annotation symbolique structurée correspondant à sa tâche de description. Il s'agit bien ici de fournir des moyens d'accès direct et de manipulation des documents audiovisuels en machine⁶ largement supérieurs aux fonctions standards de type commande de magnéto-scope, afin d'en permettre une exploitation.

Il apparaît alors qu'il y a plusieurs enjeux dans l'organisation des connaissances de description.

Il faut d'abord mettre à disposition de l'annotateur un ensemble d'*éléments d'annotation abstraits* permettant de définir les éléments d'annotations qui pourront être utilisés et leurs attributs. Ce niveau d'organisation ne dépasse pas celui d'un thésaurus organisant des mots-clé.

Il est ensuite souhaitable de mettre en place des connaissances permettant de décrire comment peuvent être mis en relation les éléments d'annotation. Ces connaissances peuvent apparaître sous deux formes : soit comme connaissances *locales* aux éléments d'annotation abstraits (par exemple comme pour les concepts d'une logique terminologique) ; soit comme connaissance *globale* extérieure aux EAA (ce qui s'apparenterait plus par exemple à une DTD expliquant comment structurer un document avec des éléments). Dans le premier cas, on spécifiera par exemple dans l'abstraction de $\langle \text{Document} \rangle$ que celui-ci peut être mis en relation avec $\langle \text{Discours} \rangle$ par l'intermédiaire de $\langle \text{Sujet principal} \rangle$. Dans le deuxième, il s'agit de préciser qu'il y a lieu de décrire un document audiovisuel avec $\langle \text{Document} \rangle$ et $\langle \text{Discours} \rangle$, les deux étant mis en relation par l'intermédiaire de $\langle \text{Sujet principal} \rangle$.

Nous ne présenterons dans ce chapitre que la mise en place d'une « base de connaissances » d'éléments d'annotation abstraits, la mise en place des connaissances de mise en relation, de structuration étant traitée dans le chapitre suivant.

Eléments d'annotation abstraits

Un (EAA) permet de définir un élément d'annotation. Il possède donc le même *nom* et contient les définitions des attributs de celui-ci, ainsi que leur caractère d'obligation (obligatoire, facultatif).

Par exemple l'EA $\langle \text{Document}; \text{Duree}:00\text{H}03\text{MIN}; \text{Format}: \text{MPEG}2 \rangle$ découlera de l'éléments d'annotation abstrait $\langle \text{EAA:Document}; \text{Durée}: \text{TypeDuree}: \text{obligatoire}; \text{Format}: \text{TypeFormat}: \text{obligatoire}; \text{Date Diffusion}: \text{TypeListeDate}: \text{facultatif} \rangle$ ⁷.

5. Certaines ayant des relations plus fortes que d'autres du fait de leur appartenance à un même flux.

6. Ce que nous avons appelé « full indexing », indexation « totale » dans [Auffret et Prié, 1999] référence au « full text », plein texte, qui fournit des moyens d'accès à tout élément d'un texte.

7. Peut alors se poser la question pour les attributs ayant une valeur unique sous la forme de chaîne de caractère, c'est à dire également exprimables comme EA (par exemple l'attribut *Réalisateur* que nous évoquions précédemment). Sans apporter de réponse unique — tout dépend de l'application — nous choisissons de privilégier la notation sous forme d'EA, et réserveront de tels attributs (comme *nom*) à des EAA généraux comme par exemple $\langle \text{Homme politique} \rangle$, qui seront utilisés quand les EA adéquats n'existent pas encore (par exemple un

D'autres attributs d'EAA, appelés *valences*, et exprimant des *possibilités de mise en relation* de l'élément d'annotation découlant de l'EAA sont également définis, et seront détaillés dans le chapitre suivant.

Un élément d'annotation abstrait est en relation d'*inscription dans le flux* R_{if} avec l'EA qui en est extrait. Inversement, celui-ci est en relation de *décontextualisation* avec l'EAA dont il est extrait (cf. figure 5.6). On aura alors par exemple $\langle EAA:Document \rangle R_{if} \langle Document \rangle$ et $\langle Document \rangle R_d \langle EAA:Document \rangle$. L'inscription dans un flux signifie que l'annotation prend substance, se réifie par la médiation du flux, et en devient indissociable, en tant qu'élément d'annotation d'une UAV issue du flux. Chaque inscription est *a priori* considérée comme unique⁸ : on ne peut avoir un même EA en relation d'annotation avec deux UAV, tout EA est unique.

A l'inscription dans le flux, l'EA se contextualise, ce qui veut dire que si le nom ne change pas par rapport à l'EAA, ses attributs peuvent être instanciés, tandis que sa valeur, son contenu sémantique « de surface » varie avec les relations qu'il entretient en contexte.

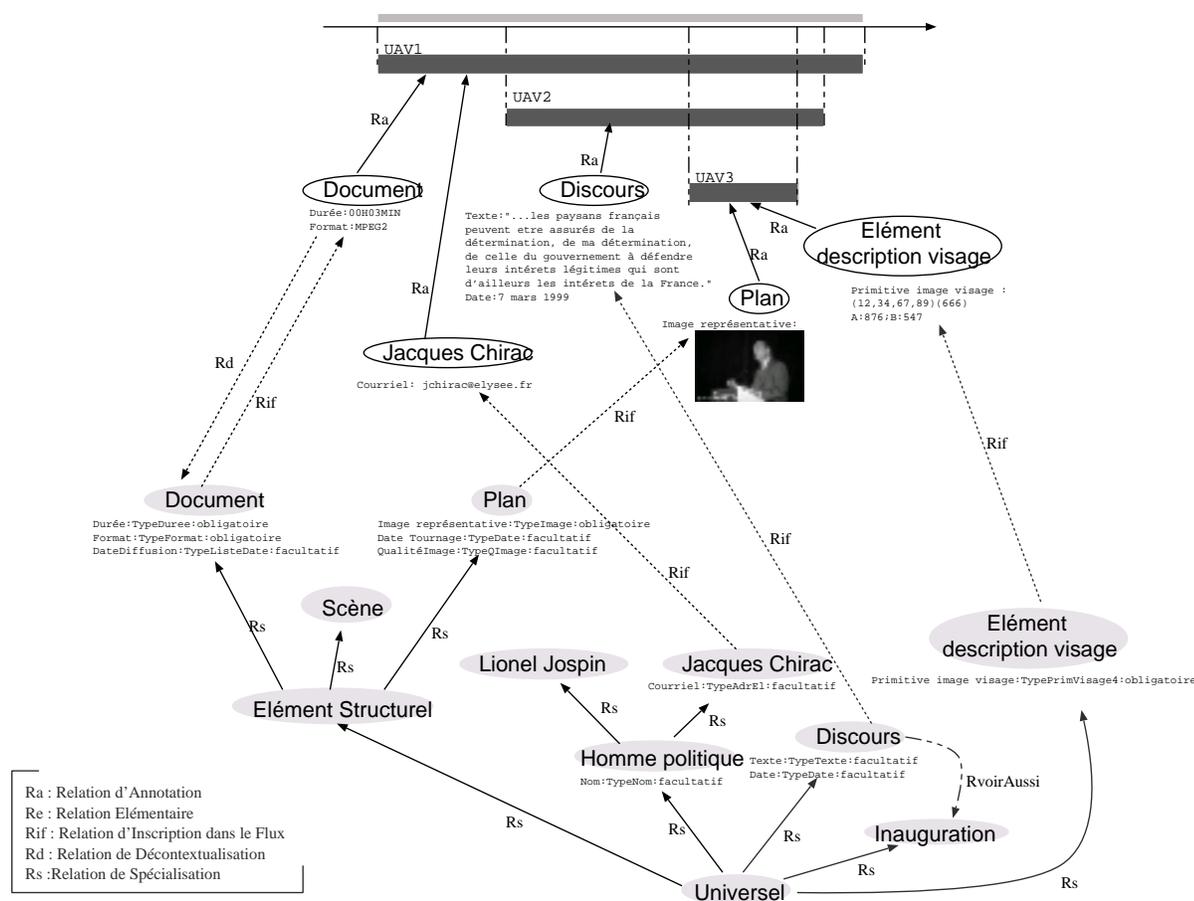


FIG. 5.6 – *Eléments d'annotation abstraits et base de connaissance*

homme politique faisant une première apparition à l'écran).

8. Ce qui nous rapprocherait d'un *hapax* tel que considéré en linguistique : tout mot est utilisé dans un contexte particulier et a un sens particulier.

Base de connaissances

Les éléments d'annotation abstraits sont organisés dans ce que nous appelons base de connaissances (BC). Une base de connaissances est donc une structure d'éléments d'annotation abstraits, basés sur des relations de concepts (à la différence des relations entre EA qui sont des relations d'instances).

Nous ne faisons pas *a priori* d'hypothèses sur les relations de la base de connaissances, si ce n'est que le graphe des EAA doit être connexe, c'est à dire qu'il doit exister un chemin (non orienté) entre deux EAA quelconques.

Une organisation de base de connaissances utilisée en recherche d'information documentaire est l'organisation en thésaurus. On a alors une hiérarchie des descripteurs suivant des relations de spécialisation/abstraction, ainsi que des relations transversales telles que *synonymie*, *voir aussi*, *etc.* On peut donc organiser les EAA en thésaurus, la figure 5.6 en donne un exemple indicatif (dans lequel on ne spécifie que certaines relations entre éléments, en négligeant les relations inverses, par exemple R_a^{-1} relation inverse de la relation d'annotation R_a).

Mais il est également possible d'organiser les EAA pour exprimer des connaissances du domaine, par exemple pour lier $\langle \text{Chirac} \rangle$ à $\langle \text{Président de la République} \rangle$, ou $\langle \text{Pédalier} \rangle$ à $\langle \text{Bicyclette} \rangle$, auquel cas on a une relation de méréonymie (partie-tout) [Winston *et al.*, 1987].

Tous les niveaux de relations sont donc possibles, et éventuellement superposables afin d'organiser les éléments d'annotation abstraits.

Nous pourrions, comme hypothèse de travail, considérer les EAA comme organisés dans un graphe (*a minima* un thésaurus) duquel il est possible d'extraire une hiérarchie en ne considérant que les relations de spécialisation/abstraction. On notera donc que cette hiérarchie n'est pas *a priori* une hiérarchie d'héritage d'attributs entre éléments d'annotation abstraits. Nous discuterons plus loin les diverses façons de considérer l'organisation des connaissances d'annotation. Rappelons simplement que notre approche d'annotation s'apparente à une écriture sur le document à l'aide d'un ensemble de termes-éléments d'annotation. Dans le cas le plus lâche, l'utilisateur écrira librement à partir des EAA de la base de connaissance qui serviront de vocabulaire contrôlé. Dans les environnements de description plus prescriptifs, la base de connaissances sera plus structurée et contraignante, par exemple toute relation mise en place sera explicitée par un EA intermédiaire, et son statut se rapprochera des véritables bases de connaissances des systèmes à base de connaissances.

5.2.3 Discussion

Nous venons de présenter la manière dont le système de représentation des Strates-IA était organisé.

Remarquons tout d'abord que pour arriver à une telle approche, il était nécessaire de s'abstraire d'une vision standard du document comme structure hiérarchique. Nous avons en effet au départ tenté d'étudier des structures générales du document audiovisuel, valables pour tous les documents, ainsi que des descripteurs généraux. Mais cette recherche ne pouvait aboutir, à moins de se spécialiser dans un domaine particulier et surtout à type d'analyse particulière des documents audiovisuels, ce qui ne correspondait pas à notre volonté de trouver un modèle suffisamment général. Nous avons donc dû passer à un niveau d'abstraction supplémentaire permettant de décrire la manière dont il était possible d'organiser toute description de document. Il était également nécessaire de réconcilier de façon harmonieuse les approches de segmentation hiérarchique et de stratification. Considérant que la stratification est plus générale que l'approche de segmentation hiérarchique, il est alors nécessaire d'utiliser une approche de stratification

permettant d'exprimer les hiérarchies d'une segmentation, d'où la mise en place des concepts d'unités audiovisuelles banalisées et d'éléments d'annotation. La mise en place d'une base de connaissances est naturelle et permet de contrôler au moins le vocabulaire d'annotation.

Le modèle de représentation des Strates-IA est à même de prendre en compte l'ensemble des modèles de représentation que nous avons étudiés. Sa validité est donc avérée théoriquement à ce niveau, comme réponse générique de modélisation de toute représentation audiovisuelle. Nous avons présenté dans [Prié *et al.*, 1998b] un tableau comparant les approches de segmentation hiérarchique standards, l'approche de stratification de [Davis, 1993] et les Strates-IA suivant les critères d'analyse des modèles de représentation audiovisuelle que nous avons évoqués dans le chapitre 3 (tableau 5.1).

	Segmentation hiérarchique	Stratification	Strates-IA
Granularité	liée à une structure hiérarchique	non limitée	non limitée
Complexité	Attributs/Valeurs	Icônes ou phrases d'icônes	non limitée
Structure	arbre	pas de structuration	non limitée

TAB. 5.1 – Comparaison entre segmentation hiérarchique, stratification et Strates-IA

Un format de représentation tel que celui des Strates-IA ne fait pas *a priori* d'hypothèses sur la manière dont il est possible d'annoter ou d'exploiter les annotations mises en place. Ainsi, on peut très bien imaginer de l'utiliser selon un modèle de base de données standard pour faire des recherches, et en extraire des portions pour les présenter à l'utilisateur, ou bien alors pour naviguer dans les documents. Ce ne sera qu'avec la notion de *contexte* que nous pourrons proposer une manière contextuelle d'exploiter le modèle.

Nous venons de montrer comment il était possible dans notre approche, de systématiser au maximum le principe de stratification afin que quelle que soit la strate détectée, celle-ci ne le soit qu'en vertu d'une analyse — et suivant une dimension d'analyse — dont l'objet d'intérêt finit par s'exprimer en un élément d'annotation annotant l'UAV correspondante. Ainsi toute annotation d'un fichier audiovisuel se résume en un ensemble d'UAV et d'EA en lien avec ces UAV, cette annotation ayant été réalisée à partir des éléments disponibles dans la base de connaissances.

On peut alors considérer que l'ensemble des documents audiovisuels annotés est un graphe composé des unités audiovisuelles, des éléments d'annotation ainsi que de leurs relations. A ce graphe est associé celui de la base de connaissances, dont les EAA sont en relation avec les éléments d'annotation. L'ensemble des connaissances du système est alors un graphe unique composé des UAV, EA et EAA et de leurs relations. La figure 5.7 présente un tel graphe, dont on peut considérer quatre sous-graphes intéressants *a priori*: les graphes d'annotation de flux $g1$, $g2$ et $g3$, ainsi que la base de connaissances. Un graphe d'annotation d'un document, un flux annoté se fonde donc dans la *base des documents annotés*.

5.3 Une présentation plus formelle des Strates-IA

Dans cette section nous décrivons un système de représentation Strates-IA de façon plus formelle que ce que nous avons déjà présenté, c'est à dire en adoptant une notation plus systé-

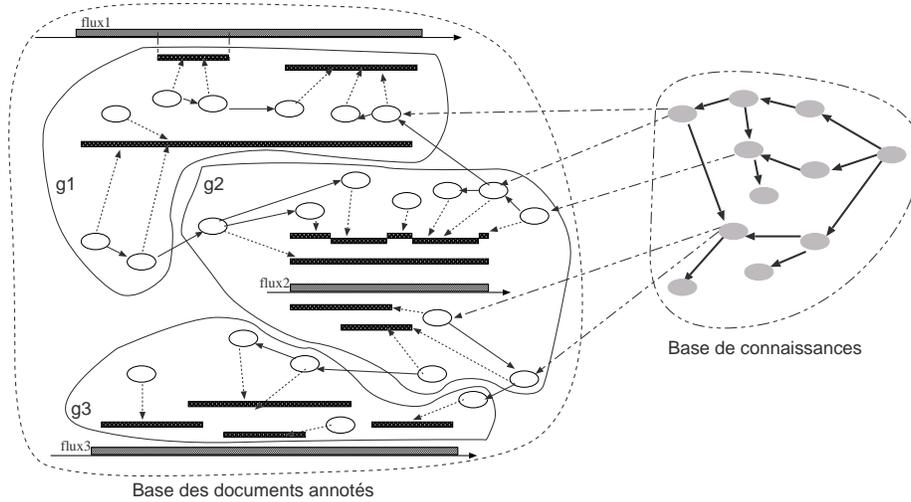


FIG. 5.7 – Une base de connaissances et sa base de documents annotés. L'ensemble représente un graphe composé de trois types de sommets (UAV, EA, EAA). A noter également que les flux annotés (UAV et EA associés se fondent dans la base des documents annotés).

matique. Comme nous considérons l'ensemble du système comme un graphe, nous procédons de la manière suivante : nous décrivons d'abord les *objets* des Strates-IA (UAV, EA et EAA), qui sont les étiquettes des sommets d'un graphe Strates-IA, que nous décrivons ensuite.

5.3.1 Objets des Strates-IA

Soit E_a l'ensemble fini des types d'attributs des éléments des Strates-IA (nous ne faisons volontairement pas de différences entre *noms d'attributs* et *types d'attributs* dans un souci de simplification). Pour tout $a \in E_a$, soit $E_V(a)$ l'ensemble dénombrable des valeurs d'attributs possibles pour l'attribut a (valeurs éventuellement complexes : tableaux, listes,...). Un couple $c = (a, v)$ avec $a \in E_a$ et $v \in E_V(a)$ est un couple attribut/valeur.

Par exemple, si $E_a = \{Nom, Date, FichierImage\}$ alors $E_V(Nom)$ est l'ensemble des chaînes de caractères de moins de 256 octets ; $E_V(Date)$ est l'ensemble des dates par exemple codées sur 8 entiers⁹ ; et $E_V(FichierImage)$ est l'ensemble des URI (Universal Ressource Identifier). On pourra alors considérer des couples $c_1 = (NomPrenom, "DupondJean")$, $c_2 = (Date, 07112004)$ ou $c_3 = (FichierImage, "http://www.insa-lyon.fr/logo.gif")$.

Un objet o des Strates-IA est un n-uplet de couples d'attributs/valeurs d'attributs

$$o = \langle c_i \rangle = \langle (a_i, v_i) \rangle_{i \in [1..n]}$$

Soit O l'ensemble des objets des Strates-IA.

Tout objet o des Strates-IA est muni d'un attribut de nom $Type$ avec $E_V(Type) = \{UAV, EA, EAA\}$, et on définit des contraintes sur les objets en fonction de leur type.

9. Tout dépend si on veut passer l'an 2000 ou non...

Unités audiovisuelles. Une unité audiovisuelle est un quadruplet

$$\langle (Type, UAV), (Nom, NomUAV), (IdentificateurFlux, IdF), (Intervalle, (t_1, t_2)) \rangle$$

Nom est un identificateur unique de l'UAV, $IdentificateurFlux$ permettant de donner une référence à une flux, par exemple une URI, et $Intervalle$ permettant de référer à une partie de ce flux.

Par exemple, on pourra considérer l'UAV :

$$\langle (Type, UAV), (Nom, Z6PO), (IdentificateurFlux, toto.mpg), (Intervalle, (134, 254)) \rangle$$

qui identifie une partie du flux *toto.mpg* définie entre les instants 134 et 254 secondes, ayant un nom-identificateur unique *Z6PO*.

Eléments d'annotation. Un éléments d'annotation est un n-uplet

$$\langle (Type, EA), (Nom, LeNomdelEA), (Attr_1 : ValAttr_1), \dots, (Attr_n : ValAttr_n) \rangle$$

Nom a pour valeur le terme de caractérisation de l'EA, $Attr_i$ a pour valeur celle de son i-ème attribut (rappelons que $Attr_i$ désigne un nom d'attribut qui contraint totalement son type, c'est pourquoi nous ne faisons pas de différence entre les deux).

Par exemple, on aura l'EA :

$$\langle (Type, EA), (Nom, Discours), (Texte : " Les paysans..."), (Date : 07031999) \rangle$$

Eléments d'annotation abstraits. Un éléments d'annotation abstrait est un n-uplet

$$\langle (Type, EAA), (Nom, LeNomdelEAA), (DefAttr : \{(Attr_1, Oblig_1), \dots, (Attr_n, Oblig_n)\}) \rangle$$

Nom a pour valeur le terme de caractérisation de l'EAA, $DefAttr$ permet de définir les types et obligations des attributs des EA qui en seront extraits.

Par exemple, un EAA lié à l'EA précédent pourra être

$$(Type, EAA), (Nom, Discours), (DefAttr : \{(Texte, Facultatif)\}), (Date, Facultatif)$$

Nous présenterons dans le chapitre suivant la manière dont il est possible de comparer deux objets des Strates-IA, c'est à dire de mettre en place des fonctions de similarités entre objets.

5.3.2 Graphe orienté étiqueté d'objets

Nous définissons l'ensemble R des types de relations entre objets, avec

$$R = \{R_a, R_a^{-1}, R_e, R_e^{-1}, R_{if}, R_d, R_{C_1}, R_{C_1}^{-1}, \dots, R_{C_n}, R_{C_n}^{-1}\}$$

où les R_{C_i} (resp. $R_{C_i}^{-1}$) sont les relations de concepts de la base de connaissances (resp. leurs relations inverses), par exemple les relations de spécialisation et abstraction R_{spec} et R_{abs} . Soit $R_C = \{R_{C_1}, R_{C_1}^{-1}, \dots, R_{C_n}, R_{C_n}^{-1}\}$.

Nous sommes alors à même de définir un graphe G des Strates-IA comme un quadruplet $\langle S, A, \mu, \nu \rangle$ avec :

- S l'ensemble des sommets de G ;
- $A \subseteq S \times S$ l'ensemble des arcs de G ;
- $\mu : S \rightarrow O$ une bijection qui associe à chaque sommet s une étiquette $\mu(s) \in O$ (l'ensemble des objets Strates-IA) ;
- $\nu : A \rightarrow L_A$ une fonction qui associe à chaque arc $a = (x, y)$ une étiquette $\nu(s) \in R$ (l'ensemble des relations Strates-IA).

répondant à une ensemble de contraintes C .

C se décompose comme suit :

- C_1 (contraintes sur les relations) : pour toute arête du graphe, celle-ci relie bien des sommets de types convenables. Par exemple R_{if} reliera bien un sommet étiqueté par un EAA à un sommet étiqueté par un EA ; une relation de R_C reliera bien deux sommet étiquetés par des EAA, *etc.*
- C_2 (contraintes de lien UAV/EA) : $\forall s_1 \in S/\mu(s_1)$ de type UAV, $\exists s_2 \in S/\mu(s_2)$ de type EA et $\nu(s_1, s_2) = R_a$ et $\nu(s_2, s_1) = R_a^{-1}$ (pour tout sommet du graphe étiqueté par une UAV, il existe au moins un sommet étiqueté par un EA en relation avec lui).
- C_3 (contraintes de lien EA/UAV) : $\forall s_1 \in S/\mu(s_1)$ de type EA, $\exists! s_2 \in S/\mu(s_2)$ de type UAV et $\nu(s_1, s_2) = R_a$ et $\nu(s_2, s_1) = R_a^{-1}$ (pour tout sommet du graphe étiqueté par un EA, il existe un unique sommet étiqueté par une UAV en relation avec lui).
- C_4 (contraintes de lien EA/EAA) : $\forall s_1 \in S/\mu(s_1)$ de type EA, $\exists! s_2 \in S/\mu(s_2)$ de type EAA avec $\mu(s_1)$ et $\mu(s_2)$ de même attribut *Nom* et $\nu(s_1, s_2) = R_d$ et $\nu(s_2, s_1) = R_{if}$ (pour tout sommet du graphe étiqueté par un EA, il existe un unique sommet étiqueté par un EAA ayant le même nom, en relation avec lui).
- C_5 (contraintes de connexité de la base de connaissances) : $\forall (s_1, s_n) \in S \times S$, $\exists c = \{s_1, \dots, s_i, \dots, s_n\}$ avec $s_i \in G$ et $\forall i \in [1, n]$ $\mu(s_i)$ de type EAA (pour tout couple de sommets étiquetés par des EAA, il existe un chemin dans G entre ces deux sommets, chacun des sommets du chemin étant étiqueté par un EAA).

La bijectivité de μ permet d'assurer une « correspondance » complète entre les sommets de S et les objets de O .

La contrainte C_5 est minimale pour la base de connaissance. Si l'on voulait par exemple assurer la présence d'une hiérarchie de spécialisation dans celle-ci, il faudrait s'assurer

- de la présence d'un sommet étiqueté par un EAA de nom *Universel*, n'étant pas en relation d'abstraction ;
- que pour tout autre sommet étiqueté par un EAA celui-ci est en relation d'abstraction avec un sommet étiqueté par un EAA, lequel est en relation de spécialisation avec lui.

Un système Strates-IA — noté *SIA* — est finalement un quadruplet :

$$SIA = \langle O, R, G, C \rangle$$

soit l'ensemble *O* des objets Strates-IA, l'ensemble *R* des relations possibles entre ces objets, un graphe *G* permettant de lier les objets de *O* entre eux à l'aide des relation issues de *R* en suivant les contraintes définies dans *C*.

5.4 Conclusion de cette partie

Nous avons dans ce chapitre d'abord présenté intuitivement notre approche de représentation en Strates Interconnectées par les annotations, avant de détailler d'une part ce qu'est un graphe d'annotation composé d'unité audiovisuelles et d'éléments d'annotation, d'autre part ce que sont les éléments d'annotation abstraits et leur organisation minimale dans une base de connaissances. Nous avons alors présenté de façon plus formelle les objets des Strates-IA, leur organisation dans un graphe Strates-IA et les contraintes associées.

L'approche des Strates-IA assure la possibilité d'écriture « informatisée » sur le flux audiovisuel, et permet une liberté totale de structuration des annotations. Nous avons pour l'instant limité la structuration de la base de connaissance à une connexité minimale car nous discuterons plus loin de quelle manière il est possible de l'organiser en fonction des types d'utilisation d'un système Strates-IA.

Nous n'avons qu'évoqué la notion de contexte dans ce chapitre. Remarquons simplement que toutes les connaissances que nous décrivons sont contextualisées. Par exemple, une unité audiovisuelle peut être décrite par son annotation directe, mais également par les éléments d'annotation avec lesquels elle est liée par l'intermédiaire des éléments d'annotation qui l'annotent. D'autre part, un élément d'annotation abstrait de la base de connaissances inscrit dans le flux en EA et mis en relation avec d'autres EA se contextualise au niveau du flux du fait de sa situation temporelle, mais également de ses relations. Dans la base de connaissances même, c'est du fait de sa position dans la structure d'organisation qu'un EAA prend sens. En fait, toutes les relations contextuelles que l'on peut considérer passent par l'utilisation de chemins d'éléments du graphe des Strates-IA.

Le chapitre qui suit est consacré à ces chemins contextuels et à leur mise en œuvre dans le cadre des Strates-IA.

Chapitre 6

Manipulations de contextes dans les Strates-IA

Nous présentons dans la première partie de ce chapitre la notion de contexte dans un système de représentation de documents audiovisuels en Strates Interconnectées par les Annotations. Un élément du graphe « dans le contexte » d'un autre élément est une extrémité d'un chemin partant de ce dernier. Afin de contrôler et de manipuler les contextes, nous mettons en place la notion de graphe potentiel et un algorithme d'opérationnalisation de la recherche de contextes. Dans la deuxième partie du chapitre, nous définissons les graphes potentiels caractérisés pour la manipulation de contextes, ainsi que quelques outils fondamentaux d'exploitation contextuelle des Strates-IA. Nous présentons alors quelques exemples d'utilisation de ces outils pour l'exploitation d'une base de documents audiovisuels annotés suivant les Strates-IA.

6.1 Contextes dans les Strates-IA

6.1.1 Un début de description

Nous présentons ici un exemple audiovisuel extrait du journal de France 2 du 13 juillet 1996¹. La figure 6.1 présente une vue globale du document : le journal est composé d'un indicatif, d'une ouverture², et de quelques reportages à chaque fois introduits par le présentateur. Le second 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 ; une rappel de 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.

La figure 6.2 présente une vue d'un début d'annotation³ du journal télévisé de la figure 6.1. Les UAV sont représentées par des cadres, dans lesquels les bornes temporelles sont indiquées en bas à droite et font 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

1. Extrait du corpus AIM (Action Indexation Multimédia), nous remercions le Département Innovation de l'Institut National de l'Audiovisuel qui a fourni une partie des matériaux de tests reproduits dans ce mémoire.

2. Dans laquelle le journaliste salue les téléspectateurs et présente les grandes lignes du journal.

3. Résultant en partie d'un travail avec une documentaliste de l'INA Centre-Est.

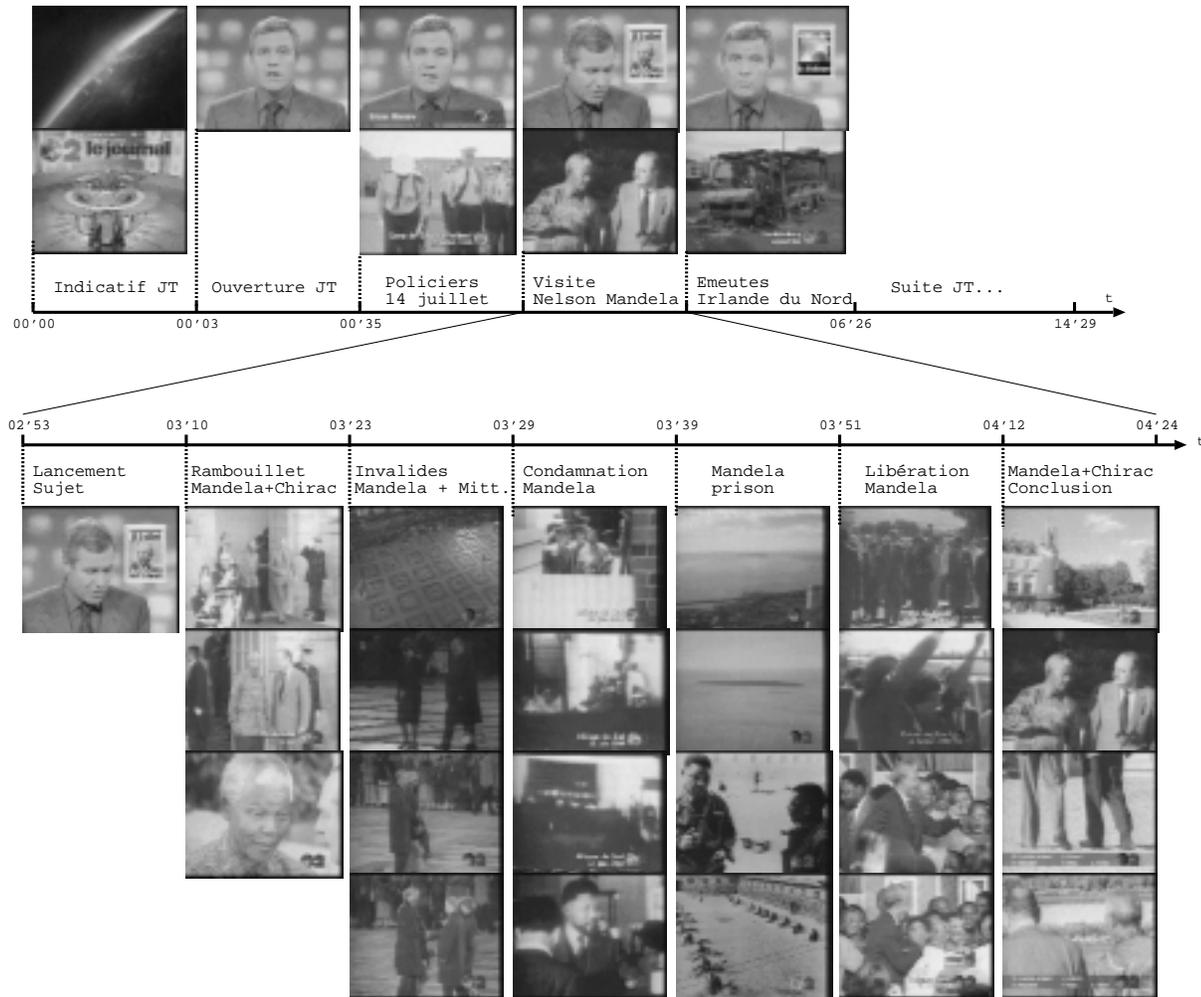


FIG. 6.1 – Vue d'ensemble du journal télévisé : la ligne de temps supérieure présente les grandes parties du journal (indicatif, ouverture, reportages), la ligne de temps inférieure détaille le contenu du reportage sur la visite de Nelson Mandela à Paris. Les suites d'imagettes se lisent de haut en bas et représentent les différents plans contenus dans le reportage

les relations élémentaires. Chaque UAV possède un numéro-identificateur unique pour le flux. On notera la double structuration du journal télévisé en ses composantes, et du document en séquences, plans, et autres documents, ainsi que la relation élémentaire mise en place entre l'EA *Extrait d'archive* et un EA *Plan* d'un autre document. Cette relation élémentaire permet donc une structuration inter-documentaire de l'annotation.

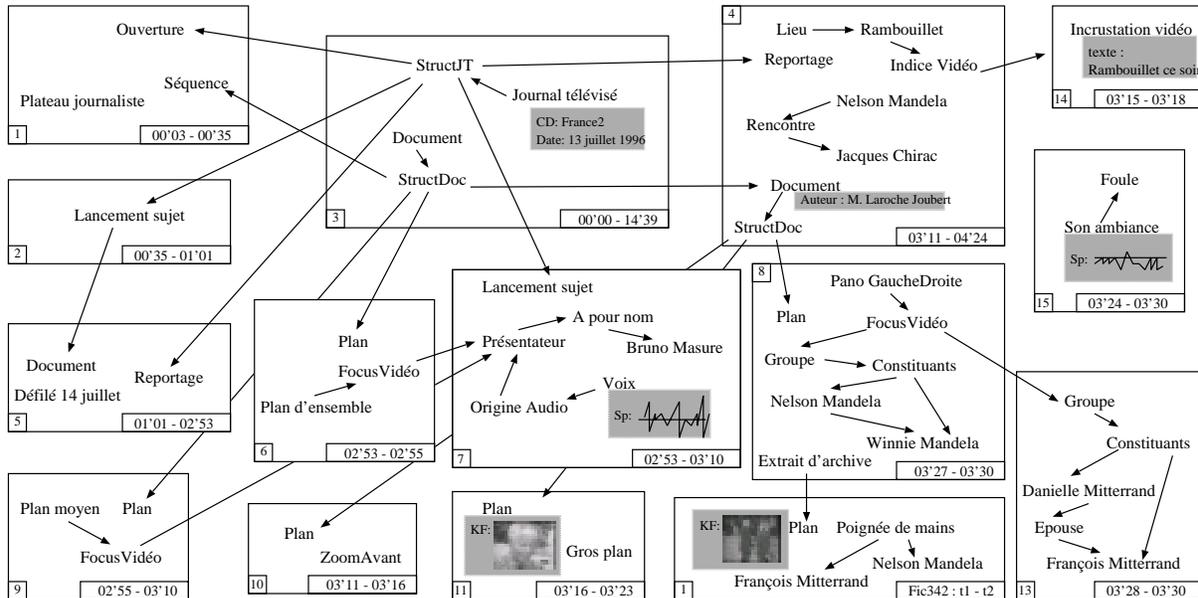


FIG. 6.2 – Une vue « emboîtée » de l'annotation d'un journal télévisé

La figure 6.3 présente une base de connaissances correspondant à l'annotation décrite figure 6.2. On notera que pour des raisons de commodité de notation, nous avons « factorisé » certains attributs dans certains EAA, tel que *Nom:TT* dans *Personnalité* ou *SP:TS* (pour un spectre sonore) dans *Son*. Il est alors admis que les EAA qui en sont spécialisation possèdent ces attributs.

La figure 6.4 présente un extrait du graphes des Strates-IA correspondant à une partie de l'annotation de la figure 6.2 et à une partie de la base de connaissances de la figure 6.3. Les bases temporelles correspondent à deux flux, on a également représentées les portées temporelles des UAV sur les flux et leurs EA primitifs.

6.1.2 Une première approche du contexte

Notre démarche se base sur une approche d'annotation de parties de documents audiovisuels à l'aide d'éléments d'annotation, exprimant des objets d'intérêt audiovisuels. Les EA sont plus particulièrement caractérisés par leur nom, à savoir le terme d'annotation, le mot-clé utilisé pour décrire ce que l'annotateur tient à exprimer à propos de la partie du document, en lien avec les relations entre EA, qui permettent de structurer l'annotation, mais également de préciser les termes utilisés. Cette écriture sur le document va ensuite donner lieu à une lecture, c'est à dire à une interprétation de l'annotation. Nous nous intéressons successivement aux trois types d'objets des Strates-IA et aux différents facteurs selon lesquels on peut les interpréter, c'est à dire à leurs interprétants dans le graphe des Strates-IA.

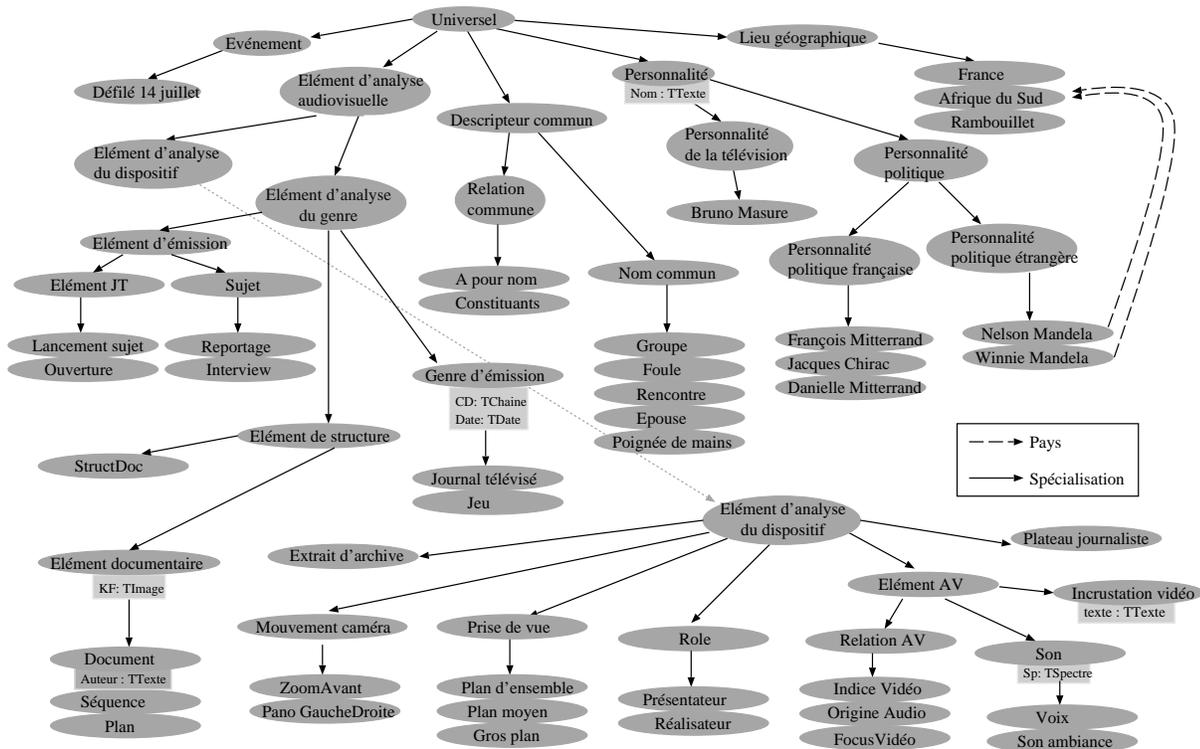


FIG. 6.3 – Un exemple de base de connaissances

Unités audiovisuelles. Nous avons vu dans la partie 4 que deux contextes audiovisuels pouvaient être considérés, le contexte temporel et le contexte sémantique.

Intéressons-nous à une unité audiovisuelle mise en place au cours d'une activité d'annotation. Cette UAV est pour le système un représentant d'une partie de flux audiovisuel, puisqu'elle permet d'y accéder de façon non ambiguë. Unité bivalente, l'UAV est donc tout d'abord décrite du côté du document audiovisuel par la partie de flux qu'elle délimite, tandis que du côté du système, c'est cette partie de flux qui est décrite par les annotations liées à l'UAV⁴. Comme de plus l'UAV est insérée dans le graphe des Strates-IA, et peut être appréhendée et manipulée en tant que telle, nous la considérons comme ayant une existence propre, et pouvant être elle-même décrite par son annotation. La partie de flux est alors documentée par la description de l'UAV.

Une unité audiovisuelle est d'abord décrite par les annotations qui lui sont liées, donc les EA en relation d'annotation R_a avec elle. Cela correspond à un contexte « immédiat », qui est le contexte de l'UAV. Si nous considérons un contexte temporel immédiat, une UAV est également décrite par les EA annotant des UAV qui la contiennent temporellement. Par exemple l'UAV 7 de la figure 6.2 est contenue dans l'UAV 3, on peut donc dire suivant ce contexte (cette relation temporelle), qu'elle est également décrite par les EA de l'UAV 3 (par exemple par $\langle \text{Journal Télévisé} \rangle$). Inversement, l'UAV 3 peut, suivant la relation inverse être décrite par les EA annotant l'UAV 7, qu'elle contient temporellement (par exemple par $\langle \text{Bruno Masure} \rangle$). Nous nous trouvons donc en présence de deux contextes temporels immédiats, basés sur l'inclusion temporelle de parties de flux. Si nous considérons les autres relations temporelles entre UAV, il

4. Il y a au moins une annotation, qui est l'élément d'annotation primitif exprimant l'objet d'intérêt repéré, une UAV ne peut exister sans annotation. On peut cependant, à la limite, envisager d'utiliser un élément d'annotation très général (*Universel*) pour définir des parties de document seulement documentées par leur existence.

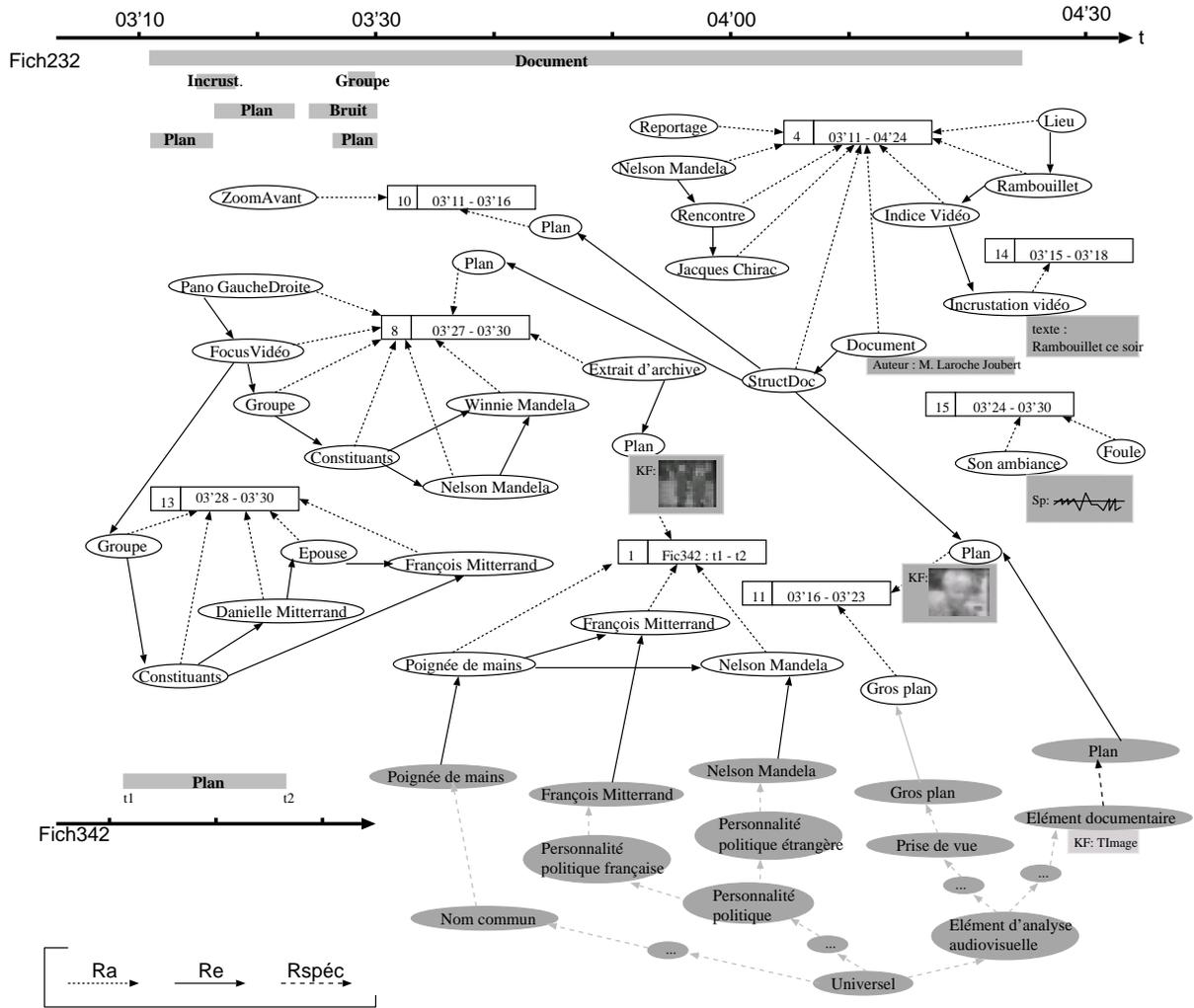


FIG. 6.4 – Extrait de graphe Strates-IA d'annotation du journal télévisé correspondant à la figure 6.2. Une partie de la base de connaissances est représentée, ainsi que quelques unités audiovisuelles (UAV n° 1, 4, 8, 10, 11, 13, 14, 15) et les éléments d'annotation qui les annotent. Les lignes de temps représentent les deux flux considérés et les unités audiovisuelles y sont décrites avec leurs éléments d'annotation primitifs

devient possible de définir tous les contextes temporels « non immédiats » utiles, par exemple la superposition (overlap), la jonction, *etc.* Toutes les combinaisons de relations temporelles telles que celles définies par [Allen, 1983] peuvent donc participer à la recherche des contextes temporels d'une unité audiovisuelle.

Nous définissons l'*annotation contextuelle* d'une UAV comme les éléments d'annotation dont on peut, en définissant un contexte, considérer qu'ils annotent cette UAV⁵. Par exemple, en définissant le contexte d'inclusion temporelle, on peut considérer que toutes les annotations d'une UAV annotent également toute UAV qui la contient temporellement. Les UAV globales jouent donc un rôle dans la détermination des unités locales, inversement, les unités locales permettent de compléter la description des unités globales.

Si le contexte temporel est lié au flux et à sa temporalité (il est toujours possible de comparer temporellement deux UAV définies sur le même flux), le contexte sémantique prend quant à lui appui sur les relations entre annotations, c'est à dire sur la structure de l'annotation. Par exemple sur la figure 6.2, le contexte structurel suivant la relation nommée par l'EA *⟨Struct Doc.⟩* permet de définir que l'UAV 9 peut être annotée contextuellement par des EA annotant l'UAV 3. Si on suit un lien de structuration d'annotation basé sur *⟨Focus Vidéo⟩*, alors l'UAV 7 se trouve dans le contexte de l'UAV 3. Enfin, si nous considérons un lien basé sur *⟨Focus Vidéo⟩*, une co-occurrence temporelle d'EA et *⟨Struct Doc.⟩*, alors l'UAV 3 se trouve dans le contexte de l'UAV 7. Certes ce contexte est équivalent au contexte d'inclusion temporelle défini précédemment, mais sa signification est différente, puisqu'ici on peut décrire le contexte par le chemin suivi dans le graphe d'annotation.

La dernière possibilité d'annotation contextuelle d'une UAV dépend encore de ses annotations, mais surtout de leurs liens avec la base de connaissances. En effet, annotée par un élément d'annotation, l'unité audiovisuelle bénéficie encore des explications sur la signification de cet élément définies par le réseau de la base de connaissances. Ainsi par exemple une unité annotée par *⟨Zoom Avant⟩* pourra également être considérée comme l'étant par *⟨Mouvement caméra⟩*, du fait de la relation de spécialisation dans la base de connaissances entre *⟨EAA:Mouvement caméra⟩* et *⟨EAA:Zoom Avant⟩*. De la même manière, une UAV annotée par *⟨Nelson Mandela⟩* peut l'être également par *⟨Afrique du Sud⟩* du fait de la relation *pays* entre *⟨EAA:Nelson Mandela⟩* et *⟨EAA:Afrique du Sud⟩*. Ce type d'annotation contextuelle profite simplement des relations de la base de connaissances, à laquelle l'UAV est connectée dans le graphe par l'intermédiaire des EA qui l'annotent.

Eléments d'annotation. Après nous être intéressés à l'annotation contextuelle des UAV, c'est à dire à leur contexte dans le graphe Strates-IA, nous pouvons appliquer un raisonnement similaire sur les éléments d'annotation. En effet, ceux-ci peuvent être également interprétés et manipulés et ici encore, plusieurs contextes peuvent être définis.

Un EA peut être considéré seul, auquel cas son interprétation sera contrainte, d'une part par l'UAV (et le flux) qu'il annote, c'est à dire par son contexte audiovisuel, mais aussi par les relations que l'EAA dont il est inscription dans le flux entretient dans la base de connaissances.

Par exemple, figure 6.2 l'EA *⟨Pano Gauche-Droite⟩* annotant l'UAV 8 s'explique d'abord par l'UAV 8 elle-même, et par la partie de flux à laquelle elle correspond, mais aussi par l'EAA *⟨EAA:Pano Gauche-Droite⟩* et les relations de celui-ci avec *⟨EAA:Mouvement caméra⟩*, ses diverses abstractions et autres spécialisations (*e.g.* *⟨Zoom avant⟩*). Si on considère les relations temporelles entre UAV, donc entre les EA qui les annotent, ceux-ci peuvent s'expliquer mutuellement

5. L'idée originale de notre approche consistait à aller dans un premier temps définir une UAV dans le contexte d'une autre, et à y sélectionner les EA d'annotation contextuelle.

(par exemple le fait qu'un EA *⟨Meurtre⟩* soit temporellement situé après un EA *⟨Insulte⟩*).

Si l'EA est considéré dans ses relations avec d'autres EA, il peut s'interpréter du fait par exemple de la présence d'autres annotations dans l'UAV (*⟨Pano Gauche-Droite⟩* annoté avec *⟨Plan⟩* l'UAV 8). Les relations élémentaires qu'il entretient avec d'autres EA permettent également de mettre en place un contexte d'interprétation. Par exemple dans l'UAV 8, *⟨Pano Gauche-Droite⟩* est en relation par l'intermédiaire de *⟨Focus Vidéo⟩* avec deux EA *⟨Groupes⟩*, eux mêmes précisés par leurs relations avec leurs constituants ; si *⟨Meurtre⟩* est en relation élémentaire avec *⟨Don Gomès⟩*, alors son interprétation est plus contrainte.

Encore une fois, il nous apparaît que la lecture et l'interprétation des annotations peut se faire en prenant en compte leurs différents contextes dans le graphe des Strates-IA, lesquels se basent soit sur des relations temporelles par le biais des UAV, soit en tirant partie de la structure de l'annotation, soit de celle de la base de connaissances. Toute connaissance inscrite dans le flux comme élément d'annotation y est en effet contextualisée aussi bien temporellement que par les relations explicites que l'on met en place. De plus, ce sont les contextes hors du flux, dans la base de connaissances qui en prescrivent la signification de départ.

Eléments d'annotation abstraits. Un élément d'annotation abstrait existe donc en premier lieu en l'absence d'annotation, et s'explique tout d'abord par le terme qui est choisi pour son nom, qui sera *toujours* interprétable et interprété, pas obligatoirement de façon unique⁶. C'est pourquoi, dans une base de connaissances comme dans un thésaurus, mais à des degrés de figement divers, les relations entre concepts (ou termes) permettent de spécifier la signification qu'il convient de mettre sous chaque terme, sous la forme de contraintes d'interprétation. La valeur sémantique d'un concept ou d'un terme dépend alors fortement du contexte dans lequel il se trouve au niveau des connaissances.

S'il est utilisé, un élément d'annotation abstrait est inscrit dans le flux sous la forme d'un élément d'annotation. La valeur en contexte (dans le flux) de l'EA peut alors naturellement varier (puisque des relations non prévues sont possibles) par rapport à la valeur hors-contexte (hors du flux) de l'EAA. En retour, sous la forme d'exemple d'utilisation, l'EA peut par sa signification contraindre l'interprétation de l'EAA, y compris du fait de sa position dans l'organisation de l'annotation. L'EAA est donc également décrit par le contexte de ses instances.

Nous l'avons vu en étudiant rapidement les possibilités d'interprétation contextuelle dans un graphe d'annotation Strates-IA, tous les objets utilisés dans ce graphe peuvent s'interpréter en fonction de la position qu'ils y occupent, c'est à dire en fonction des relations qu'ils peuvent entretenir avec d'autres objets. Le graphe des Strates-IA est un réseau de sens, avec des objets syntaxiques banalisées — les UAV — qui peuvent par conséquent bénéficier d'une sémantique associée à toute annotation contextuelle ; des termes d'annotation — les EA — qui forment un texte écrit sur le document, et à ce titre sont interprétables en fonction de relations implicites comme explicites dans l'annotation elle-même ; enfin, des termes abstraits d'annotation — les EAA — décrits hors du flux, et prenant éventuellement un sens nouveau dans celui-ci. Il apparaît qu'interpréter un objet de ce graphe n'est possible qu'en considérant son contexte, par exemple en extrayant un sous-graphe qui lui soit lié, lequel contexte servira d'interprétant local. Le contexte chez nous sera donc principalement lié à l'idée de chemins dans le graphe, lesquels seront fortement liés à la tâche courante des utilisateurs.

6. « ... le sens d'un mot ne lui est pas immanent : il est toujours le produit d'une interprétation, cette interprétation fut-elle fondée sur des inférences par défaut, qui permettent d'hériter des traits sémantiques d'un type lexical pré-construit » [Rastier, 1995b].

Ceci nous permet de justifier *a posteriori* notre volonté de considérer le système comme un graphe, car alors tous les liens de contextes entre sommets quelconques s'expriment de la même manière et de façon naturelle, ce qui ne serait pas forcément le cas avec d'autres formalisations.

6.1.3 Définition du contexte

Soit un système Strates-IA $SIA = \langle O, R, G, C \rangle$ tel que nous l'avons défini au chapitre précédent, avec $G = \langle S, A, \mu, \nu \rangle$.

Soit un objet $o \in O$ (O l'ensemble des objets) correspondant à un sommet de G $s \in S$ et $\mu(s) = o$ (tel qu'il serve d'étiquette pour le sommet s). Un *contexte* de o est un objet o' de O tel qu'il existe un chemin du graphe G permettant de mettre en relation les sommets étiquetés par o et o' . On dira que o' est *en contexte* avec o .

Le *contexte global* de o est l'ensemble des objets o_i en contexte avec o .

Cette dernière définition est peu opératoire en l'état. En effet, G étant connexe, tout objet est en relation avec tout autre objet par l'intermédiaire d'un ou plusieurs chemins. Ceci exprime simplement le fait que le système est cohérent et unique, c'est à dire qu'on met en place un ensemble de connaissances de description (EAA) liées entre elles, qui s'expliquent mutuellement, ce qui est également le cas des connaissances utilisées par les descriptions (EA et UAV). Toutes les relations contextuelles entre objets leur permettent de s'expliquer (de se définir) mutuellement (*cf.* plus haut).

Afin cependant de contrôler les contextes pour qu'ils puissent exprimer des relations *utiles* entre éléments, il convient de mettre en place des outils permettant de les définir et de les manipuler. Les graphes potentiels nous permettront de définir formellement les contextes et de fournir le contrôle voulu. Avant de les présenter, il nous faut présenter ce qu'est la similarité entre objets des Strates-IA.

Sur la similarité entre objets

Nous définissons dans cette partie comment il est possible de comparer deux objets des Strates-IA.

Pour chacun des types d'attribut de E_a , on définit une ou plusieurs fonctions de similarité, c'est à dire qu'on associe à chaque type $a \in E_a$ un ensemble non vide de fonctions $f_{a_i}, i \in [1..n]$ prenant en entrée deux valeurs d'attributs $(v_1, v_2) \in E_V(a) \times E_V(a)$ et rendant en sortie une valeur booléenne 0 ou 1.

Par exemple, l'unique fonction de similarité associée aux attributs de type *Type* sera : $f_{Type_1} : E_V(\textit{Type}) \times E_V(\textit{Type}) \rightarrow \{0, 1\}$ avec $f_{Type_1}(t_1, t_2) = 1$ si $t_1 = t_2$ et $f_{Type_1}(t_1, t_2) = 0$ sinon.

Il y a lieu de définir plusieurs fonctions de similarité pour un type d'attribut donné afin prendre en compte les « vraies » mesures de similarité liées aux primitives images et son (calculées). Par exemple, on pourra imaginer de comparer deux signatures de points d'intérêt suivant plusieurs méthodes ; ou bien on comparera deux textes soit de façon exacte, soit sur quelques-uns de leurs termes représentatifs (par exemple ils contiennent tous les deux les termes *Chirac* et *Paysans*).

Nous nous restreignons ici à des fonctions à valeurs booléennes afin de simplifier les calculs, c'est à dire que nous n'envisageons pas *a priori* d'ordonner des résultats : soit une primitive est jugée similaire à une autre, soit non.

Une fois définies les comparaisons entre attributs, nous pouvons mettre en place des comparaisons au niveau des objets. Une fonction de similarité entre objets est représentée par un ensemble de couples $\{(TypeAttribut, FonctionSimilarité)\}$, elle aura pour argument un couple d'objets $(o_1, o_2) \in O \times O$ et rendra des valeurs booléennes.

Son principe est ici le suivant : pour chacun des types d' o_1 , si il n'existe pas d'attribut de même type dans o_2 alors retourner 0 sinon comparer les valeurs d'attributs avec la fonction de similarité *ad-hoc*. Si le résultat est 1 alors passer à l'attribut suivant, sinon retourner 0. Quand tous les attributs de o_1 sont passés en revue, retourner 1.

Cette fonction n'est donc pas symétrique, le résultat de la comparaison de o_1 avec o_2 ne sera pas forcément le même que celui de la comparaison de o_2 avec o_1 .

Une fonction de comparaison simple est celle qui compare par exemple deux EA, en vérifiant que leurs types et noms sont les mêmes. Il est possible d'y ajouter une vérification que deux valeurs de primitives sont considérées comme similaires, ce qui nous permet alors de comparer des EA exprimant des primitives issues du traitement du signal.

Par exemple, on pourra comparer deux objets

$$o_1 = \langle (Type, EA), (Nom, Plan), (SignaturePointsInteret, sign_1) \rangle$$

et

$$o_2 = \langle (Type, EA), (Nom, Plan), DateTournage : 18031998, (SignaturePointsInteret, sign_2) \rangle$$

à l'aide de la fonction de comparaison définie par

$$\{(Type, cmpType), (Nom, cmpNom), (SignaturePointsInteret, cmpSPI_2)\}$$

si *cmpType* et *cmpNom* sont des fonctions de comparaison exactes, et *cmpSPI₂* permet de comparer des signatures de points d'intérêt.

Remarquons que notre définition des similarités entre objets ne se prétend pas définitive, et qu'il est possible de définir autant de principes de similarité que l'on veut, pourvu que le résultat soit booléen.

Nous choisissons des similarités booléennes entre nœuds (et non par exemple des similarités « floues » à valeur dans $[0, 1]$) car nous chercherons à exprimer des comparaisons entre graphes en terme d'isomorphismes. Alors il faudra comparer des nœuds de façon stricte.

6.1.4 Graphes potentiels pour exprimer des contextes

Si nous revenons à notre problème, celui-ci est donc le suivant : étant donné un objet o d'un système Strates-IA, comment contrôler l'ensemble des contextes qu'on pourra lui associer ? Pour répondre à cette question, il nous faut être capable de décrire dans le graphe un chemin de façon abstraite, c'est à dire hors du système, aussi bien de façon « syntaxique » (les sommets et les arcs qui les lient) que « sémantique » (les étiquettes associées aux sommets et aux arcs).

Il s'agit tout d'abord de décrire l'ossature du chemin, c'est à dire les sommets qui le composent, et les arcs entre ces sommets. La deuxième étape consiste en l'étiquetage de l'ossature du graphe (ou *chemin virtuel*), à l'aide d'objets « abstraits » pour étiqueter les sommets, et d'un ensemble de relations étendues pour prendre en compte les relations temporelles.

Objets abstraits. Un objet abstrait est un objet des Strates-IA (qui possède donc *a minima* un *type*), dont tous les autres attributs peuvent ne pas avoir de valeur, ou avoir des valeurs négligées dans les faits, ce qui revient d'une certaine façon à l'abstraire.

En effet, il s'agira de comparer les chemins, décrits de façon abstraite, aux chemins du graphe, donc de comparer les étiquettes, c'est à dire les objets abstraits aux objets concrets. Pour cela, une fonction de similarité entre objets sera utilisée, laquelle ne prendra en compte obligatoirement que le *type* de l'objet abstrait. On conçoit donc que si on ne considère que le type, il n'y ait pas lieu de définir d'autres attributs, ou que si ceux-ci sont définis, alors ils ne seront pris en compte que si la fonction de similarité le fait.

Par exemple, une UAV abstraite pourra être un objet $u = \langle (Type, UAV) \rangle$ et associée à une fonction de similarité f ne prenant en compte que l'attribut *type*, pourra être comparée aux objets de O , rendant *vrai* pour toutes les UAV de O , et *faux* pour les EA et les EAA. Un EA $e = \langle (Type, EA) \rangle$ pourra être considéré comme équivalent à tous les EA si la même fonction de similarité lui est associée. Par contre, un objet $e' = \langle (Type, EA), (Nom, Chirac) \rangle$ associé à une fonction comparant exactement *Type* et *Nom* sera considéré équivalent à tous les EA inscriptions dans le flux de l'EAA $\langle EAA:Chirac \rangle$.

Un objet abstrait associé à une fonction de similarité qui lui soit adaptée est appelé *objet générique*. Comme le plus souvent la généralité portera sur l'attribut *Nom*, nous valuerons celui-ci par $*$, $*$ pouvant prendre l'ensemble des valeurs de $E_V(Nom)$. Par exemple l'EA générique paradigmatique sera $\langle * \rangle$, l'EAA sera $\langle EAA:* \rangle$, et en étendant à l'UAV, on aura $\langle UAV:* \rangle$.

Relations entre objets. Les relations entre objets génériques sont les mêmes que les relations que nous avons déjà définies. Un EA générique sera par exemple en relation d'annotation R_a avec une UAV générique. Nous introduisons cependant un ensemble de relations supplémentaires possibles entre éléments d'annotation génériques, qui sont toutes les relations temporelles possibles entre parties de flux, typiquement les treize relations de Allen (*before, after, touches, overlaps, etc.*).

Les relations temporelles entre EA génériques permettent de définir des contextes temporels en étendant la notion de chemins virtuels. Soit R_T l'ensemble des relations temporelles.

La figure 6.5 présente un ensemble de chemins potentiels. Gp_1 par exemple décrit un chemin virtuel entre un EA générique $\langle Document \rangle$ et une UAV totalement générique, le chemin passant par un EA $\langle StructDoc \rangle$ et un EA totalement générique. Le graphe Gp_2 décrit un chemin suivant une relation de spécialisation entre un EAA $\langle EAA:Homme politique français \rangle$ et un EAA totalement générique. Le graphe Gp_3 décrit un chemin contextuel potentiel sur le même schéma que Gp_1 . Gp_4 décrit un chemin contextuel potentiel avec une composante temporelle signifiant que l'EA générique $\langle Plan \rangle$ doit être en relation d'inclusion temporelle avec un EA générique ayant un attribut *Auteur* avec « M. Laroche-joubert » pour valeur. Cela signifie d'une part que l'UAV qu'annote $\langle Plan \rangle$ doit être incluse temporellement dans l'UAV qu'annote e , d'autre part que ce dernier EA générique est un EA associé à une fonction de similarité opérant sur les attributs *Type* et *Auteur*.

Associations de chemins La description d'un chemin virtuel peut être précisée. Plus particulièrement, il est possible de spécifier pour chacun de ses éléments un contexte, c'est à dire un autre chemin contextuel, et ce de façon récursive. Ce principe est illustré par le graphe Gp_5 figure 6.5 où l'EA $\langle Groupe \rangle$ faisant partie du chemin entre une UAV et un EA totalement générique est spécifié par sa relation contextuelle avec un EA $\langle Pano Gauche Droite \rangle$.

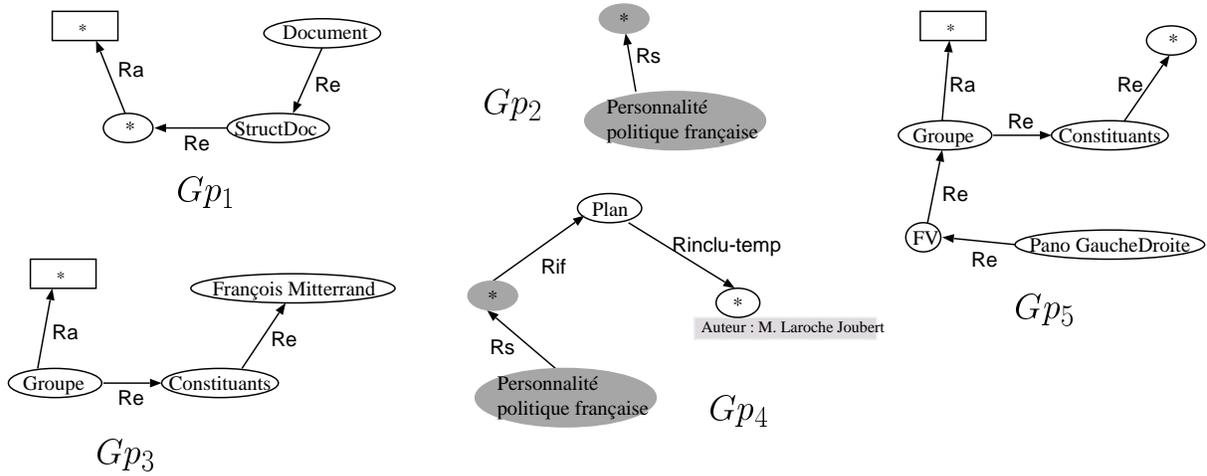


FIG. 6.5 – Quelques graphes potentiels

Il apparaît donc qu'un chemin virtuel est décrit par une ossature étiquetée de chemin, et que récursivement chacun des sommets du chemin peut être précisé par d'autres chemins, dont les sommets peuvent être à leur tour précisés, *etc.* Nous pouvons alors définir ce qu'est un graphe potentiel.

Graphe potentiel. Un *graphe potentiel* est un graphe permettant de spécifier abstraitement des contextes, des relations contextuelles entre sommets d'un graphe Strates-IA.

Soit un système Strates-IA $G_{SIA} = \langle O, R, G, C \rangle$ avec $G = \langle S, A, \mu, \nu \rangle$. Alors soit O_p l'ensemble des objets du graphe potentiel, définis suivant les mêmes contraintes d'attributs et de valeurs que les objets de O . Soit f_{O_p} l'application qui à tout objet $o_p \in O_p$ associe⁷ une fonction de similarité f_{o_p} permettant de le comparer aux objets de O . Le degré de généralité d'un sommet du graphe potentiel dépendra donc de la fonction de similarité qui lui est associée. On peut alors comparer tout objet de O_p (*i.e.* du graphe potentiel) à des objets de O (*i.e.* du graphe général des Strates-IA) : $\forall o_p \in O_p, \forall o \in O, f_{O_p}(o_p, o) = 1$ si $f_{o_p}(o_p, o) = 1$ (si o_p est jugé similaire à o suivant la fonction de similarité qui lui est associée), $= 0$ sinon.

Un graphe potentiel est défini comme $G_p = \langle O_p, f_{O_p}, R_p, G_{G_p}, C_p \rangle$ avec O_p ensemble des objets génériques du graphe potentiel, f_{O_p} associant à tout objet générique une fonction de similarité, $R_p = R \cup R_T$, $G_{G_p} = \langle S_p, A_p, \mu_p, \nu_p \rangle$ avec S_p ensemble des sommets du graphe potentiel, A_p ensemble de ses arcs, μ_p et ν_p associant respectivement aux sommets et aux arcs des étiquettes issues de O_p et de R_p . C_p est l'ensemble des contraintes sur les relations du graphe potentiel, qui sont presque équivalentes à celles contenues dans C à la différence que les relations entre EA génériques peuvent être soit élémentaires, soit appartenir à R_T .

Instancier les graphes potentiels pour réaliser les contextes. L'*instanciation* d'un graphe potentiel G_p dans un graphe Strates-IA G_{SIA} correspond à la recherche dans ce dernier des sous-graphes qui soient équivalents à G_p suivant les contraintes de structure (sommets et arcs) et les contraintes de similarité entre étiquettes de sommets et de relation. Une *instance* d'un graphe potentiel est un sous-graphe de SIA qui lui est équivalent suivant ces contraintes.

Plus formellement, si on a $G_{SIA} = \langle O, R, G, C \rangle$ avec $G = \langle S, A, \mu, \nu \rangle$ et

7. f_{O_p} permet de regrouper les diverses fonctions de similarité f_{o_p} associées aux objets de O_p .

$G_p = \langle O_p, f_{O_p}, R_p, G_{G_p}, C_p \rangle$ avec $G_{G_p} = \langle S_p, A_p, \mu_p, \nu_p \rangle$. Alors les instances de G_p dans G_{SIA} sont les sous-graphes $g_i = \langle S_i, A_i, \mu_i, \nu_i \rangle_{i \in [1, n]}$ de G (avec μ_i (resp. ν_i) la restriction de μ (resp. ν) à S_i (resp. A_i)) tels que pour tout $i \in [1, n]$ il existe une fonction bi-univoque $f_i : S_p \rightarrow S_i$ avec :

- $a = (x, y) \in A_p \Leftrightarrow a' = (f_i(x), f_i(y)) \in A_i$ et $\nu_p((x, y)) = \nu_p(f_i(x), f_i(y))$
- $\forall x \in S_p \quad f_{O_p}(\mu_p(x), \mu_i(f_i(x))) = \text{Vrai}$.

Par exemple les graphes potentiels de la figure 6.5 s'instancient dans le graphe défini par les figures 6.2 (pour les UAV et EA) et 6.3 (pour les EAA) de la manière suivante : Gp_1 a quatre instances, Gp_2 et Gp_4 en ont trois, Gp_3 une seule, et enfin Gp_5 permet de définir quatre instances.

On note $I(G_p, G_{SIA})$ l'ensemble des instances de G_p dans G_{SIA} .

Hypothèse de la correspondance initiale. Le problème de la recherche des instances d'un graphe potentiel dans un graphe Strates-IA se ramène à la recherche des isomorphismes de sous-graphe partiels entre G_p et G_{SIA} . Ce problème est dans le cas général NP-complet, c'est pourquoi nous ajoutons comme contrainte supplémentaire sur tout graphe potentiel qu'un au moins de ses sommets sera associé de façon unique à un des sommets de G_{SIA} . Ainsi, $\exists (x, y) \in S_p \times S$ tel que $\forall f_i$ associée à une instance de $I(G_p, G_{SIA})$, $f_i(x) = y$.

Ce sommet initial du graphe potentiel correspondra par exemple à un objet générique EAA nommé, qui pourra alors être associé sans ambiguïté à l'EAA de la base de connaissances correspondant. Ce sommet pourra également correspondre à une UAV totalement déterminée, ou bien à un EA précis.

La contrainte que nous posons est naturelle : les graphes potentiels servent à exprimer des contextes virtuels entre sommets d'un graphe Strates-IA, tandis que les instances expriment des contextes réels. L'instanciation, la recherche de contextes doit alors se faire à partir de sommets connus dont on cherchera le contexte. Cette recherche sera locale dans le graphe, c'est à dire que son ou ses points de départ seront connus, à la différence d'une approche qui chercherait des motifs de façon globale.

Parmi les graphes potentiels de la figure 6.5, seuls Gp_2 et Gp_4 sont valides au sens de la restriction que nous venons de poser. Les autres graphes potentiels peuvent devenir valides, pour peu qu'on spécifie explicitement certains de leurs sommets comme correspondant à des sommets précis de G , ou bien qu'on les complète par exemple au niveau des EA nommés en les mettant en relation avec des EAA génériques nommés, qui eux peuvent être mis en correspondance non ambiguë avec les EAA de la base de connaissances, comme indiqué sur la figure 6.6. La mise en place de sommets génériques EAA nommés dans les graphes potentiels correspond à une recherche d'annotations dans leur contexte, donc qui partent de ceux-ci.

Nous évoquerons plus en détail comment les graphes potentiels peuvent être manipulés et rendus valides. Retenons pour l'instant que pour tout graphe potentiel à instancier, on connaît sans calcul au moins une correspondance entre un sommet du graphe potentiel et un sommet du graphe Strates-IA.

6.2 Instanciation de graphes potentiels

Nous nous intéressons dans cette section à l'instanciation de graphes potentiels du point de vue algorithmique et de la théorie des graphes. Pour présenter la manière dont l'instanciation se réalise, nous allons simplifier notre problème en considérant des graphes étiquetés simplement (et

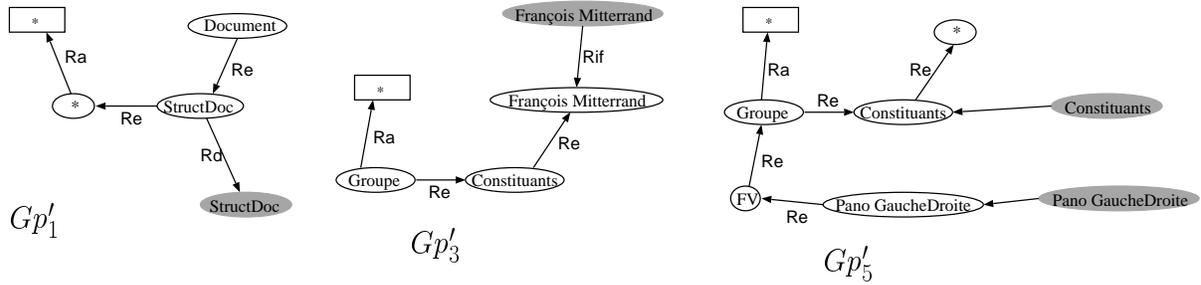


FIG. 6.6 – Rendre valide des graphes potentiels

non par des objets Strates-IA avec les contraintes associées). Par contre, les règles de vérification de cohérence utilisées seront décrites dans leur forme générale ; c'est à dire en se plaçant dans le cadre des graphes potentiels définis ci-dessus.

6.2.1 Isomorphismes de sous-graphes partiels

Un graphe G est défini par le couple $G = (X, A_X)$, où X désigne l'ensemble des sommets de G et $A_X \subseteq X \times X$ est l'ensemble des arcs de G . x_j est successeur de x_i si $(x_i, x_j) \in A_X$; l'ensemble des successeurs de x_i est noté $Succ(x_i)$. x_i est prédécesseur de x_j si $(x_i, x_j) \in A_X$; l'ensemble des prédécesseurs de x_i est noté $Pred(x_i)$.

Etant donné un graphe $G = (X, A_X)$, un sous-graphe de G est un graphe $G' = (X', A_{X'})$ tel que $X' \subseteq X$ et $X' \neq \emptyset$ et $A_{X'} = A_X \cap X' \times X'$. Un graphe partiel de G engendré par le sous-ensemble d'arcs $B \subseteq A_X$ est le graphe $G' = (X, B)$. Le sous-graphe partiel engendré par $X' \subseteq X$ et $B \subseteq A_{X'}$ est le graphe partiel de $G' = (X', A_{X'})$ engendré par $B \subseteq A_{X'}$.

Le problème de l'isomorphisme est un problème fondamental en informatique. On peut le formuler de la façon suivante : soient deux graphes $G_X = (X, A_X)$ et $G_Y = (Y, A_Y)$. G_X et G_Y sont *isomorphes* s'il existe une bijection $f : X \rightarrow Y$ telle que :

$$\forall x_i, x_j \in X : (x_i, x_j) \in A_X \Leftrightarrow (f(x_i), f(x_j)) \in A_Y$$

Une autre variante de ce problème appelée *problème de l'isomorphisme de sous-graphes partiels* connu aussi sous le nom d'*inclusion de graphes* consiste à rechercher une relation entre G_X et un sous-graphe partiel de G_Y . L'étude de ce problème répond à la question suivante : G_Y contient-il un sous graphe partiel isomorphe à G_X ? c'est à dire existe-t-il un sous-ensemble Y' de sommets de Y et un sous-ensemble $A_{Y'}$ d'arcs tels que $|Y'| = |X|$, $|A_{Y'}| = |A_X|$ et il existe une bijection $f : X \rightarrow Y'$ telle que :

$$\forall x_i, x_j \in X : (x_i, x_j) \in A_X \Leftrightarrow (f(x_i), f(x_j)) \in A_{Y'}$$

6.2.2 Méthodes de recherche d'isomorphismes de sous-graphes

Le problème de l'isomorphisme de sous-graphes se rencontre dans de nombreux domaines au nombre desquels :

- la chimie organique [Rouvray et Balaban, 1979] pour la reconnaissance de molécules qui possèdent une sous-structure particulière ;

- le raisonnement à partir de cas pour la recherche de cas dans une base de cas [Poole, 1993, Bradtke et Lehnert, 1988] ;
- les réseaux sémantiques et les grammaires de graphes [Ehrig, 1992] ;
- l'apprentissage, où les informations symboliques sont représentées par des graphes de concepts et l'on désire par exemple savoir si un concept est plus général qu'un autre [Fisher, 1990, Cook et Holder, 1994] ;
- la reconnaissance de formes et la vision par ordinateur, où l'étude de ce problème a été motivée par le désir de retrouver des objets dans des images [Cho et Kim, 1992, Wong *et al.*, 1989, Wong, 1992].

Les méthodes développées pour résoudre le problème de l'isomorphisme de sous-graphes peuvent être classées dans l'une des deux approches suivantes.

L'approche basée sur la *recherche de cliques*⁸ *maximales* dans un graphe appelé graphe d'association décrivant toutes les correspondances possibles entre sommets de G_X et de G_Y ⁹. Dans cette première approche, nous pouvons citer les travaux de [Falkenhainer *et al.*, 1989, Myaeng et Lopez-Lopez, 1997, Horaud et Skordas, 1989] (cités dans [Messmer, 1996]). Les algorithmes issus de cette approche sont très généraux, puisque le problème est considéré de façon globale mais peu efficaces dans la pratique de par leur complexité et ne s'appliquent qu'à des graphes de petite taille.

L'approche dite de *relaxation*, basée sur l'utilisation combinée de méthodes de filtrage et d'algorithmes d'énumération de l'espace des solutions. [Kitchen et Rosenfeld, 1979, Kim et Kak, 1991] décrivent des algorithmes illustrant cette seconde approche. Ces algorithmes présentent l'avantage de pouvoir s'implanter facilement sur des architectures parallèles et d'utiliser ainsi la puissance des machines multiprocesseurs. La proposition de [Ullman, 1976] (que nous avons implanté) est similaire à cette approche : on construit un tableau de dimension $N_X \times N_Y$, où N_X (resp. N_Y) désigne le nombre de sommet du graphe G_X (resp. G_Y) en exécutant une procédure récursive ayant ce tableau — modifié à chaque itération — pour paramètre. La complexité est alors rédhibitoire pour les graphes de grande taille.

Plus récemment, un algorithme de recherche d'isomorphismes également basé sur la deuxième approche a été proposé par [Cordella *et al.*, 1998], appliqué à des graphes générés aléatoirement et comparé aux algorithmes présentés par [Ullman, 1976, Messmer, 1996].

De manière informelle¹⁰, cet algorithme est basé sur le principe suivant : pour rechercher toutes les solutions, l'algorithme construit un arbre où chaque sommet représente un isomorphisme partiel et un arc correspond à l'ajout d'une paire de sommets conduisant à un autre isomorphisme partiel de cardinalité supérieure. La racine de cet arbre est l'ensemble vide et la profondeur de l'arbre en cours de construction est directement fonction du cardinal de l'isomorphisme partiel associé. Pendant la construction des solutions, si l'ajout d'une paire de sommets à un isomorphisme partiel (à un sommet de l'arbre) conduit à un état incohérent, alors l'algorithme abandonne toute exploration depuis cet état ; c'est à dire que le chemin s'interrompt sur cet état

8. Une clique d'un graphe G est un sous-ensemble S de ses sommets tel que le sous-graphe de G engendré par S soit complet.

9. Plus précisément, ce graphe d'association a pour sommets des couples de correspondances (x, y) entre sommets de G_X et de G_Y , et ses arêtes expriment une cohérence locale de relations entre correspondances deux à deux considérées.

10. Nous décrivons cet algorithme dans la suite.

Graine : Une graine est une correspondance $[x, y]$ valide donnée par l'utilisateur (correspondance de départ). Dans le modèle des Strates-IA, une graine associe un sommet du graphe potentiel à un sommet connu du graphe Strates-IA, et correspond à l'hypothèse de la correspondance initiale.

On suppose que les correspondances initiales de la recherche d'isomorphisme de sous-graphe partiel présenté figure 6.7 sont les couples $[a, 2]$ et $[d, 9]$ (indiqués en gras), il y a donc deux graines.

État : Soient deux graphes $G_X = (X, A_X)$ et $G_Y = (Y, A_Y)$. Un état e du processus d'appariement des deux graphes est défini par un ensemble de correspondances valides $V(e)$

$$V(e) = \{[x, y] \in X \times Y \mid x \text{ est en correspondance avec } y\} \quad (6.1)$$

et est appelé *isomorphisme partiel associé à l'état e* . On notera par $V_X(e)$ (resp. $V_Y(e)$) la projection de $V(e)$ sur X (resp. Y) :

- $V_X(e) = \{x \in X \mid \exists y \in Y \mid [x, y] \in V(e)\}$
- $V_Y(e) = \{y \in Y \mid \exists x \in X \mid [x, y] \in V(e)\}$

Un état composé uniquement d'une graine est appelé *état initial*. L'ensemble des graines présentes dans un état e est noté $R(e)$ (gRaines) :

$$R(e) = \{[x, y] / x \text{ soit en correspondance initiale avec } y\} \quad (6.2)$$

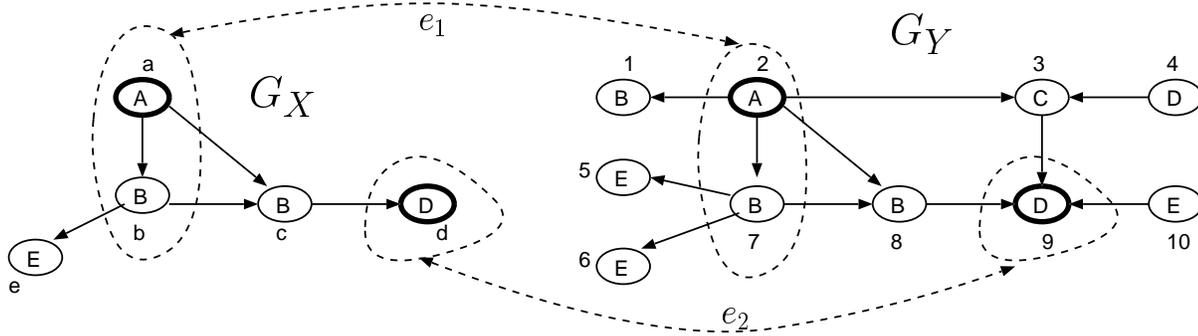


FIG. 6.8 – Deux états e_1 et e_2 dans le processus de recherche d'isomorphisme partiel de G_X dans G_Y

Sur l'exemple de la figure 6.8, on peut considérer deux états e_1 et e_2 . e_2 , composé de la graine $[d, 9]$ est un état initial. e_1 est composé de deux correspondances G_Y , $V(e_1) = \{[a, 2], [b, 7]\}$ est l'isomorphisme partiel qui lui est associé. Les projections de e_1 sur X et Y sont respectivement $V_X(e_1) = \{a, b\}$ et $V_Y(e_1) = \{2, 7\}$. On a $R(e_1) = \{[a, 2]\}$ et $R(e_2) = \{[d, 9]\}$. Remarquons que pour un état e correspondant à l'isomorphisme final (5 correspondances), $R(e) = \{[a, 2], [d, 9]\}$.

Extension d'un état : $I_X(e)$ (resp. $I_Y(e)$) dénote l'ensemble des sommets incidents intérieurement aux sommets de $V_X(e)$ (resp. $V_Y(e)$). $E_X(e)$ (resp. $E_Y(e)$) dénote l'ensemble des sommets incidents extérieurement aux sommets de $V_X(e)$ (resp. $V_Y(e)$). Plus formellement, on a :

- $I_X(e) = \{x \in X - V_X(e) \mid \exists x' \in V_X(e) \mid (x, x') \in A_X\}$

- $E_X(e) = \{x \in X - V_X(e) \mid \exists x' \in V_X(e) \mid (x', x) \in A_X\}$
- $I_Y(e)$ et $E_Y(e)$ peuvent être définis de façon similaire.

On désigne par $I(e)$ (resp. $E(e)$) l'ensemble des paires de sommets incidents intérieurement (resp. extérieurement) aux sommets de $V(e)$ et défini par :

$$E(e) = E_X(e) \times E_Y(e) \quad (\text{resp. } I(e) = I_X(e) \times I_Y(e)).$$

On note par $C(e)$ l'ensemble des correspondances candidates à l'extension de l'isomorphisme partiel associé à l'état e . La définition de l'ensemble $C(e)$ est réalisée par la séquence suivante (sachant qu'on aura toujours $I(e) \neq \{\}$, puisque les graphes considérés dans le cas de notre étude sont connexes) :

$$\text{Si } E(e) \neq \{\} \text{ alors } C(e) \leftarrow E(e) \text{ , sinon } C(e) \leftarrow I(e). \quad (6.3)$$

Par exemple pour les états de la figure 6.8, on a $C(e_1) = \{[c, 8][e, 5][e, 6]\}$ et $C(e_2) = \{[c, 8], [c, 3]\}$, si on prend en compte les critères de cohérence définis plus loin.

Soit $C_X(e) = \{x \in X \mid \exists y \in Y \mid [x, y] \in C(e)\}$ la projection de l'ensemble $C(e)$ sur X . Pour tout sommet $x \in C_X(e)$ on appelle extension de $C(e)$ suivant x la suite des correspondances de $C(e)$ ayant le sommet x comme origine et notée par $C(e \rightarrow x)$. On note par $|C(e \rightarrow x)|$ la dimension (nombre d'éléments) de la suite $C(e \rightarrow x)$ et par $\delta(e)$ la dimension de la plus petite extension associée à l'état e :

$$\delta(e) = \min\{|C(e \rightarrow x)| \text{ avec } x \in C_X(e)\} \quad (6.4)$$

Pour les états de la figure 6.8, l'extension de e_1 suivant c est $C(e_1 \rightarrow c) = \{[c, 8]\}$, et son extension suivant e est $C(e_1 \rightarrow e) = \{[e, 5][e, 6]\}$. De la même manière, $C(e_2 \rightarrow c) = \{[c, 8][c, 3]\}$ ¹¹. Les dimensions d'extension sont $\delta(e_1) = 1$ et $\delta(e_2) = 2$.

Critères de cohérence structurelle : le contrôle de cohérence d'une correspondance $[x, y]$ avec un isomorphisme partiel $V(e)$ consiste à vérifier quelques propriétés structurelles telles que les propriétés de connectivité et d'incidence.

Propriété de connectivité : cette propriété exprime le fait que x peut être mis en correspondance avec y si chacun des voisins de x qui est dans $V_X(e)$ est déjà mis en correspondance avec un voisin de y et inversement. D'où les règles qui vérifient le respect de ces contraintes d'adjacence :

- $(\forall x' \in V_X(e) \cap \text{pred}_X(x) \Rightarrow \exists y' \in \text{pred}_Y(y) \mid (x', y') \in V(e))$
- $(\forall y' \in V_Y(e) \cap \text{pred}_Y(y) \Rightarrow \exists x' \in \text{pred}_X(x) \mid (x', y') \in V(e))$
- $(\forall x' \in V_X(e) \cap \text{succ}_X(x) \Rightarrow \exists y' \in \text{succ}_Y(y) \mid (x', y') \in V(e))$
- $(\forall y' \in V_Y(e) \cap \text{succ}_Y(y) \Rightarrow \exists x' \in \text{succ}_X(x) \mid (x', y') \in V(e))$

11. On pourrait penser que $[c, 10]$ fait partie de l'extension de e_2 suivant e , ce qui serait vrai si on n'appliquait un critère de cohérence d'ordre 2 prenant en compte le nombre de relations entrantes et sortantes pour le sommet e , et vérifiant qu'il y en a suffisamment pour les candidats à la correspondance avec e . Dans le cas présent, les sommets 8 et 3 ont bien deux arcs entrants et un sortant, ce qui n'est pas le cas du sommet 10.

Propriété d'incidence : les deux règles suivantes permettent de vérifier cette propriété. La première règle indique si le nombre de successeurs de x incidents intérieurement à $V(e)$ est égal au nombre de successeurs de y incidents intérieurement à $V(e)$. La deuxième règle traite les prédécesseurs, la troisième et la quatrième sont consacrées au sommets incidents extérieurement.

- $(|(succ_X(x) \cap I_X(e)| = |(succ_Y(y) \cap I_Y(e))|)$
- $(|(pred_X(x) \cap I_X(e)| = |(pred_Y(y) \cap I_Y(e))|)$
- $(|(pred_X(x) \cap E_X(e)| = |(pred_Y(y) \cap E_Y(e))|)$
- $(|(succ_X(x) \cap E_X(e)| = |(succ_Y(y) \cap E_Y(e))|)$

Propriété d'incidence d'ordre 2: Cette propriété indique si le nombre de successeurs (resp. prédécesseurs) de x hors de l'état e est bien égal au nombre de successeurs (resp. prédécesseurs) de y :

- $(|pred_X(x) \cap (X - E_X(e) - V_X(e))| \leq |pred_Y(y) \cap (Y - E_Y(e) - V_Y(e))|)$
- $(|succ_X(x) \cap (X - E_X(e) - V_X(e))| \leq |succ_Y(y) \cap (Y - E_Y(e) - V_Y(e))|)$

La vérification de la propriété d'incidence d'ordre 2 n'est pas nécessaire, elle permet simplement de couper l'espace de recherche en éliminant les mauvaises extensions d'état (*cf.* la note 11).

Critères de cohérence sémantique : la cohérence sémantique d'une correspondance $[x, y] \in C_X(e) \times C_Y(e)$ s'exprime par le calcul d'une similarité entre étiquettes de sommets réalisé par la fonction $f_{SymEtiqNoeud}$ et étiquettes d'arcs (suivant $f_{SymEtiqRel}$) :

- cohérence des sommets : $f_{SymEtiqNoeud}(x, y) = Vrai$;
- cohérence des arcs :
 $\forall (x', y') \mid x' \in V_X(e) \cap pred_X(x) \wedge y' \in V_Y(e) \cap pred_Y(y) \Rightarrow f_{SymEtiqRel}((x', x), (y', y)) = Vrai$
 $\forall (x', y') \mid x' \in V_X(e) \cap succ_X(x) \wedge y' \in V_Y(e) \cap succ_Y(y) \Rightarrow f_{SymEtiqRel}((x', x), (y', y)) = Vrai$

Nous avons explicitement déjà pris en compte les critères de cohérence structurels et sémantiques en présentant les extensions des états e_1 et e_2 . Par exemple $[c, 3]$ ne fait pas partie de l'extension de e_1 puisque les étiquettes de sommets c et 3 sont différentes (critère de cohérence sémantique). Egalement, $[c, 1]$ ne fait pas partie de l'extension de e_1 : bien que les étiquettes de e et de 1 soient similaires, le nombre des successeurs de 1 est nul, n'est pas au moins égal au nombre de successeurs e , donc la propriété d'incidence d'ordre 2 permet d'éliminer $[c, 1]$.

Etape : Une étape P_k est définie par l'ensemble des états $\{e_i(k); i \in [1, N_k]\}$ atteints à l'itération k du processus d'appariement. L'étape initiale notée P_0 est définie par l'ensemble des états initiaux $\{e_i(0); i \in [1, N_0]\}$ spécifiés par l'utilisateur.

Sur l'exemple de la figure 6.8, une étape peut être $P_1 = \{e_1, e_2\}$, tandis que l'étape initiale contenait les deux états initiaux correspondant aux deux graines.

Structure d'un état : Chaque état e de P_k est décrit par les cinq éléments suivants :

- l'isomorphisme partiel $V(e)$ (eq. 6.1),
- l'extension $C(e)$ qui contient les extensions suivant les sommets de G_X extérieurs à l'état (eq. 6.3),
- le nombre $\delta(e)$ (eq. 6.4),
- l'ensemble $R(e)$ des graines présentes dans e (eq. 6.2),
- un indicateur d'extensibilité booléen : $extensible(e)$, mis à la valeur *Vrai* à la création de l'état.

L'indicateur d'extensibilité d'un état permet de spécifier si celui-ci peut être étendu en un nouvel état au cours du processus d'appariement, ou s'il ne peut que *fusionner* avec d'autres états.

Fusion d'états : à une étape donnée, pour savoir si l'extension d'un état e par une correspondance $[x, y]$ crée un isomorphisme partiel connexe à un autre isomorphisme partiel, il faut savoir s'il existe déjà dans l'étape un état e' contenant la correspondance $[x, y]$ et dont l'ensemble des graines est disjoint de l'ensemble des graines présentes dans e . Si c'est le cas, l'extension de e par $[x, y]$ provoque la fusion des deux états (les deux états e et e' sont dits connexes). Le nouvel état résultant de la fusion de ces deux états connexes est obtenu en fusionnant leur ensemble de correspondances $V(e'') \leftarrow V(e') \cup V(e)$ et leur ensemble de graines $R(e'') \leftarrow R(e') \cup R(e)$. L'état e devient non extensible — gelé¹² — en positionnant son attribut d'extensibilité à *False*.

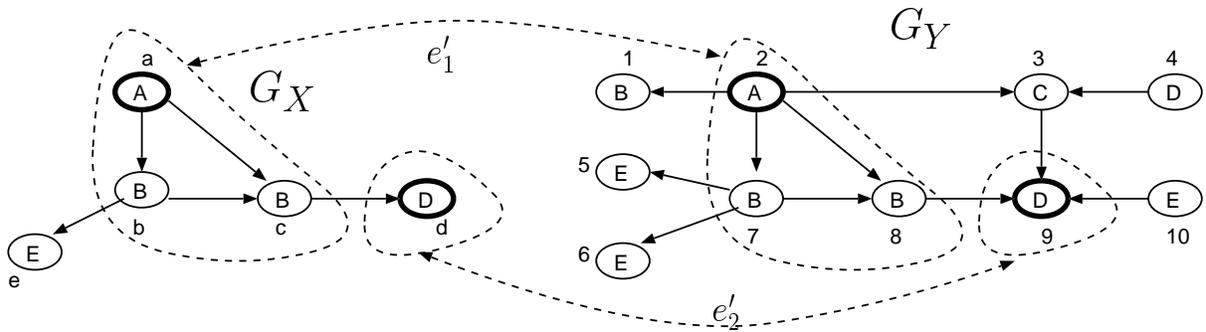
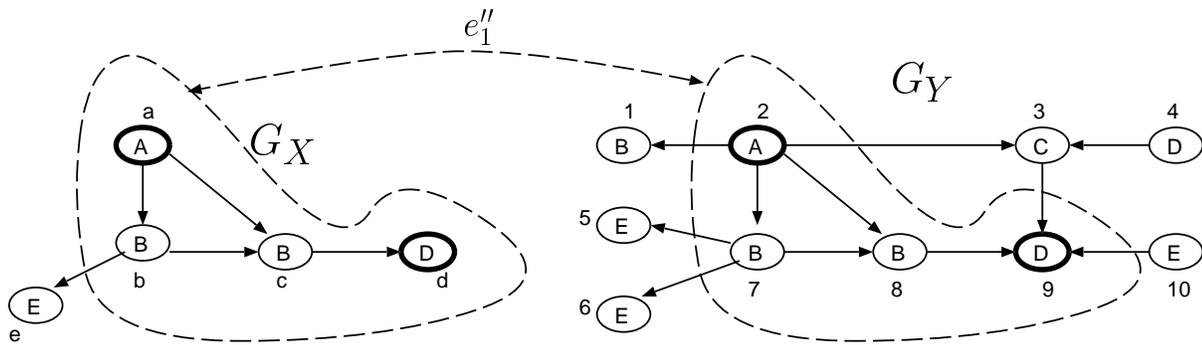


FIG. 6.9 – Etape après l'extension de e_1 suivant c

Par exemple, à supposer figure 6.9 que l'état e_1 de l'étape P_1 ait été étendu suivant c pour donner un nouvel état e'_1 (dans une étape P_2), tandis que e'_2 est la copie exacte de e_2 . On a alors $P_2 = \{e'_1, e'_2\}$, et l'état e'_1 est caractérisé par l'isomorphisme partiel $V(e'_1) = \{[a, 1], [b, 7], [c, 8]\}$; son extension $C(e'_1) = \{C(e'_1 \rightarrow e), C(e'_1 \rightarrow d)\}$ avec $C(e'_1 \rightarrow e) = \{[e, 5], [e, 6]\}$ ainsi que $C(e'_1 \rightarrow d) = \{[d, 9]\}$; $\delta(e'_1) = 1$ et $R_{e'_1} = \{[a, 1]\}$.

Comme $[d, 9]$ est contenue dans $V(e'_2)$, alors l'extension de e'_1 suivant d mène à la fusion de e'_1 et e'_2 , pour donner un nouvel état e''_1 qui est caractérisé par l'isomorphisme partiel $V(e''_1) =$

12. Le gel d'un état consiste à interdire toute extension future de celui-ci autre que résultant d'un autre état fusionnant avec lui. Le fait que l'état fusionne signifie qu'il transmet toutes ses propriétés d'extension au nouvel état fusionné, et abandonne donc celles-ci pour lui-même.


 FIG. 6.10 – Fusion des états e'_1 et e'_2 en e''_1

$\{[a, 1], [b, 7], [c, 8], [d, 9]\}$; son extension $C(e''_1) = \{C(e''_1 \rightarrow e)\}$ avec $C(e''_1 \rightarrow e) = \{[e, 5], [e, 6]\}$, $\delta(e''_1) = 2$ et $R_{e''_1} = \{[a, 1], [d, 9]\}$.

6.2.4 Algorithme de multi-propagation

Les différentes notions que nous venons de définir nous permettent de présenter dans sa totalité l'algorithme de multi-propagation, dont les exemples déjà présentés ont sans doute permis de comprendre les grandes lignes.

L'étape 0 est initialisée avec les états liés aux graines de départ. Le déroulement général de l'algorithme consiste alors à passer d'une étape k à une étape $k+1$ en étendant un certain nombre d'états de l'étape k suivant leurs relations d'extension, et en conservant les autres. Quand un état atteint la taille du graphe qu'on cherche à instancier, cet état représente un isomorphisme solution (son isomorphisme partiel est un isomorphisme solution).

On parle d'algorithme de *propagation* au sens où l'extension d'un état se fait par propagation le long des arcs des graphes. L'algorithme de [Cordella *et al.*, 1998] choisit un état quelconque de départ, une seule graine donc, et étend tous les états d'une étape de façon maximale, suivant toutes leurs possibilités d'extensions. Au bout d'autant d'étapes qu'il y a de sommets dans le graphe à instancier, cet algorithme de propagation aboutit à toutes les solutions, contenues dans la dernière étape.

Dès le moment où l'on dispose de plusieurs correspondances de départ (graines) à partir desquelles on peut propager les recherches de solutions, il y a lieu de parler de *multi-propagation* de la recherche d'isomorphisme. Dans le cas général, si on part d'une étape k , on retrouvera dans l'étape $k+1$:

- les états de l'étape k qui n'ont pas été touchés par l'extension ;
- les états de l'étape k qui ont servi à une fusion (donc ont simplement été marqués non extensibles) ;
- les nouveaux états résultant des extensions simples (l'isomorphisme partiel $V(e)$ d'un état $e \in P_k$ est étendu d'une correspondance en $V(e) \cup \{[x, y]\}$) ou bien des extensions avec fusion.

A la différence de l'approche présentée dans [Cordella *et al.*, 1998] qui étend à chaque étape *tous* les états suivant *toutes* les extensions possibles, dans notre approche on n'étend qu'*un* seul

état à chaque passage d'une étape à l'autre, ce qui permet d'engendrer moins d'états dans l'arbre des solutions (au pire le même nombre).

L'heuristique de choix de l'état à étendre à chaque étape est le critère principal de notre algorithme. A chaque itération, notre algorithme choisit un état $e \in P_k$ à étendre vérifiant le critère suivant :

$$\delta(e) = \inf\{\delta(a); a \in P_k\}$$

Ce choix de l'extension minimale revient à minimiser le nombre de nouveaux états générés dans le passage d'une étape k à une étape $k+1$, donc à progresser de la façon la plus économique possible vers la solution.

Algorithme 1 *IsoGP*($G_X = (X, A_X), G_Y = (Y, A_Y), P_0$)

```

k ← 0
répéter
  Pk+1 ← {}
  Soit e ∈ Pk extensible / δ(e) = inf{δ(a); a ∈ Pk}
  Pk+1 ← Pk - e
  Choisir un sommet x ∈ GX tel que |V(e → x)| = δ(e)
  pour tout [x, y] correspondance de V(e → x) faire
    P' = ExtensionFusion(Pk+1, [x, y], e)
    pour tout e' ∈ P' faire
      si |e'| = |X| alors
        Retourner e' comme solution
      sinon
        Pk+1 ← Pk+1 + e'
      fin si
    fin pour
  fin pour
  k ← k + 1
jusqu'à (∀e ∈ Pk ⇒ Extensible(e) = Faux)

```

L'algorithme 1 présente la procédure principale de recherche de sous-graphes partiels, qui prend en entrée les graphes G_X et G_Y et l'étape initiale P_0 comprenant les n états initiaux (n graines). A chaque étape k , le choix de l'état e à étendre fait, on reporte tous les états présents dans P_k dans P_{k+1} à l'exception de l'état e . Pour chaque correspondance $[x, y]$ de $V(e \rightarrow x)$, on appelle la fonction *ExtensionFusion*($P_{k+1}, [x, y], e$) (cf. algorithme 2) qui calcule l'extension de l'état e suivant $[x, y]$.

La taille de chacun des états retournés par *Extension_Fusion* est ensuite testée. Si celle-ci correspond à la taille du graphe potentiel, alors l'état décrit un isomorphisme solution, et cette solution est rendue immédiatement, sinon on ajoute l'état à l'ensemble des états de P_{k+1} .

L'algorithme s'achève lorsqu'il ne reste plus que des états gelés dans l'étape courante, c'est à dire lorsque tous les isomorphismes de sous-graphes partiels solutions ont été trouvés.

La procédure *ExtensionFusion* permet de retourner :

- soit l'ensemble des états obtenus par fusion de e avec des états de P_{k+1} connexes (ie. ayant $[x, y]$ dans leurs correspondances et toutes leurs graines différentes de celles de e) ;
- soit un état résultant d'une simple extension de e par la correspondance $[x, y]$;

- soit $\{\}$ si les états nouveaux générés n'ont pas d'extension possible et ne peuvent donc mener à des solutions finales.

Algorithme 2 *ExtensionFusion*($P, [x, y], e$)

```

NouveauxEtats  $\leftarrow \{\}$ 
pour tout  $t \in P$  | extensible( $t$ ) = Vrai faire
  si  $R(t) \cap R(e) = \{\}$  et  $[x, y] \in t$  alors
     $e' \leftarrow t \cup e$ ;  $R(e') \leftarrow R(t) \cup R(e)$ 
    extensible( $t$ )  $\leftarrow$  Faux
    NouveauxEtats  $\leftarrow e'$ 
  finsi
fin pour
si NouveauxEtats =  $\{\}$  alors
   $e' \leftarrow e + [x, y]$ ;
  NouveauxEtats  $\leftarrow$  NouveauxEtats +  $e'$ 
finsi
pour tout  $t \in$  NouveauxEtats faire
  si  $|t| \neq |X|$  ( $t$  n'est pas solution) alors
     $C(e') \leftarrow$  CalculExtension( $e'$ )
    si  $C(e') = \{\}$  alors
      Enlever  $t$  de NouveauxEtats
    finsi
  finsi
fin pour
Retourner NouveauxEtats

```

CalculExtension calcule l'ensemble des extensions possibles pour un état, ainsi que son extension minimale.

6.2.5 Un exemple complet

On considère les deux graphes orientés présentés figure 6.11 $G_X = (X, A_X)$ avec $X = \{a, b, c, d, e\}$ et $G_Y = (Y, A_Y)$ avec $Y = \{1, 2, 3, 4, 5, 6, 7, 8, 9, \}$. Les arcs de G_X et G_Y ne sont pas étiquetés. On cherche les sous-graphes partiels de G_Y isomorphes au graphe d'entrée G_X à partir des initiaux e_{01} et e_{02} avec $V(e_{01}) = \{[a, 3]\}$ et $V(e_{02}) = \{[b, 4]\}$.

L'exécution de l'algorithme sur notre exemple se fait suivant les étapes représentées dans le graphe états-transitions de la figure 6.12. Sur les arcs sont indiqués les sommets suivant lesquels des extensions sont réalisées. Un arc en pointillés indique une simple recopie d'un état d'une étape k dans l'étape $k + 1$. Un arc avec une croix schématise une extension non viable $\delta = 0$. Un sommet entouré d'un rectangle en pointillé indique la non extensibilité de l'état associé. Les solutions finales trouvées sont les S_i .

Le passage de l'étape P_0 à l'étape P_1 se fait par une extension simple de l'état e_{01} suivant c . Le passage de l'étape P_1 à l'étape P_2 voit l'extension de l'état e_{12} suivant c et sa fusion subséquente avec e_{11} , de laquelle découle l'état e_{22} tandis que e_{11} (et ses diverses recopies) sont gelés. Le déroulement suivant ne pose pas de problème particulier¹³, et l'algorithme s'arrête car l'unique état présent dans l'étape P_5 n'est pas extensible. On remarquera également que l'ensemble de

13. Pour peut qu'on n'oublie pas la vérification de la cohérence d'ordre 2.

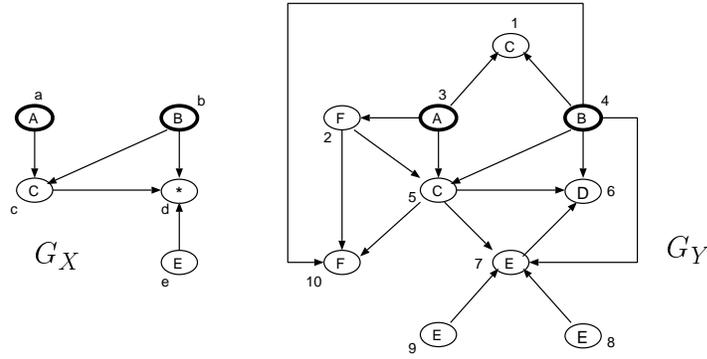


FIG. 6.11 – G_X a trois instances dans G_Y

toutes les graines des états de cette étape ne couvre plus l'ensemble des graines de départ, c'est à dire que toutes les potentialités offertes par les différents états de départ ont été exploitées au maximum. A l'issue de cette exécution, on obtient trois isomorphismes de sous graphes partiels de G_X dans G_Y :

- $S_1 = \{[a, 3], [c, 5], [b, 4], [d, 6], [e, 7]\}$
- $S_2 = \{[a, 3], [c, 5], [b, 4], [d, 7], [e, 8]\}$
- $S_3 = \{[a, 3], [c, 5], [b, 4], [d, 7], [e, 9]\}$

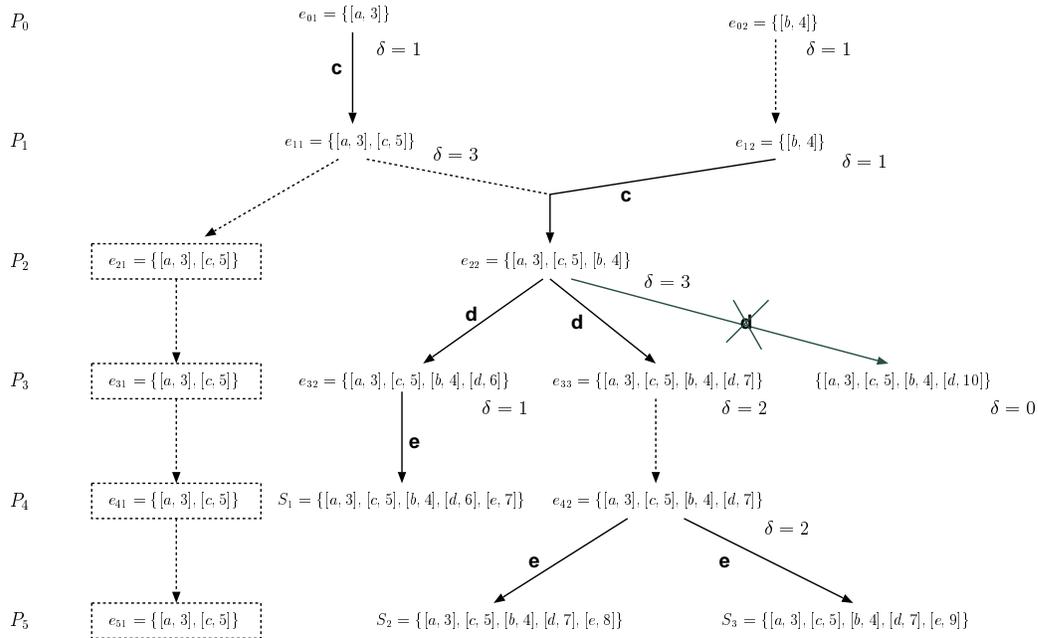


FIG. 6.12 – Graphe états-transition de déroulement de l'algorithme de multi-propagation

6.2.6 Expérimentation et discussion

Les notions d'étape, d'état et d'extension d'état ainsi que les critères de cohérence syntaxique et sémantique étaient déjà présentés dans [Cordella *et al.*, 1998]. Dans cet algorithme, on part d'une correspondance choisie au hasard, avant d'étendre maximale dans toutes les directions, ce qui correspond à un arbre état-transition de profondeur $|X|$ et de largeur maximale, dont les solutions sont les feuilles, éventuellement doublonnées. Cette approche correspond à une exploration en largeur des solutions par un algorithme de propagation simple. Nous avons également mis au point et implanté un tel algorithme sous sa forme récursive, où à chaque retour de récursion les isomorphismes partiels doublonnés étaient éliminés. La proposition de [Cordella *et al.*, 1998] nous a alors permis d'appréhender le problème sous une forme non récursive, que nous avons modifiée pour permettre une multi-propagation minimisant les états générés et éliminant naturellement les doublons, c'est à dire guidée par une heuristique déterminant à chaque étape la direction de propagation à privilégier. La notion de graine est à cet égard fondamentale, puisqu'il s'agit d'étendre plusieurs isomorphismes partiels en même temps, de telle sorte qu'ils se « rejoignent » pour donner des solutions.

La phase d'expérimentation répond aux tests validant notre approche et permettant d'analyser les performances générales de l'algorithme mis en œuvre. Pour cela, nous avons utilisé la bibliothèque LEDA [Mehlhorn et Näher, 1992, Mehlhorn *et al.*, 1998] développée à l'Institut Max Planck à Saarbrücken, qui offre une gamme très complète de fonctionnalités relatives à la représentation et aux traitements de graphes. Un composant générique graphe est fourni avec les opérateurs permettant de réaliser les traitements de base les plus courants. Nous avons évalué notre algorithme et l'algorithme présenté dans [Cordella *et al.*, 1998] sur une série de graphes généraux aléatoires répondant aux contraintes des Strates-IA (UAV, EA, EAA, relations). Les graphes potentiels composés de 10 sommets en moyenne, sont construits par extraction de sous-graphes partiels des graphes généraux ainsi générés, et possèdent au moins un sommet UAV ou EAA ayant une correspondance unique).

Un test élémentaire sur un graphe général $G_Y = (Y, A_Y)$ et un graphe potentiel $G_X = (X, A_X)$ générés aléatoirement, consiste à prendre des mesures (temps d'exécution, nombre d'états générés, *etc.*) en exécutant :

- une fois notre algorithme avec n états initiaux (n graines) ;
 - n fois l'algorithme [Cordella *et al.*, 1998] en fixant à chaque exécution un nouvel état initial.
- A partir de cette série d'exécutions, une moyenne des mesures visées est calculée.

Il est ainsi possible de comparer l'algorithme de multi-propagation prenant n graines en compte en même temps à une moyenne des n lancements de l'algorithme de propagation.

Le tableau 6.1 résume les résultats des séries de tests réalisées en donnant la taille en nombre de sommets du graphe général généré, le nombre de solutions trouvées, le numéro d'état correspondant à la première solution, le nombre total d'états générés par notre algorithme, et enfin le nombre total d'états générés par l'algorithme présenté dans [Cordella *et al.*, 1998].

Ces premiers résultats sont importants car ils confortent la stratégie choisie pour réduire l'espace des solutions explorés par notre algorithme. On montre ainsi qu'en pilotant le choix des états à l'extension par quelques paramètres simples, il est possible d'obtenir des solutions finales plus rapidement que l'algorithme présenté dans [Cordella *et al.*, 1998] et des espaces de solutions

Taille G_Y	Nombre de solutions	Etat première solution	Nombre total d'état MP	Nombre total d'états Cordella et al. [1998]
130	48	23	118	5803
250	17	17	68	2696
500	26	15	36	706
1500	14	19	40	2001
2000	20	14	28	615
8000	43	27	93	4875
15000	37	27	70	4527
30000	51	27	105	4597
80000	84	19	132	3837

TAB. 6.1 – Comparaison, en fonction de la taille du graphe d'accueil du nombre d'états générés par les algorithmes de multi-propagation et de propagation

beaucoup plus réduits. Par exemple, en moyenne la première solution est fournie par notre algorithme avant d'atteindre le tiers de la taille totale de l'arbre d'états ; alors qu'avec l'algorithme de propagation, on est contraint d'attendre la construction totale de l'arbre d'états pour en extraire la première solution. De plus, on voit clairement que la réduction du nombre total d'états générés est très sensible.

Ces résultats doivent être nuancés dans le cas général de la recherche d'isomorphismes de sous-graphes partiels. Ils impliquent en effet qu'au moins une correspondance de départ soit connue, ce qui n'est pas le cas général. D'autre part, les tests ont été menés sur des graphes liés aux Strates-IA et non sur des graphes quelconques.

Remarques. Il apparaît tout d'abord que nous avons fait le choix d'étendre un seul état par étape, suivant une unique direction, et l'heuristique de choix considère qu'une bonne direction est une direction qui générera le moins d'états possibles dans la nouvelle étape. Cette heuristique, qui donne de bons résultats peut être discutée. Par exemple, en gardant l'extension d'un unique état, il serait possible de considérer une heuristique variable qui, à partir d'un certaine étape du déroulement de l'algorithme stipulerait de favoriser les états considérés comme étant les plus près d'une solution (*i.e.* possédant un nombre de graines proche du nombre de graines total).

Ensuite, l'algorithme que nous proposons s'intègre dans le domaine de la modélisation de documents audiovisuels, mais reste général. Ceci provient d'une volonté de considérer l'ensemble du système comme un unique graphe, et de s'attacher à décrire les opérations sur ce système comme étant liées à des manipulations de contextes exprimés par des graphes potentiels. Deux conséquences en découlent : en premier lieu, la généralité de l'algorithme permet d'envisager de l'appliquer à d'autres bases de données s'apparentant à des graphes. Ainsi on peut envisager d'en étudier l'application à des bases de données semi-structurées à partir non plus d'une unique graine (racine d'un arbre de données), mais à partir de plusieurs graines différentes. En deuxième lieu, les particularités des Strates-IA devraient permettre d'envisager de nouvelles heuristiques et améliorations de performance par rapport aux principes généraux présentés dans cette section. Celles-ci pourraient s'appuyer sur les types de sommets et de relations inter-sommets de l'application. Par exemple, la complexité liée aux fonctions de calcul des similarités entre sommets peut

être prise en compte. Dans le cadre de notre application, il peut ainsi être utile d'étendre un état dans une direction impliquant un calcul de similarité basé sur des attributs issus de traitements d'image, s'il y a peu de candidats à la comparaison. Au contraire, si l'extension d'un état implique des calculs coûteux sur les attributs d'un grand nombre de sommets, il peut y avoir lieu de retarder ceux-ci, et d'explorer des branches nécessitant des calculs de similarité plus simples. De la même manière, si une relation liée à l'état est temporelle, l'extension de l'état peut être plus longue à calculer¹⁴. On peut donc imaginer de ne pas favoriser les extensions menant à des calculs de relations temporelles.

Il apparaît en conséquence que l'heuristique de choix de l'état à étendre est paramétrable et qu'adapter celle-ci à l'application permet d'améliorer encore les performances.

Remarquons enfin que le fait que l'algorithme fournisse les solutions dès qu'elles sont arrivent, et non en fin de calcul, est un point important en recherche d'information : il est possible pour l'utilisateur de commencer à traiter les résultats sans que la recherche soit totalement terminée. L'algorithme est alors « anytime » ce qui correspond à une caractéristique importante en aide à la décision.

6.3 Outils de manipulation fondés sur les graphes potentiels

Nous avons dans la section 6.1 introduit la notion de contexte dans les Strates-IA, ainsi que les graphes potentiels permettant d'instrumenter la manipulation et la recherche de contextes. L'instanciation d'un graphe potentiel se ramène à une recherche d'isomorphisme de sous-graphes partiels entre le graphe potentiel et le graphe des Strates-IA, et nous en avons proposé section 6.2 un algorithme.

Il convient maintenant de s'intéresser à la manière dont les graphes potentiels vont être utilisés et manipulés par l'utilisateur comme par la machine, et quels outils sont disponibles pour les tâches de mise en place et d'exploitation d'un système Strates-IA. Ces outils devront être principalement adaptés à l'utilisateur. Notre approche d'annotation comme écriture sur le flux nécessite en effet de gérer la manière dont il va être possible de lire les annotations, c'est à dire d'envisager une tâche d'exploitation comme régissant une contextualisation.

6.3.1 Graphes potentiels caractérisés

Caractériser pour utiliser

Un graphe potentiel permet d'exprimer un contexte virtuel « brut » en mettant en relation des objets génériques des Strates-IA. Par l'instanciation, il réalise le contexte abstraitement décrit en contextes concrets entre objets du graphe Strates-IA. Afin de pouvoir exploiter ces contextes concrets, il faut pouvoir les caractériser, c'est à dire spécifier opératoirement quels sommets du graphe potentiel (et donc des graphes-instances) seront utiles pour la tâche dans laquelle le graphe potentiel est utilisé.

Un *graphe potentiel caractérisé* est donc un graphe potentiel dont certains sommets d'intérêt, jugés utiles, sont nommés, et qui permettent de l'exploiter. Par exemple, un graphe potentiel exprimant un chemin contextuel devra posséder *a minima* une origine et une extrémité qui permettront de le désigner.

14. Il faudra alors faire appel à un calcul de voisinage temporel des EA. Ce genre de calcul dépend bien évidemment de la taille des flux et du nombre d'UAV qui y ont été définies.

Soit G_P un graphe potentiel, alors le caractériser revient à nommer certains de ses sommets (chaque nom étant unique). On notera un graphe potentiel caractérisé par

$$G_{PC} = (G_P, \{x_i\})$$

où :

- G_P est un graphe potentiel ;
- x_i est un nom (une *étiquette de sommet d'intérêt*) associé à un sommet de G_P ¹⁵.

Par exemple, le graphe potentiel caractérisé G_{PC_1} de la figure 6.13 est un graphe potentiel dont deux points sont caractérisés par des étiquettes de sommets d'intérêt *orig* et *extr* correspondant aux sommets 1 et 4. L'étiquette de sommet d'intérêt *PanoSurGroupe* est associée au sommet 1 du graphe potentiel constituant G_{PC_2} , *ConstituantGroupe* au sommet 6, *orig* au sommet 4 et *extr* au sommet 6.

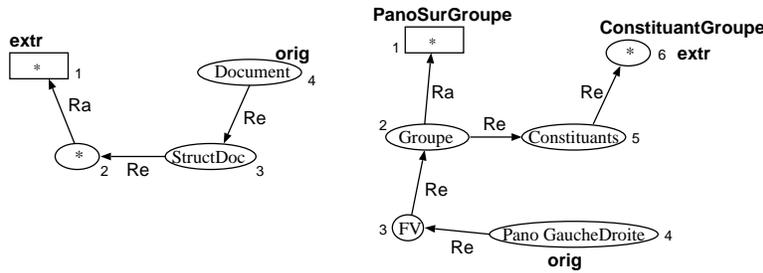


FIG. 6.13 – Premiers exemples de graphes potentiels caractérisés

Manipulations de graphes potentiels caractérisés

Avant de donner des exemples plus précis de graphes potentiels caractérisés, nous en présentons rapidement quelques opérations de manipulation. Ces opérations seront directement adaptées de celles liées aux graphes potentiels.

Construction et instanciation de graphes potentiels. On munit les graphes potentiels d'opérateurs de construction :

- *création* d'un graphe potentiel à un seul sommet et étiquettage par un objet générique ;
- *mise en relation* de graphes potentiels par une relation et étiquettage de celle-ci conformément aux contraintes de relation des Strates-IA ;
- *étiquettage d'un sommet* permettant de modifier un objet générique associé à un sommet ;
- *jonction* des deux graphes potentiels par identification de sous-graphes communs ;
- *suppression de relation* entre deux sommets.

¹⁵. Il est également possible de noter $G_{PC} = (G_P, \mu_C)$ avec μ_C fonction associant éventuellement à chaque sommet de G_P une étiquette de sommet d'intérêt.

Une dernière opération de construction de graphe potentiel est plus complexe, puisqu'il s'agit alors de *potentialiser un sous-graphe* du graphe général Strates-IA G_{SIA} . A partir d'un sous-graphe connexe g quelconque de G_{SIA} , on met alors en place un graphe isomorphe, dont les étiquettes sont des objets génériques obtenus par abstraction des objets étiquettes de g (*i.e* en supprimant ou en changeant des attributs, et en associant à chaque objet générique une fonction de similarité idoine). Cette opération se révèle d'intérêt pour prendre en compte des annotations qui ont été mises en place afin de s'en inspirer pour construire ses propres requêtes.

Aux opérations de modification qu'on vient de présenter, on ajoute un opérateur d'instanciation qui permet de calculer les sous-graphes du graphe général G_{SIA} instances du graphe potentiel. Ce dernier doit être valide, c'est à dire posséder un moins une étiquette de sommet qui puisse être mise sans ambiguïté en correspondance avec un sommet du graphe général. Dans le cas contraire, l'instanciation ne pourra s'effectuer.

Construction et instanciation de graphe potentiels caractérisés. Un graphe potentiel caractérisé étant principalement constitué d'un graphe potentiel (auquel on ajoute des étiquettes de sommets d'intérêt), il peut être construit uniquement avec les opérations présentées plus haut.

Cependant, l'accès au graphe potentiel dont est composé un graphe potentiel caractérisé peut également se faire à partir des sommets caractérisés, particulièrement les opérations d'étiquetage de sommets ou de mises en relation. Par exemple, si un graphe potentiel caractérisé possède un sommet caractérisé par *ConstituantGroupe*, alors cette étiquette de sommet d'intérêt permet d'accéder à l'objet générique correspondant directement, par exemple pour le préciser en $\langle Nelson Mandela \rangle$. En d'autres termes, les sommets caractérisés et leurs noms sont des primitives de manipulation, des index de sommets de graphes potentiels qui se révéleront utiles au cours des tâches d'utilisation des graphes potentiels.

L'instanciation d'un graphe potentiel caractérisé G_{PC} consiste simplement à instancier le graphe potentiel G_P qu'il contient.

De la même manière que précédemment, l'accès aux instances peut être contrôlé par les sommets caractérisés. Par exemple, si après l'instanciation du graphe G_{PC_2} de la figure 6.13, on obtient trois instances, alors l'accès aux sommets intéressants de celle-ci — en l'occurrence ici les UAV annotées par un groupe qui soit l'objet d'un panoramique gauche-droite — se fait en utilisant le sommet caractérisé par *PanoSurGroupe*.

Nous présentons dans la suite quelques exemples de graphes potentiels caractérisés, et illustrons les opérations que nous avons présentées. Les exemples que nous présentons ne sont que deux types très utiles de graphes potentiels caractérisés. On ne perdra pas de vue que ceux-ci peuvent être créés à volonté dès que le besoin s'en fait sentir.

Exemples de graphes potentiels caractérisés

Relations contextuelles potentielles. Une *relation contextuelle potentielle* (RCP) est un graphe potentiel caractérisé par deux sommets. Le premier est nommé *origine (orig)*, le second est nommé *extrémité (extr)* de la relation contextuelle potentielle.

Si rcp est une relation contextuelle potentielle, on a $rcp = (gp, \{orig, extr\})$.

Suivant les types des objets étiquetant les sommets origine et extrémité, les relations contextuelles potentielles peuvent être interprétées différemment. Par exemple, en se basant sur ce que nous avons présenté dans la partie 6.1.2. :

- si l'origine est UAV et l'extrémité EA ou EAA, alors on s'intéresse au contexte de l'UAV, particulièrement aux EA qui pourraient l'annoter contextuellement, soit en provenance du

flux, soit de la base de connaissances. Un exemple de telle RCP est le graphe potentiel caractérisé G_{PC_1} de la figure 6.13.

- si l’origine est EA, alors on peut s’intéresser particulièrement aux RCP à extrémité EA ou EAA, qui fournissent des explications sur l’utilisation de l’EA, lesquelles sont *en contexte* (dans le flux) ou bien *hors contexte* (dans la base de connaissances). Par exemple, le graphe potentiel caractérisé G_{PC_2} de la figure 6.13 permet de s’intéresser au contexte dans le flux des EA $\langle PanoGaucheDroite \rangle$ (*orig*), contexte concernant le cas où c’est un groupe qui est filmé, auquel cas les EA correspondant à *extr* sont les constituants des groupes filmés ;
- si l’origine est EAA, et l’extrémité EA ou UAV, cela peut correspondre à un contexte d’explication d’utilisation ; tandis que si l’extrémité est un EAA, on s’intéresse à la place de l’EAA origine dans la base de connaissances (on en trouvera des exemples avec les filtres de désignation que nous présenterons plus bas).

Bien évidemment, l’interprétation des contextes et leurs mises en place dépendent de la tâche en cours. Par exemple, supposons que nous désirions annoter contextuellement une UAV u , c’est à dire définir un contexte désignant une autre UAV u' , et plus particulièrement certains des éléments d’annotation annotant u' , qu’on pourra alors considérer annoter contextuellement u .

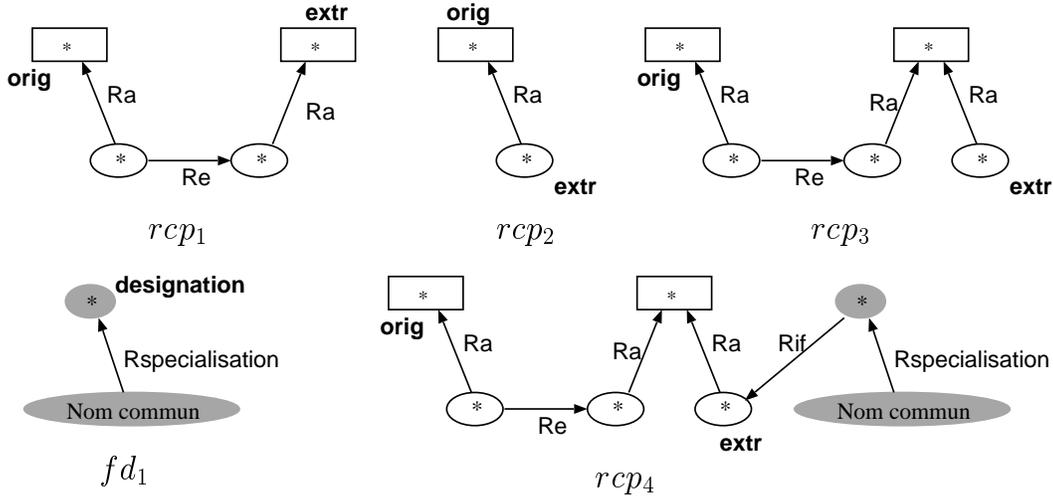


FIG. 6.14 – Exemples de graphes potentiels caractérisés

Définissons alors une RCP à origine UAV et à extrémité UAV telle que rcp_1 figure 6.14, qui exprime que l’UAV faisant partie du contexte de l’UAV origine devra être en relation quelconque (non temporelle cependant) avec elle. Si nous appliquons cette RCP à une UAV connue, par exemple en créant une UAV rcp'_1 (figure 6.15) équivalente à rcp_1 à la différence que l’objet générique correspondant à *orig* est objet UAV ayant un attribut identifiant l’UAV 8 du flux *Fic232* de la figure¹⁶ 6.15, alors l’instanciation de rcp'_1 donnera deux instances. Les extrémités de ces instances, c’est à dire les sommets du graphe Strates-IA G_{SIA} en correspondance avec le sommet *extr* de rcp'_1 sont alors les UAV u_1 (UAV 1 du flux contenu dans le fichier *fic342*, voir figure 6.15) et u_2 (UAV 13 du flux principal représenté).

Supposons maintenant que nous désirions considérer comme annotations contextuelles les EA annotant ces deux UAV (et ne participant pas à la relation entre elles). Nous pouvons opérer

16. Qui n’est autre qu’une version légèrement étendue du graphe de la figure 6.4.

7 permet alors de récupérer comme extrémités les EA $\langle \text{Epouse} \rangle$ et $\langle \text{Poignée de main} \rangle$ ¹⁷ (voir figure 6.15).

Filtres de désignation. Un *filtre de désignation* est un graphe potentiel caractérisé à sommets étiquetés par des objets de type EAA, caractérisé par un sommet d'intérêt appelé *point de désignation* (*désignation*). Par exemple figure 6.14, fd_1 est un filtre de désignation.

Si fd est un filtre de désignation, alors $fd = (gp, \{des\})$.

Instancier un filtre de désignation permet de récupérer un ensemble d'EAA, qui sont dans ses instances les sommets en correspondance avec le sommet nommé *désignation*. La figure 6.16 présente deux autres filtres de désignation : fd_2 permet de désigner uniquement l'EAA $\langle \text{EAA:Zoom avant} \rangle$ dans la base de connaissances de la figure 6.3, page 132, tandis que fd_3 désigne les EAA $\langle \text{EAA:Nelson Mandela} \rangle$ et $\langle \text{EAA:Winnie Mandela} \rangle$.

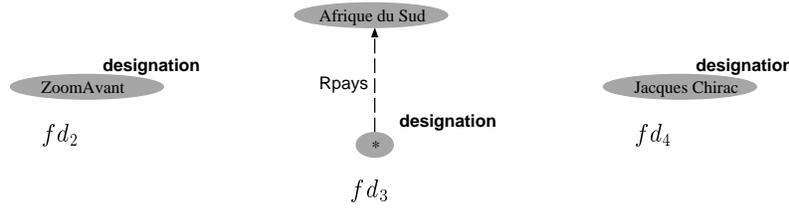


FIG. 6.16 – Deux filtres de désignation

6.3.2 Dimensions d'analyse

Définition

Nous avons dans le chapitre 5 présenté une dimension d'analyse comme un regroupement des analyses qui permettent de repérer des objets d'intérêt de même type.

Dans le cadre des Strates-IA les objets d'intérêt sont réifiés en annotations représentées par des éléments d'annotation, eux mêmes abstraits en éléments d'annotation abstraits ; une dimension d'analyse permettra toujours de rassembler un certain nombre d'EAA. Nous définissons donc une dimension d'analyse comme l'association d'un nom et d'une *méthode de désignation d'EAA*.

Une *méthode de désignation* (MD) permet de désigner un ensemble d'EAA de la base de connaissances. Elle est donc d'abord un ensemble de filtres de désignation tels que nous les avons définis en 6.3.1.

Par exemple, la méthode de désignation $md_1 = \{fd_3, fd_4\}$ permet de désigner les EAA $\langle \text{EAA:Nelson Mandela} \rangle$, $\langle \text{EAA:Winnie Mandela} \rangle$ et $\langle \text{EAA:Jacques Chirac} \rangle$.

On associe aux méthodes de désignation des opérations permettant d'ajouter, d'enlever des filtres de désignation, mais il est également possible de les manipuler à leur propre niveau. Ainsi on pourra définir une méthode de désignation comme l'association d'autres méthodes de désignation.

Une méthode de désignation md est alors un ensemble de méthodes de désignation et/ou d'ensembles de filtres de désignation :

$$md \leftarrow \{fd_i\} \mid md + \{fd_j\} \mid \{md_k\}$$

17. Les exemples ici présentés ne sont bien sûr qu'indicatifs et nous ne prétendons pas qu'ils correspondent à une pratique avérée.

On aura par exemple $md_2 = \{fd_2, md_1\}$.

Etendre une méthode de désignation consiste à remplacer récursivement les méthodes de désignation qui la composent par les méthodes de désignation ou filtres de désignation, et ainsi de suite jusqu'à ce que la MD ne soit plus composée que de filtres de désignation.

On aura alors $Ext(md) = \{fd_i\}$, par exemple $md'_2 = Ext(md_2) = \{fd_2, fd_3, fd_4\}$.

Résoudre une méthode de désignation consiste à calculer l'ensemble des éléments d'annotation abstraits que celle-ci désigne.

On aura alors $Res(md) = \{eaa_i\}$, par exemple $Res(md'_2) = Res(md_2) = \{ \langle EAA:Nelson Mandela \rangle, \langle EAA:Winnie Mandela \rangle, \langle EAA:Jacques Chirac \rangle, \langle EAA:Zoom Avant \rangle \}$.

Remarquons que les opérations telles que ajout, retrait ou intersection de MD opèrent sur les définitions de MD, et non sur leurs extensions, c'est à dire sur celles-ci telles qu'elles sont définies lorsqu'on les utilise. Ainsi par exemple, si $m_a = \{m_1, m_2, m_3\}$, $m_b = \{m_1, m_4\}$ et $m_1 = \{m_2, m_4\}$ alors on aura $m_a \cap m_b = \{m_1\}$.

Nous définissons plus précisément une dimension d'analyse d est l'association d'un nom, d'un texte, et d'un méthode de description¹⁸.

$$d = \langle Nom, Description_texte, md \rangle$$

Une dimension d'analyse sera *résolue* en l'ensemble des EAA qu'elle permet de désigner. Ainsi, si $d = \langle Nom, Description_texte, md \rangle$, alors $Res(d) = Res(md)$.

Avec la dimension d'analyse $da_1 = \langle Politiquesdivers, \text{« sommets France/Afrique du Sud 1995-1999 »}, md_1 \rangle$, on aura $Res(da_1) = \{ \langle EAA:Nelson Mandela \rangle, \langle EAA:Winnie Mandela \rangle, \langle EAA:Jacques Chirac \rangle \}$.

Contraintes sur les valeurs d'attributs d'EAA contenus dans une dimension d'analyse

Une dimension d'analyse permet de décrire des EAA à utiliser pour l'annotation, c'est à dire que la seule prescription porte sur le nom de l'EA qui sera mis en place, toute l'information sur les autres attributs de celui-ci étant portée par l'EAA.

Il est possible de spécifier un peu plus les dimensions d'analyse de telle sorte qu'elles imposent des valeurs sur certains attributs des EA qu'elles permettront de générer. On peut ainsi leur associer des contraintes sur les attributs et les valeurs d'attributs des EA qu'elles permettront de mettre en place. Une condition nécessaire à la définition de ces contraintes est que les attributs soient présents en tant que définitions dans tous les EAA désignés par celle-ci.

Par exemple, $daa = \langle DA : Emissions : \{ \langle EAA:Journal Télévisé \rangle, \langle EAA:Jeu \rangle \} : Date = 05081999 \rangle$ est une dimension associée à des contraintes spécifiant que l'annotation pourra conduire à mettre en place des EA $\langle Journal Télévisé \rangle$ et/ou $\langle EAA:Jeu \rangle$ avec une valeur d'attribut *Date* fixée à 05081999.

Discussion

La notion de dimension d'analyse est fondamentale dans les Strates-IA. En effet, permettant de désigner un ensemble d'EAA, une dimension d'analyse représente une vue de la base de connaissances correspondant à une *visée de description* du flux audiovisuel. Elle est donc en

18. Eventuellement, on pourra avoir une dimension d'analyse « informelle » $\langle Nom, Texte, \{ \} \rangle$ simplement décrite par un nom et une description textuelle de la façon de la construire, par exemple « tous les EAA liés à la politique ». La résolution de la DA fait alors appel à l'utilisateur à un moment ou à un autre.

quelque sorte un contexte dans lequel l'utilisateur se place afin d'appréhender le flux et d'y projeter ses attentes¹⁹. La dimension d'analyse peut alors être pensée comme technique, à la fois comme guide de repérage des objets d'intérêt et de caractérisation de ceux-ci. Une dimension d'analyse est donc liée à la tâche et à la volonté de description en cours, et en représente le pendant formel.

A ce titre une dimension d'analyse doit être fortement évolutive, c'est à dire qu'il doit être possible de la compléter, de la restreindre, de la fusionner avec une autre pour l'adapter à la visée de description. Les quelques opérations de manipulation que nous avons présentées correspondent à cette nécessaire évolutivité.

Certaines dimensions d'analyse sont construites manuellement par l'utilisateur qui sélectionne les EAA convenables, elles sont donc composées de filtres de désignation à portée réduite, ne désignant qu'un EAA pour chaque filtre (comme par exemple fd_2 ou fd_3 figure 6.16). D'autres dimensions d'analyse prennent avantage de l'organisation de la base de connaissances, c'est à dire des relations entre EAA par l'intermédiaire de filtres de désignation à portées plus larges (par exemple fd_1 ou fd_3). Ainsi une dimension d'analyse ayant pour méthode de désignation le filtre de désignation fd_1 permet de désigner tous les « Noms communs » de la base de connaissances en utilisant l'organisation hiérarchique de cette dernière. Il va alors de soi que les choix d'organisation de la base de connaissances ont une influence sur les dimensions d'analyse qui seront utilisées pour l'exploitation de cette dernière. Par exemple, organiser hiérarchiquement la base de connaissances revient, au moins au niveau des feuilles, à définir implicitement des filtres de désignation efficaces et les dimensions d'analyse qu'ils permettent de construire. Regrouper les « hommes politiques » utiles à l'annotation sous un même père hiérarchique revient à mettre en place implicitement une dimension d'analyse liée aux hommes politiques. Cependant, au delà de ces dimensions d'analyse « naturelles » à la base de connaissances (puisqu'implicites), toute construction ou manipulation de DA est possible et nécessaire pour exploiter la base de connaissances d'une manière adaptée à la tâche.

6.3.3 Schémas de description

Nous avons jusqu'ici défini les graphes potentiels caractérisés comme des graphes potentiels manipulables, et les dimensions d'analyse comme des marques de la visée d'analyse d'un utilisateur annotant un flux. Les dimensions d'analyse permettent de guider un utilisateur selon une certaine description en ne lui proposant que des ensembles limités et nommés d'EAA²⁰ à inscrire dans le flux.

Si nous sommes donc pour l'instant capables de fournir un cadre précis pour l'annotation suivant certaines dimensions d'analyse, nous ne pouvons ni spécifier ni prescrire une description structurée d'un document ou d'une partie de document. Cela peut pourtant s'avérer nécessaire, par exemple lorsque l'annotation n'est pas personnelle, mais est réalisée dans une institution suivant les canons correspondant à la volonté de description de celle-ci. Il importe alors qu'au minimum certaines méthodes d'écriture comme de lecture sur le flux, donc de contextualisation soient définies.

Ces méthodes seront exprimées suivant des schémas de description.

19. Comme pour un texte, la lecture et l'interprétation d'un flux audiovisuel sont des phénomènes temporels qui impliquent une constitution du sens dans la durée, la dimension d'analyse est alors structurante de la pensée, de la visée, de la cohérence interprétative analytique de l'utilisateur (pour paraphraser [Terray, 1996]).

20. Bref un vocabulaire limité.

Définition

Considérée du point de vue pratique de la description, une dimension d'analyse est un ensemble de termes utilisables pour l'annotation de façon quelconque. Pour exprimer une structuration de la pratique de description, il y a lieu de la considérer comme *ensemble de termes jouant le même rôle par rapport à d'autres ensembles de termes*, c'est à dire qu'il faut exprimer des relations entre les dimensions d'analyse dont les EA extraits seront mis en relation.

Par exemple, pour exprimer le fait que la description devra mettre en relation des EA issus de la dimension d'analyse $da_1 = \langle DA:Homme\ politique \rangle$ avec des termes issus de $da_2 = \langle DA:Prise\ de\ parole\ publique: \{ \langle Discours \rangle, \langle Conférence\ de\ presse \rangle, \langle Interview \rangle \} \rangle$, on mettra da_1 et da_2 en relation dans un schéma de description.

Un *schéma de description basique* est un graphe orienté étiqueté dont les sommets sont doublement étiquetés par noms de dimensions d'analyse et des booléens.

On a alors $SD = \langle S, A, \mu_1, \mu_2, \nu \rangle$ avec S ensemble des sommets du graphe, A ensemble des arcs, μ_1 associant à chaque sommet une dimension d'analyse, μ_2 associant à chaque sommet un booléen, et ν associant à chaque arc une étiquette de relation.

Pour un sommet quelconque $s \in S$, l'interprétation de $\mu_2(s)$ est la suivante :

- si $\mu_2(s) = Faux$, alors la dimension d'analyse spécifique de mettre en place un EA de façon quelconque ;
- si $\mu_2(s) = Vrai$, alors la mise en place de l'EA doit obligatoirement s'accompagner de la création d'une nouvelle unité audiovisuelle dont il sera EA primitif.

Soient s_1 et s_2 deux sommets de S tels que $(s_1, s_2) \in A$ alors les étiquettes de relations possibles $\nu((s_1, s_2))$ et leurs interprétations sont les suivantes :

- $\nu((s_1, s_2))$ peut être une des relations élémentaires $R_e[m, n]$, $m \leq n$ permettant de spécifier le nombre de relations élémentaires à mettre en place ($R_e[0, 1]$, $R_e[1, 1]$ et $R_e[1, n]$ seront les plus utilisées). Par exemple, si ea_1 est issu de la dimension d'analyse $\mu_1(s_1)$, et que $\nu((s_1, s_2)) = R_e[1, n]$, alors il y a lieu de le mettre en relation élémentaire R_e avec un ou plusieurs ea_{2i} issus de $\mu_1(s_2)$.
- $\nu((s_1, s_2))$ peut appartenir à R_T l'ensemble des relations temporelles. Alors il faut que l'UAV annotée par ea_1 soit dans la même relation temporelle avec l'UAV qu'annote ea_2 .
- Si $\nu((s_1, s_2)) = R_a$ la relation d'annotation, alors ea_2 doit annoter la même unité audiovisuelle que ea_1 .

La figure 6.17 présente un schéma de description sd_1 (indépendant des exemples que nous avons jusqu'ici présentés). Les sommets s_i du graphe entourés d'un rectangle signifient que $\mu_2(s_i) = Vrai$ donc que pour les EA extraits de leurs dimensions d'analyse étiquettes doivent être EA primitifs.

Les dimensions d'analyse utilisées dans le schéma de description sd_1 sont les suivantes :

- $\langle DA:Document: \{ \langle Document \rangle \} \rangle$

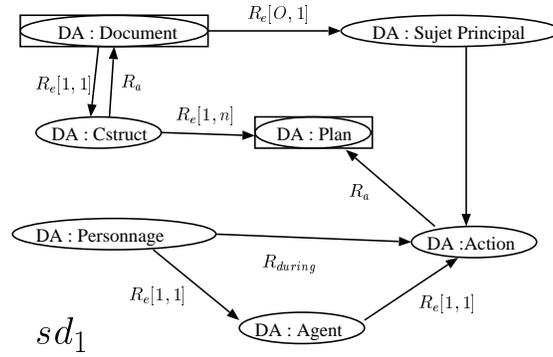


FIG. 6.17 – Un schéma de description de base

- $\langle DA:Plan:\{\langle Plan \rangle\} \rangle$
- $\langle DA:Cstruct:\{\langle Cstruct \rangle\} \rangle$
- $\langle DA:Agent:\{\langle Agent \rangle\} \rangle$
- $\langle DA:Sujet\ principal:\{\langle Sujet\ principal \rangle\} \rangle$
- $\langle DA:Personnage:\{\langle Le\ Corbeau \rangle, \langle Le\ Renard \rangle, \langle Le\ Loup \rangle, \langle L'agneau \rangle, \langle Le\ Ch\^ene \rangle, \langle Le\ Roseau \rangle\} \rangle$
- $\langle DA:Action:\{\langle Poss\^eder \rangle, \langle Discourir \rangle, \langle Flatter \rangle, \langle Ecouter \rangle, \langle Manger \rangle, \langle Boire \rangle\} \rangle$

Nous pr\^esentons dans la suite trois \^etapes d'une utilisation de sd_1 pour la description d'un flux correspondant \^a un document (par exemple un dessin anim\^e) illustrant une fable bien connue. A chaque \^etape, les parties utilis\^ees du sch\^ema de description sont soulign\^ees en gras.

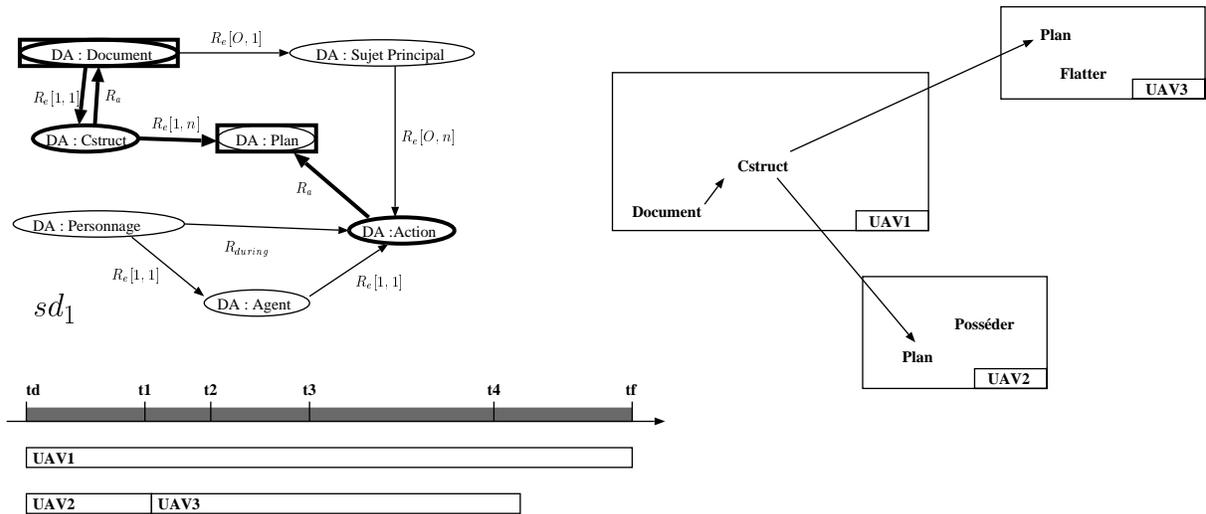


FIG. 6.18 – Utilisation de sd_1 : d\^ebut d'annotation, mise en place de la structure et des « actions »

La figure 6.18 correspond au d\^ebut de l'annotation, c'est \^a dire \^a la mise en place de la structure (relations entre $\langle Document \rangle$ et $\langle Plan \rangle$) dans le flux, conduisant \^a la cr\^eation de trois

unités audiovisuelles, comme spécifié dans le schéma de description. De la même manière, à chaque UAV définie primitivement par $\langle Plan \rangle$ est associé un EA issue de la dimension d'analyse $\langle DA:Action \rangle$.

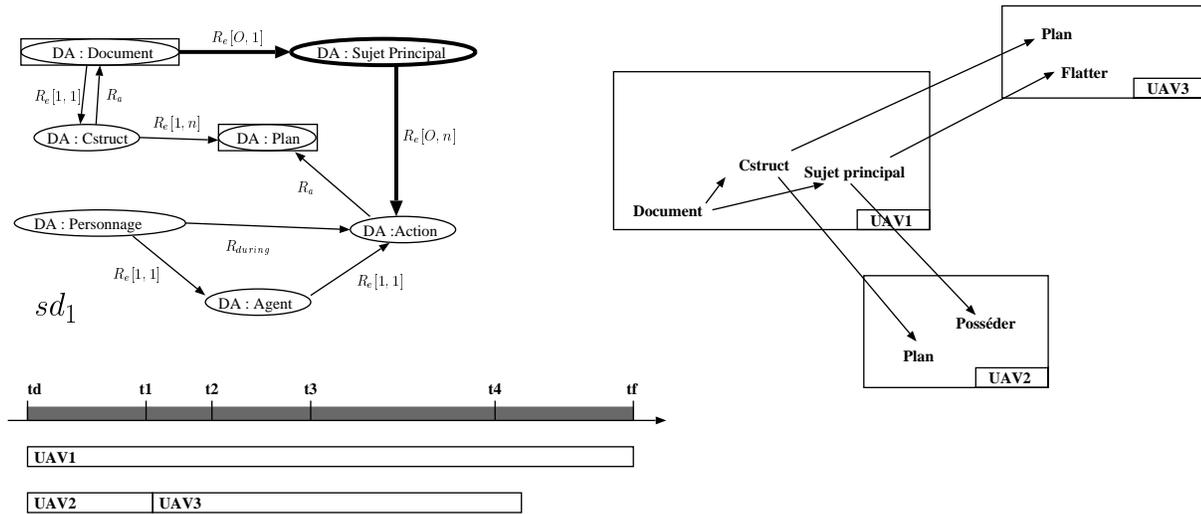


FIG. 6.19 – Utilisation de sd_1 : relations entre $\langle Document \rangle$ et les « actions » qui le composent

La figure 6.19 correspond simplement à l'utilisation de la mise en relation facultative permettant de lier $\langle Document \rangle$ à des sujets principaux, qui sont issus de $\langle DA:Action \rangle$.

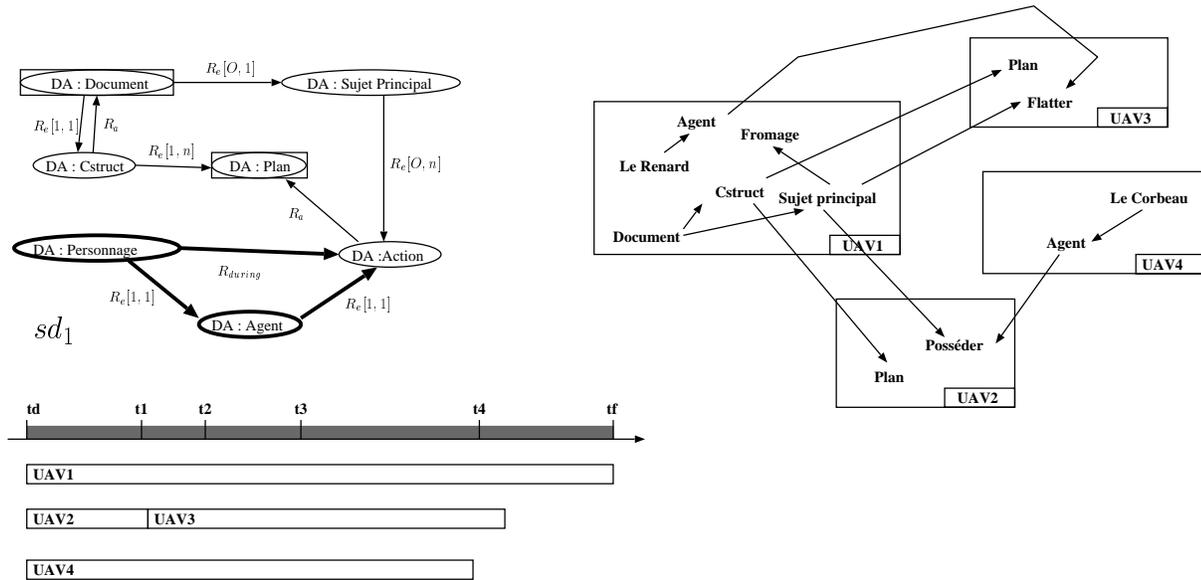


FIG. 6.20 – Utilisation de sd_1 : mise en place des « personnages » et complétion libre de l'annotation

La figure 6.20 correspond à l'utilisation de la relation temporelle R_{During} dans le schéma de description entre les EA issus de $\langle DA:Personnages \rangle$ et ceux issus de $\langle DA:Action \rangle$. On remarquera qu'une certaine latitude a été laissée pour l'annotation, ainsi par exemple la mise en place de l'EA $\langle Le\ corbeau \rangle$ extrait de $\langle DA:Personnage \rangle$ a pu conduire à la mise en place d'une nouvelle UAV,

mais ce n'était pas obligatoire. Après la mise en place des relations médiatisées par $\langle Agent \rangle$, l'annotateur a choisi d'ajouter l'EA $\langle Fromage \rangle$ à l'annotation de l'UAV1, et de le mettre en relation avec $\langle Sujet principal \rangle$, ce qui est en dehors du schéma de description sd_1 , ou plutôt est inclu dans le schéma de description le plus général imaginable, spécifiant que l'annotateur est libre d'utiliser les éléments d'annotation à sa guise.

L'annotation présentée correspondrait alors à l'utilisation du schéma spécifique sd_1 associé à un schéma libre.

Schémas de description généraux

Un *schéma de description général* (ou *plan de description*) est un ensemble de schémas de description. Il spécifie quels sont les schémas de description à utiliser pour une annotation.

Par exemple, si $sdg = \{sd_1, sd_4\}$ avec sd_4 réduit à une dimension d'analyse

$$da = \langle DA:Sentiments: \{ \langle Vanité \rangle, \langle Orgueil \rangle, \langle Joie \rangle, \langle Ironie \rangle \} \rangle.$$

Alors utiliser sdg pour annoter un flux revient à utiliser sd_1 pour mettre en place une structure d'annotation, mais aussi à utiliser également les EAA extrait de da pour compléter l'annotation.

Manipulation

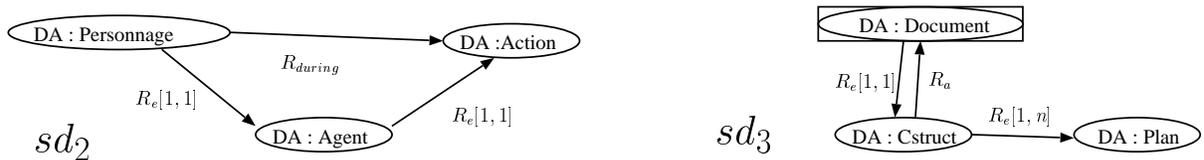


FIG. 6.21 – Quelques schémas de description à partir desquels a pu être construit sd_1

Parmi les opérateurs de construction de schémas d'annotation basiques, citons :

- la possibilité de joindre des schémas de description par certains sommets, par exemple pour préciser certains points d'annotation. Par exemple sd_1 a pu être construit à partir, entre autres, de sd_2 tel que présenté figure 6.21 ;
- la possibilité d'extraire des schémas de description d'autres schémas de description, par exemple sd_3 correspond à un schéma de description plus général que sd_1 et a pu en être extrait ;
- la possibilité d'extraire des schémas de description de l'annotation elle-même, en en repérant les régularités d'annotation. Par exemple, sd_1 a pu être mise en place en observant des annotations libres de dessins animés illustrant des contes et en en repérant les régularités.

Nous venons de le voir, schémas de description et (sous-)graphes d'annotation sont donc très proches, les uns pouvant être considérés comme abstractions des autres, qui en sont des spécialisations.

Discussion

Les schémas de description permettent de spécifier une structuration de l'annotation, de façon prescriptive. Cela répond donc à notre volonté d'avoir un moyen de contrôler l'annotation, *i.e.* l'écriture sur le flux dans le modèle des Strates-IA.

L'expressivité des schémas de description n'est pourtant pas totale : il est par exemple ardu d'exprimer dans un schéma de description que l'ensemble des UAV annotées par $\langle Plan \rangle$ dont est composé un document doit temporellement couvrir la totalité de l'UAV annotée par $\langle Document \rangle$. Ce genre de problèmes pourrait être résolu en complexifiant la définition des schémas de description, par exemple en y définissant des relations supplémentaires, ou en attachant des vérifications procédurales aux dimensions d'analyse. Le schéma de description n'en restera pas moins un graphe de dimensions d'analyses correspondant à des sous-graphes d'annotation réels.

C'est cette dernière possibilité qui se révèle en effet la plus importante : connaissant un schéma de description ayant servi à annoter des flux, c'est à dire des schémas de mise en contexte, il est possible de construire des graphes potentiels caractérisés exprimant des schémas de contextualisation. Il nous reste donc à présenter quels sont les liens entre les graphes potentiels et les schémas de description.

6.3.4 Graphes requêtes

Un *graphe requête* (ou graphe potentiel généralisé) est un ensemble de graphes potentiels caractérisés défini en intension.

Informellement, la définition d'un graphe requête est similaire à celle des graphes potentiels caractérisés, à la différence près que les sommets étiquetés auparavant par des EAA peuvent l'être par des dimensions d'analyse.

La figure 6.22 présente un graphe requête gr_1 défini à partir du schéma de description sd_1 . Seule une partie de sd_1 a été utilisée, et la dimension d'analyse $\langle DA:Action \rangle$ a été spécialisée en $\langle DA:Discourir \rangle$.

Passer d'un graphe requête à un ensemble de graphes potentiels caractérisés revient à générer n graphes potentiels pour chaque sommet étiqueté par une dimension d'analyse décrite par n filtres de désignation. Les sommets précédemment étiquetés par des dimensions d'analyse sont alors identifiés avec les points de désignation des filtres de désignation dont elles sont composées.

Par exemple, le graphe requête gr_1 possède deux sommets étiquetés par des dimensions d'analyse $\langle DA:Discourir \rangle$ et $\langle DA:Personnages \rangle$. La première DA est décrite par un unique filtre de désignation, tandis que la seconde est décrite par cinq filtres de désignation présentés figure 6.22. Le graphe requête permet de mettre en place 5×1 graphes potentiels caractérisés $gpc_{i,i=1..5}$.

Il apparaît donc que pour une dimension d'analyse « naturelle » à la base de connaissances (*i.e.* fondée sur des propriétés d'organisation explicite de celle-ci, comme le serait une dimension d'analyse décrite par fd_5), le passage d'un graphe requête à un graphe potentiel caractérisé est immédiat et économique. Dans le cas contraire — quand la dimension d'analyse est fortement personnelle ou liée à une tâche non prévue — ce passage peut générer de nombreux graphes potentiels caractérisés, et donc entraîner un nombre important d'instanciations. Nous montrons dans la suite que l'algorithme de multi-propagation s'adapte aisément à la prise en compte de dimensions d'analyse « non-naturelles », pour peu que la base de connaissances soit virtuellement complétée par une super-organisation liée à la dimension d'analyse, c'est à dire à la tâche en cours.

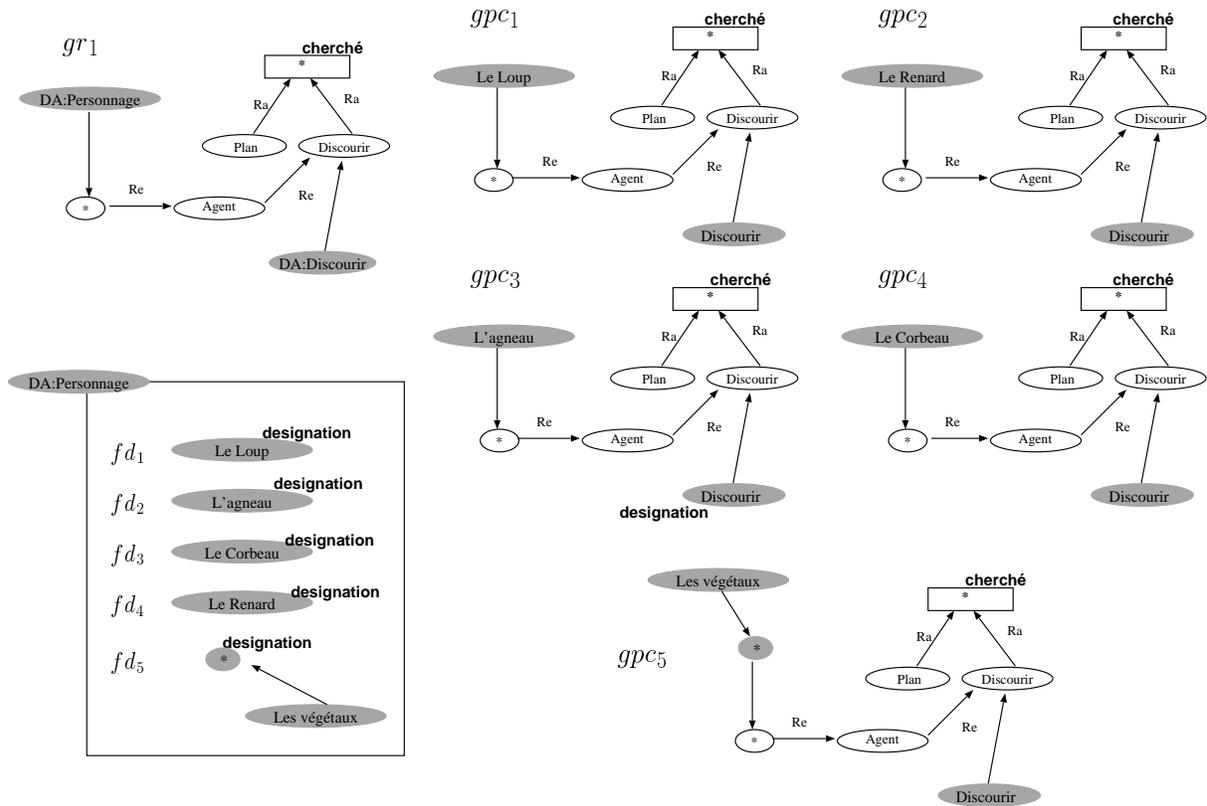


FIG. 6.22 – Un graphe requête et les quatre graphes potentiels caractérisés qu'il définit

L'instanciation d'un graphe requête peut en effet se faire de deux manières, pour des résultats équivalents.

La première méthode consiste en la mise en place des graphes potentiels caractérisés que le graphe requête décrit et de leurs instanciations, en identifiant les sommets étiquetés par les dimensions d'analyse avec les différents points de désignation des filtres de désignation qu'elles contiennent. Par exemple, on instanciera les gpc_i de la figure 6.22 et on récupérera les différentes sous-graphes instances comme résultats après cinq processus d'instanciation.

La deuxième méthode consiste à imposer pour chaque dimension d'analyse un certain nombre d'états de départ *ayant la même graine*, et signifiant les correspondances entre les EAA désignés par la dimension d'analyse avec les EAA de la base de connaissances. Par exemple l'étape P_1 dans le cas de l'utilisation de gr_1 aura sept états initiaux :

- un état e_{11} désignant la correspondance de l'EAA $\langle EAA:Discourir \rangle$, et ayant donc cette graine ;
- six états e_{1i} désignant les correspondances des EAA $\langle EAA:Le\ Loup \rangle$, $\langle EAA:Le\ Renard \rangle$, $\langle EAA:L'agneau \rangle$, $\langle EAA:Le\ Corbeau \rangle$, $\langle EAA:Le\ Chêne \rangle$, $\langle EAA:Le\ Roseau \rangle$ ayant une même graine. Cette graine virtuelle désigne une correspondance virtuelle au niveau d'un EAA virtuel dont les six EAA seraient par exemple spécialisation. Cela revient en fait à générer virtuellement dans la base de connaissances un EAA dont les différents EAA de la dimension d'analyse seraient spécialisation, c'est à dire à mettre en place une organisation virtuelle de la base de connaissances dans laquelle la dimension d'analyse serait « naturelle ». Dans cette base de connaissance modifiée, un graphe potentiel virtuel s'instancie de façon naturelle, comme illustré figure 6.23.

On aura alors $P_O = \{e_{01}, e_{02}\}$ et $P_O = \{e_{11}, e_{12}, e_{13}, e_{14}, e_{15}, e_{16}\}$ avec $V(e_{01}) = \{[f, 9]\}$ et $V(e_{02}) = \{[a, 1']\}$, $V(e_{11}) = \{[a, 1'], [b, 11]\}$, etc.

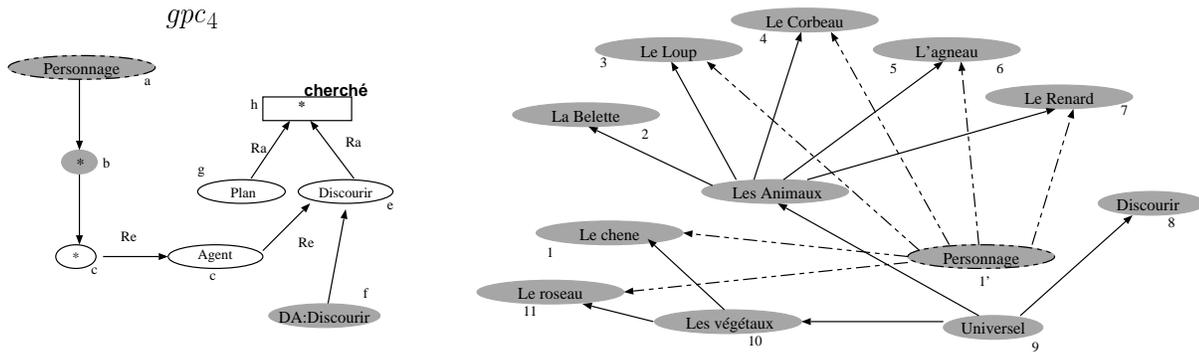


FIG. 6.23 – Un graphe potentiel caractérisé virtuel s'instancie dans une base de connaissances virtuelle

Les graphes requêtes permettent donc, dans le cas où une dimension d'analyse possède plusieurs filtres de désignation, de ne pas générer autant de graphes potentiels qu'il y a de combinaisons possible de filtres de désignation, mais d'adapter virtuellement la base de connaissances à la tâche en cours, décrite par ses dimensions d'analyse. Ils s'inscrivent donc naturellement dans les outils d'exploitation de graphes Strates-IA.

Les graphes-requêtes peuvent être construits manuellement, ou bien générés à partir de graphes potentiels caractérisés (ce qui est immédiat), des descriptions elles-mêmes (par abstraction de description et réutilisation), ou bien encore à partir des schémas de description. Ce dernier processus revient alors à choisir une partie d'un schéma de description, et à l'utiliser pour annoter un flux imaginaire dont on mettra en place les UAV totalement génériques, certains EA étant précisés directement dans le flux, d'autres par leurs liens avec la base de connaissance, soit sous la forme d'EAA extraits des dimensions d'analyse du schéma de description, soit encore sous la forme de dimensions d'analyse²¹.

Un graphe requête n'étant qu'une façon de décrire un ensemble de graphes potentiels caractérisés, dans la suite on utilisera les termes « graphes potentiels caractérisés » ou « graphe requête » indifféremment, en précisant uniquement s'il y a lieu ce que nous entendons par là.

6.3.5 Valences

Une *valence* (notion que nous avons déjà évoquée dans la partie 5.2.2) est un attribut d'élément d'annotation abstrait exprimant une possibilité de mise en relation contextuelle pour les EA qui en sont extraits.

La mise en relation peut se faire soit en utilisant un EA pour en exprimer la sémantique, soit directement par une relation élémentaire.

Une valence est donc composée :

- d'une relation contextuelle potentielle dont l'origine est un EA et l'extrémité un autre EA (éventuellement sous la forme d'un graphe requête) ;
- d'une éventuelle indication sur le type de mise en relation, c'est à dire la présence ou non d'un filtre de désignation.

Une valence s'utilise éventuellement lorsque l'EAA dont elle est attribut est inscrit dans le flux. La relation contextuelle potentielle est alors appliquée au nouvel EA *ea*, et instanciée. S'il y a des instances, leurs extrémités ea'_i sont proposées à la mise en relation avec l'EA *ea*, soit directement si la valence n'a pas plus de précisions, soit par l'intermédiaire de l'EAA proposé.

La figure 6.24 propose un exemple de valence très simple, composée d'une relation contextuelle potentielle (ici présentée sous la forme d'un graphe-requête) et de l'EAA $\langle EAA:Agent \rangle$. La relation contextuelle potentielle exprime qu'il y a lieu de chercher dans le contexte temporel d'inclusion un EA découlant de la dimension d'analyse $\langle DA:Personnage \rangle$. $\langle EAA:Agent \rangle$ exprime le fait que si l'utilisateur décide d'utiliser la valence, alors cet EA sera mis en relation avec $\langle Manger \rangle$ par l'intermédiaire d'un EA $\langle Agent \rangle$.

Une autre utilisation des valences peut consister, si la relation potentielle ne s'instancie pas, à proposer à l'utilisateur de mettre en place un EA tel qu'il y ait une instance, puis de mettre en relation ce nouvel EA avec l'EA extrait de l'EAA dont est issue la valence.

Remarquons qu'on peut considérer les valences illustrant les possibilités de liaisons d'un EA en contexte comme des schémas de description locaux de voisinage des EAA (donc des EA), c'est à dire des indications sur les contextes potentiels de voisinage de tel ou tel EA. Ce schéma local n'a *a priori* pas obligatoirement d'existence en contexte, et peut de toute façon

21. On le constate donc, l'approche des Strates-IA est unifiée aussi bien au niveau des descriptions que des descriptions de la manière de décrire ou des requêtes, on conçoit donc que les possibilités de réutilisation de l'expérience de description soient multiples, toutes les connaissances étant exprimées dans des formats comparables.

Nous supposons qu'une base de connaissances a été mise en place et que des schémas de description existent, ainsi qu'au minimum les dimensions d'analyse qui sont utilisées par ces derniers.

6.4.1 Annoter — indexer

A l'introduction d'un flux dans le système est créée une UAV le représentant, annotée par exemple par l'EA (*Flux*). A partir de cette annotation primale, qui n'est pas réalisée par l'utilisateur, toute annotation devient *sur-annotation* d'une partie de flux désignée par une UAV. Quelle que soit la suite de l'annotation, on reste donc dans le cadre de manipulation du système, qui est le graphe des Strates-IA.

Toute annotation se fera dans le cadre d'un schéma de description, lequel sera plus ou moins prescriptif.

Par exemple, une annotation totalement libre suivra un schéma de description lâche ne contenant qu'une dimension d'analyse, laquelle spécifiera tous les éléments d'annotation abstraits, c'est à dire une vue totale de la base de connaissances. A l'opposé, un schéma totalement prescriptif donnera exactement le format de structuration du document à atteindre. Le rôle de l'annotateur se limitera à l'instanciation du schéma en un graphe d'annotation, donc au choix des limites temporelles des unités audiovisuelles, et à celui des EAA inscrits dans le flux parmi ceux proposés par les dimensions d'analyse.

Entre ces deux extrêmes, tout est imaginable : par exemple un schéma de description général sera composé d'un schéma de description minimal qui guidera le début de l'annotation, et d'un schéma très lâche autorisant l'utilisateur à compléter à volonté l'annotation. Lorsque l'annotation est libre, l'utilisateur peut définir ses propres dimensions d'analyse et ses propres schémas de description, y compris en cours d'annotation. Cela revient à spécifier ses propres visées de description pour annoter ensuite en les suivant.

L'utilisation de plusieurs schémas de description différents, par exemple pour réindexer un flux déjà indexé est possible : plusieurs descriptions différentes, éventuellement générées par des annotateurs différents sont donc naturellement gérées par le modèle. La relation élémentaire permet de plus, de façon libre de mettre en relation n'importe lesquels des éléments d'annotation en place dans le graphes Strates-IA, y compris ceux ayant été mis en place suivant deux schémas de description différents.

A l'inscription dans le flux d'un élément d'annotation, l'utilisateur doit donner une valeur à ses attributs obligatoires, ainsi qu'éventuellement aux autres. Il peut également le contextualiser en le mettant en relation élémentaire avec d'autres EA. Cette mise en relation peut consister :

- à continuer à suivre un schéma de description, si celui-ci est prescriptif, auquel cas c'est le système qui spécifie les éléments à mettre en relation, et la manière de le faire ;
- à utiliser une valence de l'élément d'annotation qui vient d'être mis en place, auquel cas le système propose une mise en relation, et éventuellement la création de l'EA cible de la relation si celui-ci n'existait pas déjà ;
- à créer ou à rechercher librement l'EA cible de la relation, en explorant le contexte de l'EA qui vient d'être inscrit dans le flux.

Des assistants d'annotation, particulièrement ceux spécialisés dans l'analyse vidéo peuvent venir en aide à l'annotateur quand celui-ci analyse le flux suivant une dimension d'analyse liée à des caractéristiques calculables. Par exemple, analyser le flux suivant une dimension d'analyse liée aux plans peut prendre appui sur un assistant réalisant le découpage automatique du flux.

De la même manière, un assistant utilisant les flux image et son peut aider l'utilisateur à repérer dans le flux les occurrences d'un personnage par exemple.

6.4.2 Rechercher

Rechercher un élément des Strates-IA (plus particulièrement une séquences audiovisuelle, c'est à dire une UAV) consiste à le décrire tel qu'on aimerait le trouver, c'est à dire à le décrire contextuellement à partir d'éléments connus.

Les modes de recherche sont multiples, et se ramènent toujours, en dernière analyse, en un ensemble de graphes potentiels caractérisés qu'il s'agit d'instancier²².

Tout d'abord, la recherche peut consister à instancier des « schémas de recherche » déjà mis en place, et qu'il convient de compléter, à charge pour un assistant automatique de mener la recherche en précisant les graphes potentiels caractérisés sous-entendus par les schémas. Par exemple, un schéma basique consiste à proposer à l'utilisateur de retrouver des documents annotés par un ensemble de termes (EA exprimant des caractéristiques de haut-niveau). Il s'agit alors de retrouver une UAV annotée directement par $\langle Document \rangle$, et annotée, dans un contexte d'inclusion temporelle (R_{it}), par les termes cherchés. Alors le système — ou plutôt un assistant de recherche spécialisé — mettra en place un graphe potentiel tel que celui de la figure 6.25, l'instanciera, et retournera les UAV résultats.

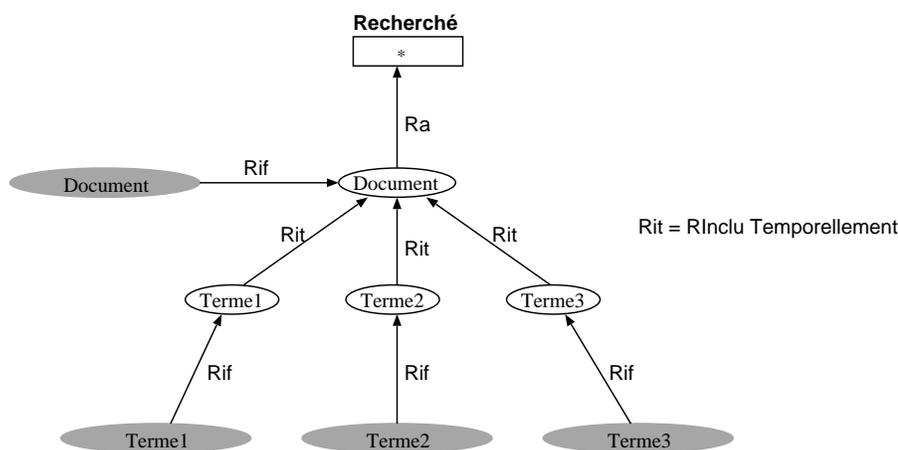


FIG. 6.25 – Un graphe potentiel caractérisé de recherche mis en place automatiquement par un assistant: l'élément recherché est un document annoté par trois « mots-clé » $Terme1$, $Terme2$ et $Terme3$. Par exemple, si on avait $Terme1 = \text{Le Renard}$, $Terme2 = \text{Le Corbeau}$ et $Terme3 = \text{Fromage}$ alors l'instanciation du graphe potentiel dans le graphe de description de la figure 6.20 permettrait de récupérer l'UAV UAV1 comme éléments Recherché.

Ensuite, la recherche avec des graphes requêtes peut se baser sur la construction par l'utilisateur de sa requête comme annotation d'un flux virtuel.

22. Remarquons que les graphes potentiels permettent d'exprimer implicitement ou explicitement des « conjonctions » d'annotations (tel et tel EA doivent être présents), des « disjonctions » (par exemple avec des graphes requête où une dimension d'analyse désigne tel EAA ou tel autre), mais ne permet pas de gérer les négations de façon simple. Cela s'accorde avec notre approche qui considère d'une part qu'on repère « positivement » des objets d'intérêts, d'autre part que les graphes de recherche sont également construits à la manière de descriptions et non dans un formalisme différent

Cette annotation se fait en utilisant un schéma de description, libre ou contraint suivant que l'utilisateur n'a pas d'idée préconçue sur l'annotation qu'il cherche, ou au contraire sait quel schéma de description a été utilisé pendant la phase d'indexation. Dans tous les cas, il met en place les annotations qu'il recherche, ainsi que la manière dont il souhaite que celles-ci soient contextualisées, de façon plus ou moins lâche, en utilisant par exemple des dimensions d'analyse pour certains nœuds du graphe requête.

Une recherche par l'exemple consiste à annoter un flux réel (l'exemple) en suivant un schéma de description, et à rendre génériques certains sommets avant d'instancier le graphe potentiel ainsi obtenu.

Une recherche à l'aide d'un exemple connu du graphe Strates-IA consiste à extraire un sous-graphe d'annotation de celui-ci afin de s'en inspirer pour mettre en place la requête²³

Dans tous les cas, rechercher des objets des Strates-IA consiste à décrire ceux-ci comme objets génériques dans leurs relations avec d'autres objets génériques, ceux-ci pouvant correspondre à des objets non connus, ou au contraire à des objets connus (auquel cas ils serviront de point de départ à la recherche d'instances). Bref, il s'agit de décrire les objets visés en les contextualisant :

- par rapport à d'autres objets génériques, par exemple figure 6.25 l'UAV totalement générique $\langle * \rangle$, $\langle Document \rangle$, $\langle Terme1 \rangle$, $\langle Terme2 \rangle$, $\langle Terme3 \rangle$ sont des objets virtuels en relations contextuelles ;
- par rapport à des objets connus, par exemple figure 6.25 $\langle EAA:Document \rangle$, $\langle EAA:Terme1 \rangle$, $\langle EAA:Terme2 \rangle$, $\langle EAA:Terme3 \rangle$ sont des objets connus, des éléments d'annotation abstraits de la base de connaissances, qui peuvent être mis en correspondance de façon unique avec des objets du graphe Strates-IA.

Dans un espace d'information tel que le graphe Strates-IA, on peut donc considérer que l'on cherche toujours quelque-chose dont on connaît le contexte, c'est à dire que l'on vise un point de l'espace depuis plusieurs points connus à la fois, de telle sorte que les différentes visées y convergent.

6.4.3 Naviguer

La navigation résulte d'un principe similaire, à ceci près qu'un des points de visée est celui où l'on se trouve réellement, et que l'on souhaite naviguer vers le point cherché.

Naviguer dans le graphe consiste donc toujours à mettre en place une relation contextuelle potentielle dont le sommet origine correspondra à l'élément du graphe où l'on se trouve. Après l'instanciation, les sommets extrémité des sous-graphes instances seront proposés à la navigation.

Plusieurs conséquences en découlent.

Première conséquence, la *navigaton prend plusieurs formes* suivant les types des éléments origine et extrémité des relations contextuelles potentielles.

Considérons pour commencer les formes de navigation les plus importantes, à savoir celles mettant en jeu des unités audiovisuelles, c'est à dire des morceaux de flux :

- aller des UAV vers les UAV correspond à une navigation dans le flux, dans le ou les documents, médiés par le substrat constitué par les EA ;

²³. Cela revient à extraire un schéma de description local d'une annotation déjà en place considérée bon exemple de ce qu'on cherche, afin de construire un graphe requête.

- aller des UAV vers les EAA correspond à un passage d’une vue du flux vers une vue abstraite issue de la base de connaissances. Par exemple, si une UAV est annotée par $\langle \text{Chirac} \rangle$, elle l’est également par $\langle \text{Homme Politique} \rangle$ en utilisant entre autre un contexte lié à une relation de spécialisation dans la base de connaissances :
- passer d’un EA à une UAV ou d’une UAV à un EA correspond au passage de l’annotation au flux, c’est à dire de la documentation vers les documents (et inversement) ;

Si l’on considère maintenant, en étendant quelque peu la concept de navigation hors des documents visés par l’annotation mais dans le système de documentation lui-même :

- aller des EAA vers les EAA correspond à une navigation exploratoire dans la base de connaissances ;
- aller d’EA en EA correspond à une navigation dans l’annotation, par exemple en l’absence du flux lui-même comme fichier audiovisuel à visualiser ;
- aller des EAA vers les UAV ou les EA correspond à une recherche d’utilisation des EAA dans le flux, voire peut s’apparenter à une recherche telle que nous l’avons définie dans la section 6.4.2.

Deuxième conséquence, il apparaît que la *navigation et la recherche par requêtes relèvent exactement du même principe*, c’est à dire que ces deux modes de recherche sont unifiés dans le cadre des Strates-IA.

Si nous reprenons le principe des « fenêtres » de [Chiaramella, 1997], nous pouvons considérer un système dans une fenêtre ou vue dans laquelle se situe l’utilisateur, et à partir de laquelle il s’ouvre, aussi bien par requêtes que par navigation, d’autres vues qui vont lui permettre de continuer son exploration-recherche. L’extension ici réalisée consiste à intégrer dans le schéma de la base de document les index et les descripteurs abstraits permettant d’y accéder, c’est à dire à considérer un système d’exploitation documentaire dans son ensemble (figure 6.26).

Les liens de navigation sont des éléments d’annotation d’unités audiovisuelles²⁴, et autant de types de navigation qu’il y a de manière de contextualiser sont alors imaginables. Médier le résultat d’une requête par l’endroit du graphe où l’on se trouve (*cf.* [Cunliffe *et al.*, 1997]) est également trivial.

Troisième conséquence, la navigation ne consiste plus simplement à aller d’un nœud à un nœud incident, mais l’on peut considérer des *chemins de navigation* de longueurs quelconques correspondant à toutes les manières de contextualiser possibles. Le calcul s’appuie sur le graphe pour *proposer* les liens, les liens proposés dépendent de la tâche en cours et ne sont pas figés, puisqu’ils participent de la documentation des documents audiovisuels et sont exploités en tant que tels.

La navigation dans les Strates-IA s’appuie donc sur un calcul des liens à proposer à l’utilisateur en fonction des visées de contextualisation de celui-ci (*i.e.* en fonction du contexte vers lequel il entend naviguer).

24. Par exemple pour spécifier des liens *explicites* comme liens, on peut utiliser un EA $\langle \text{Lien} \rangle$, l’éclairage de tels liens d’une UAV relevant alors de l’application d’une relation potentielle contextuelle à l’UAV courante, et du *calcul* des extrémités correspondantes.

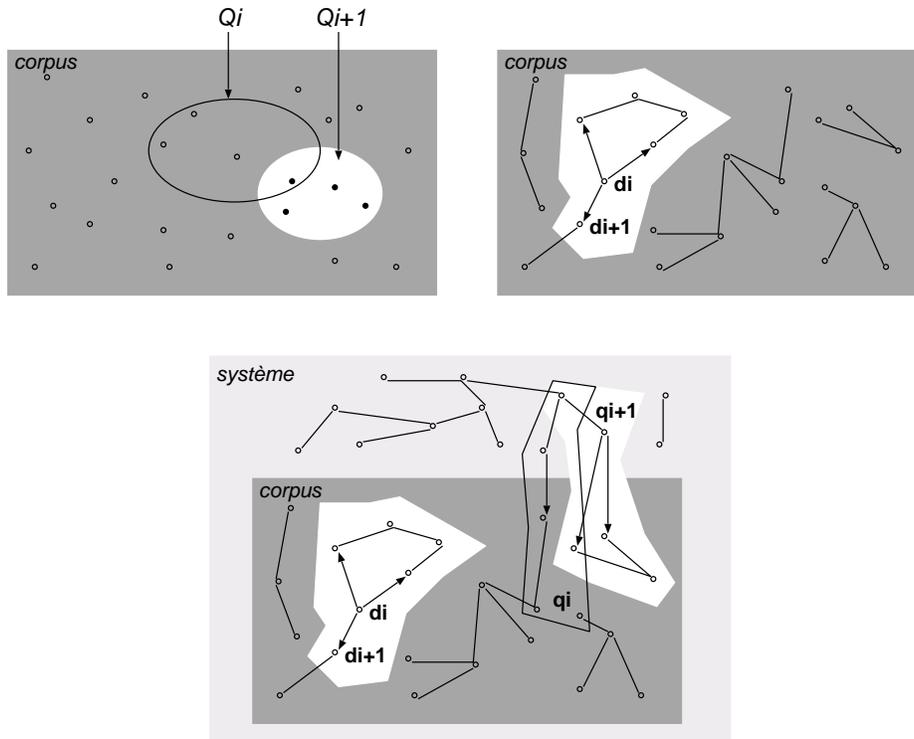


FIG. 6.26 – *Etendre le système pour considérer navigation et requêtes de la même manière*

6.4.4 Analyser

L'analyse d'un élément Strates-IA consiste en son éclairage suivant un ou plusieurs contextes particuliers. Suivant les types de contextes considérés, les objectifs de l'analyse ne sont pas les mêmes.

Annoter contextuellement une UAV consiste à chercher dans son contexte quels éléments pourraient l'annoter. Il ne s'agit pas comme ci-dessus de naviguer jusqu'à ceux-ci, mais bien de les considérer comme annotant l'UAV considérée.

Cela peut par exemple se faire en appliquant à l'UAV des relations contextuelles potentielles, spécifiées explicitement ou implicitement. Dans le dernier cas, on peut par exemple déterminer un contexte en utilisant une relation contextuelle potentielle UAV-UAV, et spécifier les types d'éléments intéressants par une dimension d'analyse, c'est à dire un ensemble de filtres de description. Alors l'assistant de construction de vue contextuelle étend les relations contextuelles en les joignant aux filtres de désignation²⁵.

Analyser une UAV suivant une dimension d'analyse consiste à lui appliquer une grille de lecture, et revient fondamentalement au même que l'annotation contextuelle que nous venons de présenter. C'est au niveau des modes de présentation de l'analyse que la différence principale se fera sentir, ainsi que sur les UAV auxquelles elle s'appliquera : on étudiera plutôt une UAV précise et de durée faible (en regard de celle du flux) afin d'y apporter des précisions ; on étudiera une UAV représentant un flux entier selon une dimension d'analyse.

²⁵. C'est avec cette idée pour objectif qu'a été originalement mis au point le modèle des Strates-IA : il s'agissait d'envisager d'emblée l'annotation d'une unité audiovisuelle comme pouvant provenir de son contexte, aussi bien temporel que structurel.

Typiquement, l'analyse d'un flux suivant une ou plusieurs dimensions d'analyse consiste à en extraire les EA appartenant aux dimensions et à les présenter sur une ligne de temps de telle sorte que l'utilisateur puisse se faire une idée des co-occurrences — temporelles ou non — des annotations. On peut envisager ce genre d'analyse comme une « linéarisation » de l'annotation suivant un certain nombre d'axes, une projection des visées d'analyse sur le flux.

Dans le cadre de l'annotation textuelle, [Stiegler, 1990] remarque que « l'application d'un même mot-clé à deux passages différents, où le mot n'est pas nécessairement présent dans les textes commentés eux-mêmes, est la création d'un lien conceptuel, d'une corrélation, l'élaboration d'une « grille de lecture » ». L'écriture sur le document est en effet déjà analyse, ainsi que le choix d'un terme plutôt qu'un autre. La co-occurrence d'un terme dans plusieurs document est une relation contextuelle entre deux UAV passant par la base de connaissances. L'analyse inter-documentaire peut se baser sur ce genre de contextes pour étudier par exemple comment deux documents audiovisuels « résonnent » entre eux.

6.4.5 Editer des documents, construire des documents pour la présentation

L'édition d'un document audiovisuel se base sur la recherche de parties de documents en vue de leur montage dans un nouveau document. Cela ne peut donc se concevoir que si des connaissances suffisantes existent dans la base.

Ces connaissances de montage se basant *a minima* sur des relations temporelles entre les UAV, le modèle des Strates-IA est à même de les prendre en compte dans des graphes de description. Il s'agirait alors de créer un graphe G_V annotant le document virtuel tel qu'on le souhaite, des relations temporelles étant mises en place entre les UAV virtuelles. Alors un ou plusieurs assistants adéquats se chargeraient de chercher dans le graphe général Strates-IA des instances de sous-parties du graphe G_V qui correspondraient à la description. Le reste du graphe G_V , c'est à dire les sous-parties non utilisées, pourrait alors correspondre à une annotation *a priori* du nouveau document.

Dans le même ordre d'idée, mais pour des applications plus simple, l'édition ou la génération automatique de présentations se basent sur la construction de documents réponses aux requêtes de l'utilisateur. Les graphes de description des nouveaux documents de présentation peuvent alors être créés automatiquement en fonction de la présentation désirée, avec des sous-parties correspondant aux diverses requêtes de l'utilisateur.

Les Strates-IA permettent donc d'une certaine manière de décrire en contexte des parties du nouveau document à mettre en place, le système cherchant dans le graphe des UAV correspondant le plus possible à celles cherchées.

6.4.6 Extraire une documentation

Les Strates-IA peuvent être utilisées comme indexation de base, descriptions de documents en vue d'une utilisation dans d'autres systèmes et suivant d'autres formats. On peut alors envisager d'extraire pour un graphe Strates-IA une hiérarchie documentaire qui correspondra à une vue du document liée à un besoin particulier, et lui sera attachée comme documentation dans un autre format (par exemple du HTML).

Dans ce cas de figure, une description en Strates-IA ne sert que de base à une description, dans un format qui peut être autre, dans laquelle une contextualisation est réalisée et présentée à l'utilisateur. Cela se rapproche d'ailleurs d'une vue d'un graphe Strates-IA présentée à l'utilisateur dans le cadre de son interaction avec le système d'information audiovisuelle. Les différentes

informations utiles présentées à l'utilisateur subissent une médiation graphique, et expriment d'une manière ou d'une autre la présentation d'un élément des Strates-IA mis en contexte avec (contextualisé par) d'autres éléments. C'est cette approche systématique de présentation d'éléments des Strates-IA comme vues contextuelles que nous adoptons dans notre prototype.

6.5 Conclusion

Nous avons présenté dans ce chapitre comment il était possible d'exploiter contextuellement les modèles de représentation de documents audiovisuel en Strates Interconnectées par les Annotations. L'exploitation se base sur un outil fondamental, le graphe potentiel, qui permet d'exprimer des contextes ; ainsi que sur d'autres outils tels que les dimensions d'analyse et les schémas de description. La dernière section du chapitre est destinée à la présentation des différentes utilisations qu'il est possible de faire d'un système fondé sur les Strates-IA.

La suite de ce document sera consacrée à la présentation des prototypes réalisés et des fonctionnalités qu'ils implantent (chapitre 7) ; à l'étude d'un cadre original de modélisation d'un système d'information dans l'objectif d'en stocker l'expérience d'utilisation de telle sorte que celle-ci soit *expliquée* (chapitre 8) ; ainsi qu'à une discussion plus générale sur les documents et les connaissances nous permettant d'y situer l'approche des Strates-IA (chapitre 9).

Chapitre 7

Réalisations

Sommaire

7.1	Prototype de validation de l'algorithme de multi-propagation	179
7.2	Prototype SESAME	182
7.2.1	Principes généraux	182
7.2.2	Fonctionnalités	183
7.2.3	Evolutions	184

Nous présentons dans cette partie les réalisations que nous avons mises en place dans le cadre du projet SESAME. Un premier prototype a été développé entre octobre 98 et janvier 99, qui a permis de tester et de valider l'approche en graphes et les algorithmes d'instanciation de graphes potentiels. Un deuxième prototype, développé entre juillet et novembre 1999 reprend les résultats du premier et fournit une interface conviviale d'annotation et d'exploration d'une base de documents.

7.1 Prototype de validation de l'algorithme de multi-propagation

Le premier prototype que nous avons réalisé, déjà évoqué en 6.2.6, a été développé en C++ en se basant sur la bibliothèque LEDA¹. Cette bibliothèque d'outils de mathématiques discrètes fournit entre autre des librairies puissantes de création et de manipulation de graphes étiquetés génériques. Nous l'avons donc choisie après avoir été tenté par GRAS², qui lui fournit les primitives nécessaires à la gestion de bases de données de graphes, mais s'est révélé trop lourd pour permettre nos expérimentations.

Le premier prototype permet essentiellement de comparer un certain nombre d'algorithmes de calcul d'isomorphismes de sous-graphes partiels, c'est à dire des instanciations de graphes potentiels. Il permet donc de générer un graphe général Strates-IA (sous la forme d'un unique graphe LEDA) en ayant indépendamment accès à la base de connaissances, mais aussi à des sous-graphes d'annotation liés à un flux. Il permet également de générer des graphes potentiels comme graphes LEDA. Les éléments de graphe considérés à ce stade se limitent à des doublets $\{Type, Nom\}$, avec *Type* pouvant prendre les valeurs *UAV*, *EA* et *EAA*. Les connaissances temporelles liées aux UAV sont gérées dans des tableaux hors du graphe LEDA, et une extension de la fonction de recherche de voisins d'un nœud suivant une relation a été étendu aux relations temporelles.

1. Voir <http://www.mpi-sb.mpg.de/LEDA/>.

2. Voir <http://www-i3.informatik.rwth-aachen.de/research/gras/index.html>.

L'interface du prototype étant uniquement en mode texte, et permettant en conséquence de ne construire les graphes que de façon fastidieuse et non explicitement liée aux flux audiovisuels, des mécanismes de génération automatique du graphe global, et d'extraction de graphes potentiels à partir de celui-ci, ont permis de générer des graphes de tailles conséquentes.

Les algorithmes d'instanciation de graphes potentiels suivants ont été implantés et testés :

- L'algorithme de [Ullman, 1976], dont nous avons déjà évoqué les performances médiocres (voir p 6.2.2) dès que la taille du tableau des correspondances graphe potentiel/graphe global devient grande (par exemple 10000), mais qui a l'avantage de fournir une base solide de comparaison.
- Une version modifiée de ce dernier calculant d'abord une sous-partie du graphe général de diamètre³ égal à celui du graphe potentiel dans lequel le graphe potentiel est alors instancié. Ceci ne permet pas de diminuer de façon conséquente la taille du graphe global — dû à la forte connexité d'un graphe Strates-IA — dès que la taille du graphe potentiel dépasse 6 ou 7 éléments ;
- Le premier algorithme de propagation que nous avons mis au point, et dont le principe est le suivant : à partir d'une correspondance donnée entre le graphe global et le graphe potentiel, on cherche à propager récursivement l'instanciation, c'est à dire à instancier le graphe potentiel privé des éléments participant déjà à des correspondances dans le graphe global lui-même privé des correspondants de ces éléments. Cet algorithme donne de bon résultats dans le cas où le graphe potentiel ne comporte pas de cycles. Dans le cas contraire, il convient d'éliminer (récursivement ou non) les doublons.
- L'algorithme de [Cordella *et al.*, 1998], qui correspond en fait à une version itérative de l'algorithme récursif de propagation, et fournit, dans le cas de cycles dans le graphe potentiel un ensemble de solutions doublonnées ;
- Notre propre algorithme de multi-propagation (présenté en 6.2), et qui donne de bons résultats en partant de l'ensemble des correspondances connues au départ, et en propageant la recherche dans une seule direction (choisie suivant une heuristique simple et améliorable) à chaque itération. La gestion des jonctions grâce à la notion de graine permet de minimiser le nombre des états générés. De plus les solutions sont fournies au fur et à mesure de leur découverte.

Le tableau 7.1 résume les caractéristiques principales des algorithmes étudiés.

Ullman	Ullman avec fenêtre	Propagation	Cordella et al. [1998]	Multi-propagation
Récursif, global, stable, lent	Inutile si la taille de la fenêtre dépasse 7	Récursif, local, doublons à éliminer	Version itérative du précédent, doublons à éliminer	Génère peu d'états, anytime, pilotable par heuristique

TAB. 7.1 – Quelques caractéristiques des algorithmes étudiés

3. Diamètre est ici considéré comme la longueur maximale des chemins d'un graphe.

La figure 7.1 présente une copie d'écran de l'affichage d'un graphe potentiel extrait automatiquement d'un graphe global généré automatiquement avec 1000 EAA, 30 flux, 1000 UAV, une moyenne de 2 EA par UAV et 30 % des EA en relation élémentaire avec d'autres EA.

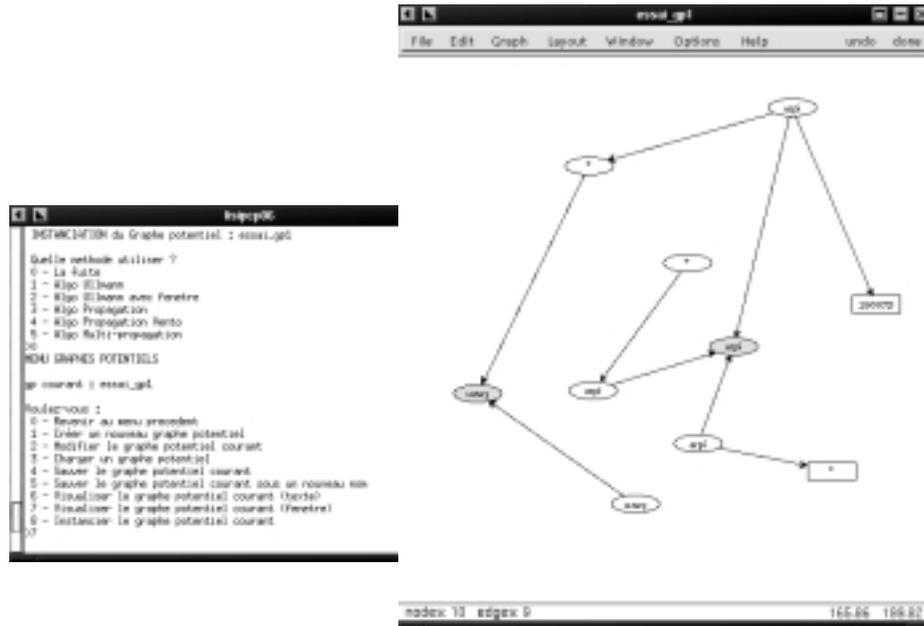


FIG. 7.1 – Interface textuelle du prototype 1 et utilisation de la fenêtre d'affichage de graphe fournie par LEDA. Les points de départ de l'algorithme de multipropagation sont ici l'UAV 290075 (rectangle) et les EAA $\langle \text{EAA:arpi} \rangle$ et $\langle \text{EAA:ugwg} \rangle$ (ovale grisé). Les EA sont les ovale clairs. Certains nœuds du graphe potentiel sont génériques (*).

Les lignes suivantes montrent un exemple d'une session de recherche d'instances du graphe potentiel de la figure 7.1 permettant de comparer deux algorithmes.

```
INSTANCIATION du Graphe potentiel : essai_gp1
```

```
Quelle methode utiliser ?
```

- 0 - La fuite
- 1 - Algo Ullmann
- 2 - Algo Ullmann avec fenetre
- 3 - Algo Propagation
- 4 - Algo Propagation Cordella
- 5 - Algo Multi-propagation

```
>5
```

```
* Resultat *
```

```
Il y a 756 instanciations du graphe potentiel essai_gp1
dans le graphe general de la base courante essai
L'operation a pris 53 ms.
```

```
INSTANCIATION du Graphe potentiel : essai_gp1
```

```
Quelle methode utiliser ?
```

- 0 - La fuite

```
1 - Algo Ullmann
2 - Algo Ullmann avec fenetre
3 - Algo Propagation
4 - Algo Propagation Cordella
5 - Algo Multi-propagation
>1
Taille G :5902 / Taille Gp :10
* Resultat *
Il y a 756 instanciations du graphe potentiel essai_gp1
dans le graphe general de la base courante essai
L'operation a pris 94555 ms.
INSTANCIATION du Graphe potentiel : essai_gp1
```

Notre premier prototype a donc permis d'atteindre les objectifs que nous nous étions fixés en termes de validation des principes généraux des Strates-IA et d'instanciation de graphes potentiels dans un graphe Strates-IA.

7.2 Prototype SESAME

Sur la base de l'expérience acquise au cours du développement du premier prototype, nous avons pu fixer les bases d'un second prototype, qui a été principalement développé au cours du stage de Projet de Fin d'Etude de Elöd Egyed-Zsigmond [Egyed-Zsigmond, 1999b]. Ce prototype SESAME permet de fournir un cadre pour l'expérimentation en vraie grandeur des Strates-IA, c'est à dire l'implantation des principaux concepts présentés, manipulés au sein d'une interface conviviale. Il implante également la sauvegarde des données en XML ainsi que la possibilité de gérer de façon uniforme les graphes Strates-IA comme des graphes LEDA ou bien comme ensemble de fichiers XML [Egyed-Zsigmond, 1999a], voir 9.2.2. Ce prototype a été développé en Visual C++ sous Windows NT, en utilisant LEDA comme librairie fournissant des structures de données aisément utilisables, ainsi que le DirectX MediaFramework de Microsoft et XML4C de IBM Alphaworks. Nous en présentons les principes généraux, ainsi que les principales fonctionnalités.

7.2.1 Principes généraux

Interface de programmation. L'architecture du prototype est fondée sur un graphe Strates-IA, dont une interface de programmation est disponible et permet un accès à des fonctions de base. Ce graphe est indépendant de l'application qui l'exploite, c'est à dire qu'il peut être représenté de manière quelconque. Ainsi, celui-ci est actuellement représenté comme un graphe LEDA unique d'objets tels que les UAV, les EA ou les EAAs. La sauvegarde du graphe se fait en utilisant les primitives de sauvegarde LEDA associées à des sérialisations des nœuds en XML, mais tout a été prévu afin qu'il soit aisé de porter les structures de données dans des fichiers XML (voir 9.2.2). L'accès au graphe Strates-IA est normalisé et passe obligatoirement par l'interface de programmation.

Vues contextuelles d'éléments du graphe. Le prototype est conçu comme une application stockant ses données dans un graphe Strates-IA. Le principe de l'application est de fournir des vues du graphe qui en autorisent soit la consultation, soit la modification.

Une vue d'un élément du graphe Strates-IA est, de façon générale, la présentation de cet élément (EAA, EA ou UAV) et d'éléments qui sont dans son contexte (déterminé par des graphes

potentiels). Tout présentation du graphe Strates-IA passe obligatoirement par une vue d'un de ses éléments et d'éléments contextualisés par rapport à lui.

Par exemple, la présentation de l'annotation d'un flux est une vue d'une UAV annotée par $\langle Flux \rangle$. Dans cette vue se retrouvent également les éléments d'annotation et les unités audiovisuelles qui se trouvent dans son contexte d'inclusion temporelle. Une vue d'un élément d'annotation peut être constituée de ses attributs, mais aussi des éléments d'annotation qui sont dans son contexte immédiat (à une distance de 1 dans le graphe).

On essaye ainsi de ne présenter des informations à l'utilisateur que de façon contextuelle. Les vues peuvent cependant être complexes : par exemple, une vue d'une UAV peut consister en un lecteur audiovisuel permettant l'annotation de l'UAV. Une vue de la base de connaissances peut être considérée comme une vue de son EAA racine et de l'arborescence des EAA qui sont dans son contexte suivant la relation de spécialisation, *etc.*

Le principe des vues du graphe que nous proposons trouve un écho dans le mécanisme propre à Visual C++ des vues de documents.

La figure 7.2 présente quelques vues d'un graphe Strates-IA. Différentes dimensions d'analyse et la base de connaissances se trouvent à gauche. Une vue de l'UAV annotée par $\langle Flux \rangle$ permet de visualiser les UAV et les EA primitifs qu'elle contient temporellement. Une autre vue, composée entre autre d'un lecteur audiovisuel permet d'annoter. Ainsi, sur l'exemple présenté, deux unités audiovisuelles sont en cours de création, la première est définie par l'EAA $\langle EAA:Personnage \rangle$ et son annotation est complétée à l'aide de $\langle EAA:Entrer \rangle$. La seconde est définie par $\langle EAA:Prêtre \rangle$ et commence à la 23ème seconde du flux.



FIG. 7.2 – Plusieurs vues d'un graphe Strates-IA : base de connaissances, dimension d'analyse, ligne de temps et vue d'annotation

7.2.2 Fonctionnalités

Nous décrivons ici les fonctionnalités offertes à terme par le prototype SESAME pour l'exploitation des Strates-IA.

Annotation. L’annotation se fait toujours dans le contexte d’une ou de plusieurs dimensions d’analyse, qui sont en permanence disponibles. La création de nouvelles UAV se fait dans une vue d’UAV telle qu’un lecteur audiovisuel ou une vue en ligne de temps (*cf.* figure 7.2). Une UAV peut soit être décrite directement par son extension temporelle, soit en en fixant une borne de départ et en suspendant la création effective à la découverte d’une borne d’arrivée.

A la mise en place d’un EA, le système impose de compléter ses attributs marqués comme obligatoires, et propose de compléter les attributs facultatifs. La mise en relation d’EA est une tâche qui peut impliquer la création de nouveaux éléments d’annotation, et l’utilisation d’une valence liée à l’EAA utilisé. L’annotation peut être contrainte par un schéma de description spécifiant les dimensions d’analyse à utiliser et les relations entre celles-ci.

Exploration. L’exploration du système des Strates-IA passe bien entendu par la présentation contextuelle de ses différents éléments suivants des vues, et le passage d’une vue à l’autre.

La navigation est pilotée par les graphes potentiels qui fournissent des contextes qu’il est possible d’atteindre.

Gestion des connaissances d’exploitation. Le système permet de mettre en place la base de connaissances, c’est à dire de spécifier les EAA et leurs relations, ainsi que leurs attributs. Les dimensions d’analyse peuvent être mises en place comme ensemble d’EAA ou/et d’autres dimensions d’analyse. Une interface de mise en place de graphes potentiels permet de construire manuellement les sommets, ou bien de réutiliser (en les abstrayant) des éléments du graphe réel. Les graphes potentiels sont stockés en tant que tels. Ils sont également utilisés pour mettre en place des valences. Une interface similaire à celle permettant de construire des graphes potentiels permet de construire des schémas de description comme graphes de dimensions d’analyse.

Requêtes. Les requêtes au système passent par le choix ou la mise en place de graphes potentiels caractérisés adaptés à la tâche en cours. Après instanciation, il est alors possible de visualiser les résultats, c’est à dire des vues des nœuds de sous-graphes instances mis en évidence par des points de caractérisation.

D’autres requêtes plus générales sont possibles, qui correspondent à un niveau supplémentaire d’abstraction. Par exemple il est possible de chercher des unités audiovisuelles telles qu’elles possèdent dans leur contexte temporel un certain nombre d’EA (*cf.* 6.4.2, p. 172). Alors l’utilisateur doit juste spécifier ceux-ci, tandis que le système se charge de la mise en place des graphes potentiels correspondants et de la présentation des résultats.

7.2.3 Evolutions

Avec ce prototype, nous avons mis en place les bases nous permettant de tester et d’expérimenter l’annotation et l’utilisation des différents concepts des Strates-IA.

Bien entendu, il reste du travail sur la gestion des différents outils, l’interface graphique, la gestion de l’expérience, *etc.* Ces principes pourront cependant être implantés de façon aisée à partir du cadre d’étude que nous avons mis en place, tout ajout à l’interface ne relevant que de ce niveau, et non de celui du graphe Strates-IA.

Indépendamment, il sera possible de faire évoluer l’algorithme d’instanciation sans toucher ni à la représentation du graphe ni à l’interface graphique. Le prototype SESAME se révèle donc modulaire dans son organisation, et permettra une évolution dans toutes les directions utiles.

Assistants. Terminons en disant un mot sur la notion d'assistants dans le système SESAME. Le concept d'assistant a été présenté dans [Prié *et al.*, 1998a], et correspond à un agent informatique aidant l'utilisateur dans ses différentes tâches d'exploitation d'un graphe Strates-IA. Des assistants peuvent se rencontrer à tous les niveaux de tâches de l'utilisateur :

- Les assistants d'annotation l'aideront à mettre en place la description du document, par exemple en contrôlant celle-ci à partir de schémas de description, en instanciant les valences, *etc.* Certains assistants aideront par exemple à la détection de visages ou au traitement automatique des sous-titres.
- Les assistants de recherche permettront de manipuler des graphes requêtes, de les construire, de proposer des contextes prédéfinis pour la navigation, de réutiliser l'expérience passée d'utilisation, *etc.*
- Les assistants d'analyse proposeront des schémas d'analyse pré-construits, et généreront des présentations adaptées, de même que les assistants de génération.
- Les assistants de construction de dimensions d'analyse, de schémas de description, de graphes potentiels caractérisés fourniront les guides nécessaires à la mise en place d'instances de outils d'exploitation.
- ...

La notion d'assistant a été validée dans le cadre d'un travail sur une autre version du prototype SESAME [Mille *et al.*, 1998], avec une approche distribuée basée sur CORBA. Un assistant d'annotation image a ainsi été construit dans ce contexte.

Tous ces assistants, présentés ici de façon très générale, ont accès à l'ensemble de connaissances du système, ainsi qu'à d'autres connaissances, par exemple les algorithmes de détection de plans, de reconnaissance de la parole, *etc.* Les assistants ne font donc pas partie du « système Strates-IA de base » (*i.e.* les Strates-IA peuvent être utilisées sans leur aide), mais participent à son exploitation en collaboration avec l'utilisateur. Ils possèdent à cet égard une connaissance plus ou moins limitée des mécanismes des Strates-IA, et doivent être considérés comme des agents intelligents, qui s'appuient sur des modèles de connaissances appropriés aux tâches de l'utilisateur qu'ils assistent.

Chapitre 8

Vers une utilisation de l'expérience pour l'assistance à l'utilisateur

Sommaire

8.1	Modéliser la tâche d'un utilisateur pour aider à sa réalisation .	188
8.2	Modéliser utilisation et tâches pour aider l'utilisateur	190
8.2.1	Définitions	191
8.3	Vers une généralisation de l'approche	195
8.4	Discussion: exploiter l'expérience	195
8.4.1	Les cas d'utilisation pour l'illustration et la validation	196
8.4.2	Apprentissage	196
8.4.3	Aide à l'utilisateur fondée sur l'expérience	197
8.5	Conclusion	198

Nous abordons dans ce chapitre l'étude¹ de l'utilisation d'un système fondé sur les Strates-IA en vue de l'organisation et de la rationalisation de l'expérience concrète d'utilisation.

L'objectif qui nous guide est en effet l'assistance à l'utilisateur fondée sur l'expérience d'utilisation. Nous présentons en conséquence dans ce chapitre les linéaments d'un cadre général pour l'assistance à la tâche d'un utilisateur exploitant les services d'un système d'information. Nous proposons un modèle de description d'un système d'information en termes de modèle d'utilisation et de modèles de tâches, lesquels permettent d'organiser et par conséquent d'expliquer des cas d'utilisation portant la connaissance d'utilisation. Ces cas pourront ensuite être utilisés pour aider l'utilisateur, et nous présentons quelques pistes de recherche allant dans cette direction.

Un système d'information en général doit en effet idéalement répondre au mieux aux questions que lui pose un utilisateur pour lui faciliter sa *tâche*. Un tel système n'est alors pas simplement une composition d'applications informatiques capables de répondre à un nombre limité de questions prédéfinies (par exemple comme dans un système comptable), mais une collection de services offerts pour répondre à des *classes de questions* correspondant à des *classes de tâches*. Le système d'information est dans ce cas général essentiellement vu comme une mémoire plus ou moins partagée que l'on interroge pour mener à bien une tâche particulière. Les sous-tâches majoritairement confiées à l'ordinateur sont liées à son rôle de mémoire. Il s'agit d'une part d'organiser les

1. Cette étude reprend et complète les idées que nous avons présentées dans [Prié *et al.*, 1999b].

nouvelles informations dans cette mémoire, et d'autre part de calculer-restituer les informations pertinentes, en collaboration avec l'utilisateur. Nous avons vu en 1.3.2 que cette collaboration impliquait également que le système puisse aider l'utilisateur en cas de besoin. Nous avons alors souligné qu'il convenait de *modéliser les tâches utilisateurs* de façon à ce que celles-ci soient manipulables par le système.

Dans le cadre d'un système d'information documentaire tel qu'un système construit sur les Strates-IA, il ne s'agit ni de gestion de base de données, ni de gestion de bases de connaissances au sens classique de ces termes, compte tenu du fait que les processus d'organisation de la mémoire et de remémoration ne sont pas liés à des tâches parfaitement identifiées à l'avance :

- un schéma de description, même extrêmement prescriptif peut ne représenter qu'une description d'organisation locale d'une base de documents, tandis que l'utilisation qui peut être faite de cette dernière n'est contrainte que par le format des graphes potentiels, qui est très libre.
- Les « inférences contextuelles »² qu'il est possible de réaliser sur le système à base de connaissances qu'est un graphe Strates-IA ne sont de la même manière contraintes que par le format d'expression des graphes potentiels.

On se rapproche en fait — du point de vue de l'organisation des connaissances — tout à la fois des bases de données semi-structurées (dans lesquelles les tâches ne sont absolument pas prises en compte) et de l'organisation des données en documents XML. Dans ces systèmes, comme dans un système Strates-IA, l'exploitation du système suivant des schémas locaux (on sait qu'à tel endroit la connaissance est organisée ainsi) comme son exploration sont les moyens privilégiés de l'utilisateur dans sa quête d'assistance à sa propre tâche. L'aide que peut lui apporter le système doit se baser sur l'expérience d'utilisation comme matériel premier d'apprentissage et de réutilisation.

Nous présentons tout d'abord dans ce chapitre comment il est possible de représenter la tâche d'un utilisateur, en nous intéressant plus particulièrement à la tâche de description et d'annotation dans les Strates-IA. Nous proposons ensuite de décrire un système informatique en modèle d'utilisation d'une part et en modèles de tâches d'autre part, afin d'être à même de décrire des cas d'utilisation permettant de stocker et d'exploiter l'expérience d'utilisation. Après une généralisation de l'approche proposée, nous fournissons quelques pistes d'exploitation des cas d'utilisation.

8.1 Modéliser la tâche d'un utilisateur pour aider à sa réalisation

Une *tâche* (humaine) correspond à ce qu'une personne doit réaliser, se propose de faire dans le cadre d'une activité.

Dans le contexte de l'ingénierie des connaissances, la modélisation de la notion de tâche a été proposée [Schreiber *et al.*, 1993] et a même fait l'objet de tentative d'ontologie de tâches génériques (*i.e.* spécialisable pour toute tâche), comme par exemple dans [Chandrasekaran *et al.*, 1998]. Cette approche est alors totalement liée au développement de systèmes à base de connaissances.

La tâche de description et d'annotation de documents audiovisuels est une tâche d'une part non réalisable totalement automatiquement, d'autre part difficile même pour un utilisateur humain. Il y a donc lieu d'assister l'utilisateur effectuant cette tâche, et de considérer celle-ci en tant que telle.

2. Voir 9.3.1.

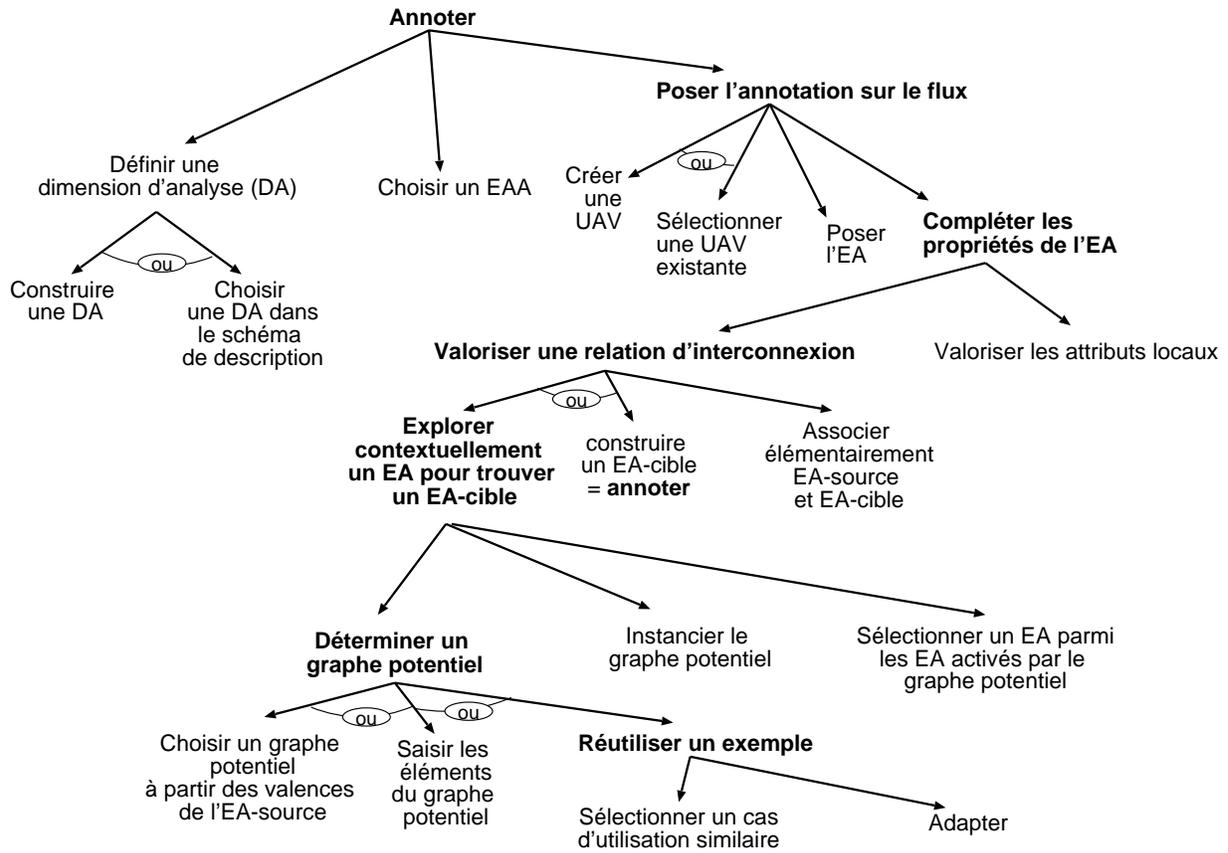


FIG. 8.1 – Diagramme de décomposition décrivant la tâche d'annotation d'un document audiovisuel. Les sous-tâches d'une tâche se lisent de gauche à droite et peuvent résulter d'un choix (OU). L'arbre est élagué, certaines sous-tâches ne sont pas décrites.

Pour illustrer comment les principes généraux des Strates-IA peuvent servir de support aux tâches d'un utilisateur, nous choisissons la tâche générique d'*annotation de document audiovisuel* qui, comme nous l'avons vu, est à la base d'autres tâches comme l'indexation, la recherche d'une séquence, l'analyse d'un document, *etc.*

Pour présenter les tâches, nous adoptons un formalisme de diagramme de décomposition (*cf.* figure 8.1) permettant de montrer comment les éléments du modèle des Strates-IA sont exploités. Chaque nœud d'un diagramme est une sous-tâche d'un utilisateur, laquelle se décompose éventuellement en succession (lisible de gauche à droite) de sous-tâches à effectuer. Certaines sous-tâches peuvent prendre une forme parmi plusieurs (OU).

Nous présentons figure 8.1 de façon simplifiée une partie d'un arbre décrivant quelques enchaînements de sous-tâches de l'annotation.

La tâche **annoter** consiste tout d'abord à choisir une dimension d'analyse explicitant un objet d'intérêt. Cette dimension d'analyse peut être construite, ou bien sélectionnée parmi les DA disponibles dans le schéma de description, et il s'agit ensuite d'y choisir un élément d'annotation abstrait, et de poser celui-ci sur le flux. Pour cela, il convient soit de compléter l'annotation d'une unité audiovisuelle pré-existante, soit d'en créer une nouvelle, que l'on annote alors. La création du nouvel élément d'annotation nécessite de valoriser les propriétés de celui-ci: ses attributs (issus par exemple de traitement d'images), mais aussi ses relations élémentaires avec d'autres éléments d'annotation. Pour choisir un autre élément d'annotation **EA-cible** à mettre en relation R_e avec notre **EA-source**, on peut soit le construire (et donc recommencer une sous-tâche **annoter**, plus précise puisque l'utilisateur est guidé par l'EA-source qu'il souhaite mettre en relation, et éventuellement par un schéma de description prescriptif), soit aller le choisir dans le contexte de **EA-source**. Pour **explorer** ce contexte, on peut par exemple déterminer un graphe potentiel décrivant la portée du contexte, et instancier celui-ci, avant de sélectionner l'**EA-cible** parmi les candidats activés. Un graphe potentiel contextuel peut être déterminé de plusieurs façons: soit il fait partie des valences de l'**EA-source**, soit on le construit directement (réservé aux utilisateurs experts); soit enfin on réutilise un exemple issu d'un cas d'utilisation similaire, que l'on adapte au cas courant (*cf.* 8.4.3). L'adaptation consiste à substituer des éléments du graphe potentiel exemple par de nouveaux éléments corrigeant ainsi les écarts de contexte.

8.2 Modéliser utilisation et tâches pour aider l'utilisateur

L'assistance à la tâche de l'utilisateur nécessite sa description sous une forme homogène à la représentation générale du système. Nous avons vu qu'une tâche peut se décomposer sous la forme d'un arbre, mais il est également possible de la décrire par ses spécifications précises comme proposé dans les outils implantant Common Kads [Martin, 1994].

Nous excluons ce dernier cas de la recherche d'information semi-structurée telle que nous l'envisageons, car les tâches ne peuvent généralement pas y être décrites de façon précise en terme procédural. La réalisation d'une tâche peut par contre *être partiellement décrite sous la forme des éléments manipulés par l'utilisateur pour l'accomplir, ainsi que par les relations qui existent entre ces éléments et qui, d'une certaine façon, l'expliquent*. En conséquence, une tâche peut être modélisée par une description des éléments qui doivent ou qui peuvent la caractériser, conduisant à la mise en place de formulaires « expliqués ». Nous présentons dans la suite quelques définitions précisant ces notions dans le cas général d'un outil informatique, en les illustrant avec le modèle des Strates-IA.

On notera que le travail présenté ici est exploratoire et que les idées qui y sont décrites le

sont de façon relativement générale. Nous ne prétendons donc pas donner une définition formelle ou définitive des objets présentés, mais plutôt donner les principes généraux et les intuitions fondamentales de l'approche.

8.2.1 Définitions

Outil informatique

Un outil informatique est un logiciel exploité pour une classe de tâches. Il est caractérisé par les *objets de base* qu'il fournit à *l'utilisateur*. Ces objets de base peuvent être décrits par un *modèle d'utilisation*.

Dans le cadre des Strates-IA, l'outil informatique de base serait par exemple un outil d'annotation manuelle de séquences audiovisuelles, aussi bien pour l'archivage que pour la recherche. Les objets de base sont ceux décrits dans les chapitres 5 et 6.

Utilisateur

Un utilisateur est un être humain exploitant les possibilités offertes par un outil informatique dans le cadre d'une tâche. La tâche considérée peut être une sous-tâche d'une tâche beaucoup plus générale. Elle correspond à un sous-but que l'utilisation de l'outil informatique permet d'atteindre plus facilement. Une tâche peut être décrite par un *modèle de tâche*.

Dans le cadre des Strates-IA, la tâche d'annotation précédemment présentée devrait par exemple être modélisée.

Modèle d'utilisation

Un modèle d'utilisation est intrinsèquement lié à l'outil exploité. Il décrit les objets de base de l'outil informatique ainsi que les relations qu'ils entretiennent. Un modèle d'utilisation minimal existe par définition (l'utilisation d'un outil sans mode d'emploi explicite minimal est exclue), et on considère qu'il n'existe qu'un seul modèle d'utilisation pour un outil informatique, celui-ci étant défini au moment de la conception de cet outil.

Dans le domaine de la bureautique, les objets de base principaux seraient par exemple le caractère, le mot, la ligne, le paragraphe, le document, le modèle de document, *etc.*

Le modèle d'utilisation simplifié des Strates-IA est résumé dans la figure 8.2. Les éléments d'annotation sont des inscriptions dans le flux d'éléments d'annotation abstraits et annotent les *unités audiovisuelles*. Les EAA sont organisés en réseau de relations. Des graphes potentiels, en s'instanciant, permettent de définir des instances qui sont des sous-graphes du graphe général. Des dimensions d'analyse regroupent des EAA désignés par des graphes potentiels. Les valences représentent des possibilités de relation entre EA et s'expriment également sous la forme de graphes potentiels, tandis que les schémas de description organisent des dimensions d'analyse.

Modèle de tâche

Nous appelons *modèle de tâche* une description des objets du modèle d'utilisation mobilisés par une tâche particulière avec les relations qu'ils entretiennent dans le cadre de cette tâche. Lorsqu'une tâche se révèle complexe, il peut être nécessaire de décomposer son modèle à l'aide des modèles de ses différentes sous-tâches mises en relation avec les objets du modèle d'utilisation qu'elles permettent d'obtenir. Chaque modèle de sous-tâche se trouve alors dans le contexte du modèle de tâche dont il fait partie, un modèle de tâche se décompose en sous-tâches selon une hiérarchie de décomposition.

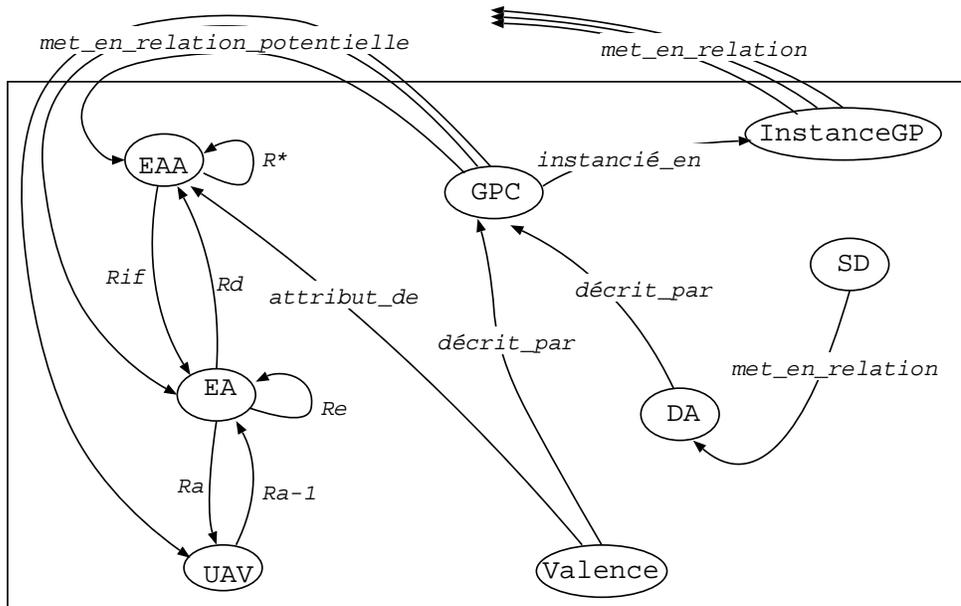


FIG. 8.2 – Modèle d'utilisation des Strates-IA : les éléments des Strates-IA et les relations minimales qu'ils entretiennent

Un exemple simple dans le domaine de la bureautique serait constitué par les modèles de document d'un traitement de texte : un modèle de lettre décrit les objets mobilisés dans la tâche d'écriture d'une lettre : expéditeur, destinataire, date, texte *etc.*

La figure 8.3 illustre comment décrire un modèle de la tâche d'annotation exploitant le modèle d'utilisation des Strates-IA. L'annotation d'une UAV (Modèle Annotation) mobilise :

- le choix d'une dimension d'analyse DA,
- le choix d'un élément d'annotation abstrait EAA qui y est sélectionné,
- la détermination de l'unité audiovisuelle UAV qui sera annotée,
- l'élément d'annotation EA issu de EAA annotant UAV, et éventuellement
- le ou les éléments d'annotation cible EAc mis en relation élémentaire avec EA. La mise en relation élémentaire est une sous-tâche qui permet d'obtenir EAc (EACible) et qui mobilise :
 - soit EAC tel qu'il a été obtenu par une sous-tâche d'annotation (qui est alors récursivement décrite par le modèle de tâche (Modèle Annotation),
 - soit :
 - un graphe potentiel GP obtenu par une sous-tâche choix GP, et
 - une instance de graphe potentiel GPi qui permet de désigner EAc. L'instance a bien évidemment pour point de départ l'élément d'annotation EA, fourni par le contexte du modèle de tâche Annotation.

La sous-tâche choix GP spécifie que le graphe potentiel désignant le contexte peut :

- être extrait d'une valence liée à l'élément d'annotation EA (passé par le contexte du modèle de tâche Mise en relation élémentaire.

- être construit manuellement au cours d'une tâche de construction,
- découler de la réutilisation d'un graphe potentiel (éventuellement adapté) d'utilisé dans un cas d'utilisation considéré comme similaire au cas en cours.

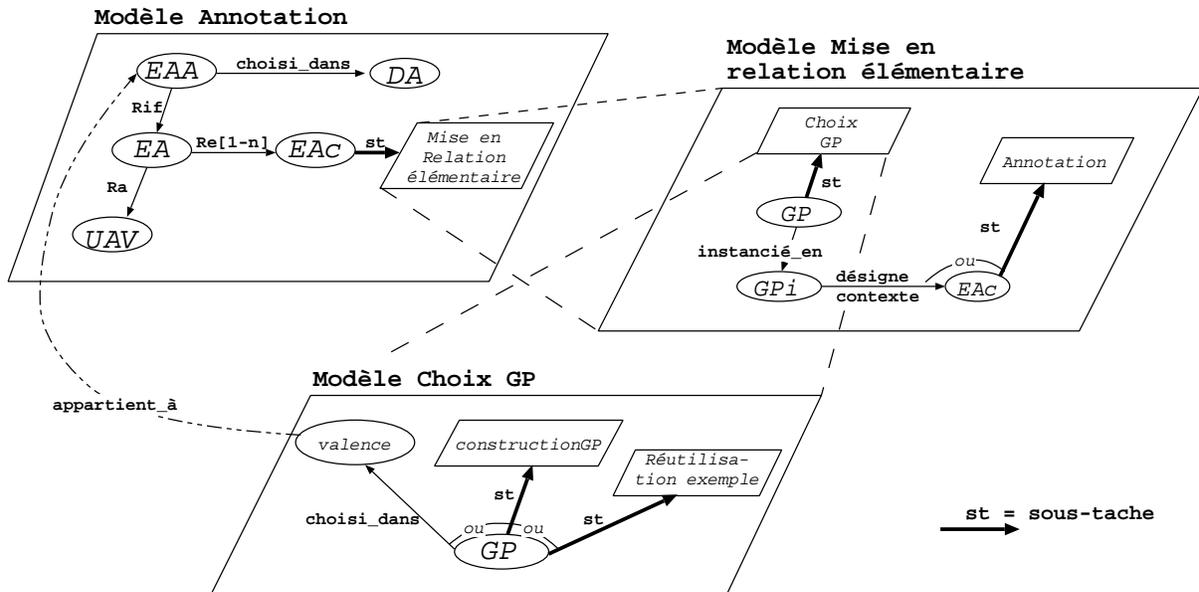


FIG. 8.3 – Modèles de tâche : Tâche d'annotation et sous-tâches s'appuyant sur le modèle d'utilisation des Strates-IA

On peut à la limite considérer le modèle d'utilisation comme un modèle de la tâche extrêmement générique *Utiliser le système Strates-IA*, qui ne décrit que les éléments qui peuvent être manipulés et leurs relations.

Cas d'utilisation

Un *cas d'utilisation*³ décrit les instances des objets du modèle d'utilisation utilisés dans le cadre d'une tâche particulière dont un modèle est disponible. Un cas d'utilisation est en quelque sorte une trace de la réalisation de la tâche correspondant au modèle de tâche utilisé. Une instance d'un modèle de tâche se compose donc des instances des éléments qui le constituent, lesquels sont éventuellement d'autres modèles de tâche instanciés.

L'exemple de cas d'utilisation présenté figure 8.4 correspond à une modèle de tâche général de **Description**, dont un modèle de sous-tâche est le modèle de tâche de la figure 8.3, et un autre correspond à une tâche très générale d'**Utilisation** des Strates-IA. Le cas d'utilisation se compose alors deux sous-cas d'utilisation correspondant aux deux modèles de sous-tâche utilisés. Le premier sous-cas représente une trace de la réalisation de la tâche d'annotation générale qui a utilisé : la dimension d'analyse DA1, dont est issu l'élément d'annotation abstrait EAA12. L'élément d'annotation EA54 est une inscription dans le flux de EAA12 et annote l'unité audiovisuelle UAV9. EA54 est en relation élémentaire avec EA451, dont la présence n'est pas expliquée (voir plus loin), et avec EA523 qui a lui été désigné par l'instance GPi11 du graphe potentiel GP11, lequel correspond à la valence *valence4* (valence de EAA12). Le second sous-cas spécifie qu'un élément

3. A ne pas confondre avec les cas d'utilisation de la méthodologie de développement objet UML.

d'annotation EA673 a été utilisé et mis en relation avec EA145 et EA451. Un graphe potentiel GP23 a également été utilisé au cours de cet épisode.

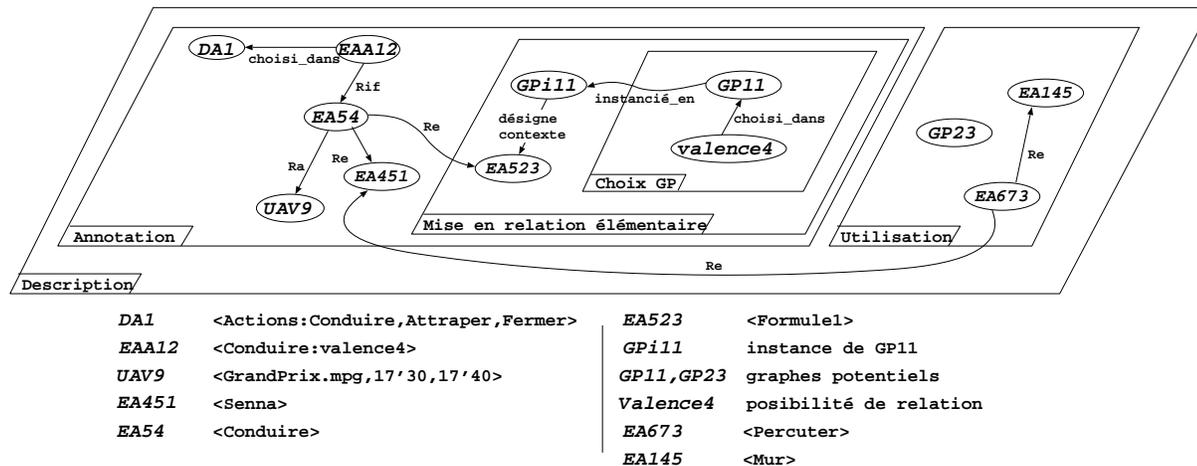


FIG. 8.4 – Un exemple de cas d'utilisation correspondant au modèle de la tâche d'annotation présenté figure 8.3

Explications d'un cas d'utilisation

Les éléments d'un cas d'utilisation sont des instances d'objets décrits dans le modèle d'utilisation et éventuellement dans un modèle de tâche (c'est à dire les tâches et sous-tâches).

Les différents éléments du cas d'utilisation profitent des *relations existantes* (inter- ou intra-modèles) entre les objets de ces modèles, qui fournissent des *explications* à leur co-occurrence. Un cas d'utilisation peut donc toujours s'expliquer par le modèle d'utilisation, et, quand il existe, par le modèle de la tâche courante.

Le cas d'utilisation de la figure 8.4 s'explique ainsi entre autres :

- par les relations internes au modèle d'utilisation : par exemple le fait qu'un graphe potentiel mette en relation potentielle des éléments EA, EAA et UAV est une explication de GP11, de la même manière que la relation d'inscription dans le flux R_{if} entre EAA12 et EA54 est expliquée par les relations générales entre EAA et EA.
- par les relations internes des modèles de tâches auxquels est lié le cas d'utilisation. Par exemple le fait que l'EAA EAA12 ait été choisi parmi les EAA désigné par la dimension d'analyse DA1 pour annoter, ou que le graphe potentiel permettant d'explorer le contexte aurait pu être mis en place de trois manières différentes (si on utilise le sous-modèle de tâche correspondant) sont des explications provenant des modèles de tâches auxquels est lié le cas d'utilisation.

Les relations internes aux cas d'utilisation, par exemple le fait qu'un EA utilisé dans une tâche d'annotation et un EA utilisé dans une tâche générale indéterminée sont en relation participent aux explications internes au cas. Par exemple figure 8.4, la relation entre EA673 et EA451 est une relation interne au cas d'utilisation.

Mémoire du système

La mémoire du système est composée des informations qu'il doit mémoriser en vue de les restituer dans le cadre des différentes tâches de l'utilisateur (nous n'y incluons pas les modèles d'utilisation et de tâches qui sont *a priori* fixés).

Les éléments constitutifs de la mémoire sont donc des instances des éléments décrits dans le modèle d'utilisation. Ils s'expliquent « universellement » selon les relations décrites dans le modèle d'utilisation, selon les relations génériques décrites dans les modèles de tâche et, *in fine* selon les relations spécifiques décrites dans les cas d'utilisation.

Par exemple dans le cadre de l'utilisation des Strates-IA, la mémoire (*cf.* figure 8.5) est composée des UAV, des EA et des EAA, des dimensions d'analyse et des graphes potentiels, *etc.* La mémoire comporte également les cas d'utilisation eux-mêmes. Ceux-ci organisent les connaissances décrites ailleurs (dans le graphe Strates-IA par exemple) en les mettant en relation, c'est à dire :

- soit en rendant saillantes (à l'intérieur du cas) des relations existantes (par exemple la relation élémentaire entre EA451 et EA54 est mise en valeur par sa présence dans le cadre du cas d'utilisation) ;
- soit en y ajoutant des relations découlant des modèles des tâches utilisés (par exemple entre EAA12 et DA1, relation importante qui permet d'expliquer de quelle dimension d'analyse l'EAA et donc l'EA provient).

Tous les éléments de la mémoire s'expliquent mutuellement par les relations directes ou indirectes qu'ils entretiennent.

8.3 Vers une généralisation de l'approche

Les définitions que nous venons de présenter ont un cadre plus large que l'étude des Strates-IA, et la même démarche a été utilisée dans le cadre d'autres projets de recherche tels que présentés dans [Corvaisier *et al.*, 1998, Mille *et al.*, 1999] par exemple. C'est cependant dans le cadre de l'étude des Strates-IA qu'elle a été considérée et étudiée comme objectif en soi.

Nous proposons de généraliser l'approche présentée distinguant modèle d'utilisation (universel), des modèles de tâches (génériques) permettant de mémoriser des cas d'utilisation (spécifiques) comme suggéré figure 8.5. La séparation entre éléments concepts et instances est artificielle et correspondrait dans les Strates-IA à celle entre EAA et EA. Plusieurs types d'éléments peuvent faire partie de la mémoire. Par exemple, dans les Strates-IA les graphes potentiels et les dimensions d'analyse y sont conservés.

Les éléments du modèle d'utilisation s'expliquent par les relations qu'ils entretiennent les uns avec les autres. Les modèles de tâches s'expliquent par les éléments du modèle d'utilisation qu'ils utilisent et par les relations qu'ils entretiennent les uns avec les autres. Les éléments d'un cas d'utilisation s'expliquent par les éléments du modèle d'utilisation qu'ils utilisent, par les éléments du modèle de tâche qu'ils respectent et enfin par les relations internes au cas d'utilisation non expliquées dans d'autres modèles.

8.4 Discussion : exploiter l'expérience

La modélisation proposée d'un système d'information en modèle d'utilisation, modèle de tâche et cas d'utilisation permet de stocker l'expérience d'utilisation du système de telle sorte

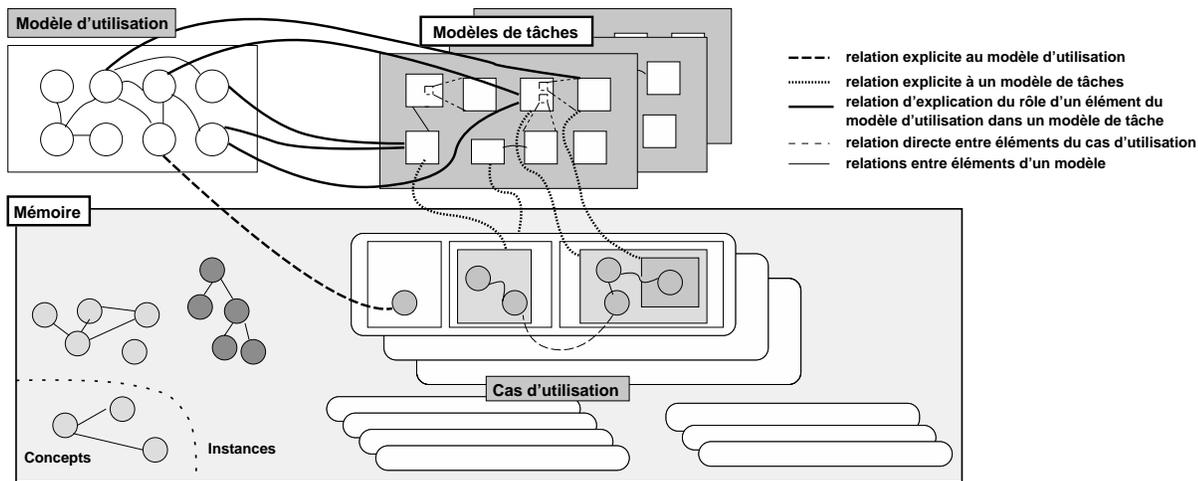


FIG. 8.5 – Synthèse de l'approche : modèle d'utilisation, modèles de tâches, mémoire (comportant des cas d'utilisation) et les explications les mettant en relation.

que celle-ci soit explicable. En effet, les modèles de tâches permettent de rationaliser les cas d'utilisation, c'est à dire à la fois de diriger le choix des connaissances stockées dans les cas et d'expliquer ces dernières, les rendant ré-utilisables.

Plusieurs directions de recherche sont ouvertes pour exploiter l'expérience sous la forme de cas d'utilisation, nous les présentons rapidement dans la suite de ce chapitre.

8.4.1 Les cas d'utilisation pour l'illustration et la validation

Les cas d'utilisation sont des illustrations de l'exploitation des connaissances pour réaliser une tâche. Ces illustrations sont autant d'exemples concrets de l'utilisation de concepts modélisés dans le système, aussi bien les éléments du modèle d'utilisation que les modèles de tâches. *A contrario*, des concepts non utilisés peuvent être remis en cause par leur non-utilisation.

Dans SESAME par exemple, l'utilisation d'un EAA ou d'une valence s'illustre et s'explique directement par un exemple dans la base, c'est à dire par les cas d'utilisation les mettant en jeu d'une manière ou d'une autre. L'utilisation de la valence **valence4** de l'EAA **EAA12** est ainsi illustrée par le cas d'utilisation présenté la figure 8.4.

Dans un cadre plus général, le processus de validation d'une « mémoire d'entreprise » pourrait s'appuyer sur ce type d'approches. Cette façon de procéder revient à valider (ou invalider) les modèles de l'entreprise par leur usage concret.

8.4.2 Apprentissage

Les cas d'utilisation peuvent être utilisés comme support concret d'apprentissage de connaissances nouvelles.

Indications statistiques locales de relations générales

Les relations mises en place par l'utilisateur entre éléments d'un cas d'utilisation peuvent être, par répétition, autant d'indications statistiques de relations plus générales intégrables dans la mémoire.

Dans le cadre de SESAME, des relations élémentaires (épisodiques par définition) répétées dans la mémoire entre EA issus d'EAA identiques peuvent pousser à induire des relations entre ces EAA. Il s'agit d'une sorte de *data mining* validé à l'avance par les différents épisodes d'annotation eux-mêmes (un épisode d'annotation étant un cas d'utilisation dans le cadre d'un système Strates-IA). La relation élémentaire, parce qu'elle autorise toute mise en relation, y compris non prévue, constitue donc un outil puissant d'apprentissage de nouvelles relations.

Intégration d'explications dans les modèles de tâches

Un apprentissage sur les modèles de tâches pourrait s'opérer par renforcement des explications utilisées pour assembler les éléments d'un cas d'utilisation.

La figure 8.6 illustre cette idée : un modèle de tâche A est utilisé pour expliquer la sous-tâche $st1$ d'un cas d'utilisation $CU1$, et une relation d'explication R , non prévue dans le modèle de tâche est utilisée pour mettre en relation un élément de $st1$ à un élément de $st2$. La répétition d'une telle explication dans plusieurs cas d'utilisation peut conduire à la création d'un nouveau modèle de tâche A' , spécialisation de A , qui pourra à son tour expliquer de nouveaux cas d'utilisation, tel que $CU2$, ou contribuer à éclairer retrospectivement $CU1$.

Par exemple, dans le cadre des Strates-IA, les valences pourraient être acquises à partir des graphes potentiels utilisés pour chercher des contextes permettant la mise en place de relation. Ce processus d'apprentissage serait alors en boucle ouverte sur l'utilisateur expert qui devrait valider les candidatures de valences avant de les ajouter. Si la notion de valence fait bien partie du modèle d'utilisation, les valeurs des valences (instances de graphes potentiels) sont clairement associées à une tâche et peuvent être utilisés en tant que tels.

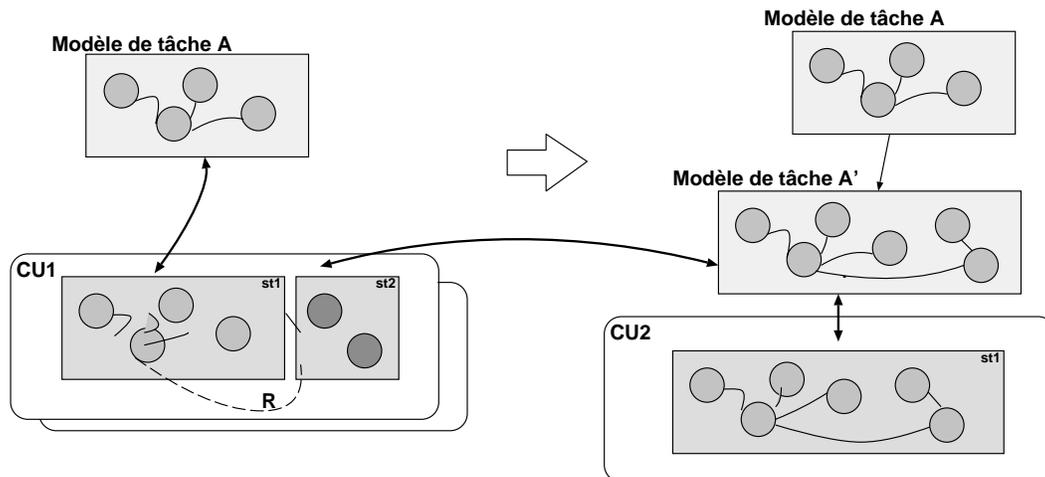


FIG. 8.6 – Intégration d'explication, l'explication symbolisée par la relation R est utilisée pour la mise en place d'un nouveau modèle de tâche A'

Dans une démarche inverse, une différenciation d'explications pourrait être mise en évidence par les explications que l'utilisateur est amené à modifier ou ajouter, conduisant à modifier des modèles de tâche existants.

8.4.3 Aide à l'utilisateur fondée sur l'expérience

Les cas d'utilisation peuvent servir directement à l'assistance à la tâche. Ainsi, on peut imaginer un assistant générique — une sorte de moteur d'assistance — qui, sollicité par un utilisateur

pour l'aider dans une certaine tâche (décrite dans un modèle de tâche) s'y spécialise. L'assistant utiliserait alors la description partielle de la tâche en cours et la base de cas d'utilisation pour proposer des pistes de solutions fondées sur l'expérience de tâches similaires. Cette approche se décline au niveau des sous-tâches, et le paradigme du *raisonnement à partir de cas* (RàPC) [AAmodt et Plaza, 1994] peut être utilisé avec profit pour ce type d'assistance.

On pourrait par exemple imaginer qu'au cours d'une session d'annotation un utilisateur se trouve à court d'idées d'annotation et fasse appel au système. Alors celui-ci (ou plutôt un assistant), analysant la session en cours construirait un cas d'utilisation correspondant à la tâche courante et cherche dans la base de cas d'utilisation un cas similaire, qu'il proposera à l'utilisateur afin de donner des pistes à celui-ci (par exemple « utiliser telle valence »).

Cette façon de voir les choses pourrait faciliter considérablement la réalisation d'assistants pour des tâches mal définies au départ, mais s'appuyant sur un modèle d'utilisation unique. L'outil, même s'il ne prend pas alors le statut de système à base de connaissance pour autant, doit permettre la manipulation explicite des connaissances qui régissent son fonctionnement, de façon à rendre la notion d'« assistant intelligent » possible. A l'instar de [Faron *et al.*, 1996], nous considérons qu'il s'agit de la seule façon de rendre de véritables services « personnalisés » aux utilisateurs.

8.5 Conclusion

Nous avons dans ce chapitre présenté une manière originale de considérer un système informatique, et plus particulièrement un système d'information en s'intéressant à son rôle de mémoire, non seulement comme permettant de stocker les connaissances de description mises en place (les documents), mais aussi l'expérience d'utilisation rationalisée par des modèles de tâche adaptés.

Nous avons ainsi présenté la tâche de description dans un système fondé sur les Strates-IA, et, sur la constatation de la complexité d'une telle tâche, nous avons proposé un cadre de modélisation permettant de distinguer modèle d'utilisation, modèles de tâches et cas d'utilisation, comme traces de réalisation des tâches. Cette approche est d'abord définie et illustrée dans le cadre de la tâche d'annotation décrite précédemment, puis est généralisée. Nous avons enfin justifié cette manière de décrire les tâches et l'expérience d'utilisation d'un système d'information en présentant des pistes de recherche prometteuses d'utilisation des cas d'utilisation comme prise en compte naturelle de l'expérience comme source d'explication et d'assistance pour l'utilisateur.

Le travail présenté ici ne prétend donc pas fournir des solutions à l'aide à l'utilisateur fondée sur l'expérience, mais n'en est que la première étape conceptuelle. Il s'agit alors de proposer un moyen de stocker l'expérience d'utilisation de telle sorte que celle-ci soit expliquée et rationalisée, donc utilisable comme connaissance.

Chapitre 9

Strates-IA : documents et connaissances

Sommaire

9.1	Structure, documents, connaissances et annotations	200
9.1.1	Evolution des structures de présentation	200
9.1.2	Structures sémantiques	200
9.1.3	Ajouter des connaissances par l'annotation	202
9.2	Descriptions documentaires	204
9.2.1	Langages de balises	205
9.2.2	Strates-IA et langage de balises	206
9.2.3	Arbres et graphes	208
9.2.4	MPEG-7	209
9.2.5	Approches orientées « connaissances »	217
9.2.6	Analyse des approches	223
9.3	Strates-IA et connaissances	226
9.3.1	Strates-IA et approches centrées sur les connaissances	226
9.3.2	Ecriture sur le flux	230
9.3.3	Gestion et évolution des connaissances Strates-IA	232
9.4	Conclusion	234

Nous avons présenté les Strates-IA et comment il est possible de les exploiter contextuellement, ainsi que de stocker une expérience d'utilisation expliquée afin d'aider l'utilisateur dans la mise en place et dans l'exploitation d'une base de documents. Il nous faut maintenant étudier cette approche en la situant dans la thématique générale des documents et des connaissances, et mener une importante discussion.

Dans ce chapitre, nous menons d'abord une courte étude des rapports entre documents et connaissances en lien avec les structures mises en place. Nous présentons ensuite quelques exemples d'approches de descriptions et de structuration de documents, notamment au travers des langages de balises, de la future norme MPEG-7 et de réseaux sémantiques permettant de réaliser des inférences. Enfin, nous interrogeons et situons les Strates-IA dans ce contexte, et en développons quelques points d'intérêt.

Une première remarque sur le terme *sémantique* pourra nous aider pour la suite de ce chapitre : ainsi, nous distinguerons entre *sémantique* au sens usuel du terme (liée à l'homme) et *sémantique formelle* (liée à la machine), qui suppose des modèles d'interprétation explicites dans le cadre de systèmes logiques.

9.1 Structure, documents, connaissances et annotations

Nous nous intéressons ici aux structures explicites liées aux documents, audiovisuels ou non, en distinguant tour à tour structures de présentation et structures de connaissances, avant de défendre l'idée que toutes participent de structures sémantiques dont l'utilisation déterminera le statut. Nous nous intéressons alors à l'annotation comme opération d'ajout de connaissances aux documents.

9.1.1 Evolution des structures de présentation

Nous définissons une *structure de présentation* comme une structure documentaire explicite liée à la présentation du document, c'est à dire au calcul de sa forme d'appropriation, à laquelle l'utilisateur est confronté.

Historiquement, la structure de présentation est donc d'abord liée à la structure physique du document, dont la numérisation entraîne la mise en place d'une forme d'enregistrement qui mime cette structure physique. Les premières structures logiques (au sens structure du document numérique) sont à la base des structures de présentation (que l'on retrouve dans les traitements de texte, le multimédia : placement des zones les unes par rapport aux autres), encodées dans des formats de présentation. Pour l'audiovisuel, cela tient par exemple simplement à une suite d'images, à un flux sonore, et à des indications sur la manière de multiplexer le tout.

La phase suivante de l'évolution historique voit la mise en place d'abstractions d'éléments de structure : les tableaux, les listes, les paragraphes, les images en sont quelques exemples. Dans le même processus sont également définies les structures organisant les éléments (ce que l'on retrouve dans les formats de sauvegarde des traitements de textes). De plus, avec la volonté de décrire de façon plus abstraite les documents, sont apparues des possibilités de décrire des *types* de documents, c'est à dire d'abstraire les structures de présentation elles-mêmes : un article est par exemple composé d'un titre, d'un résumé, de sections composées de sous-sections, et d'une bibliographie. Des langages ont alors été conçus permettant de décrire des langages de description de *structures documentaires* en général. Ainsi HTML permet de décrire des hyperdocuments pour le Web, et est un langage instance, une définition de type de document (DTD) du méta-langage SGML.

La possibilité de décrire des types de documents a plusieurs conséquences.

La première conséquence est que la structure documentaire n'a plus obligatoirement de rapport bi-univoque avec la structure physique de présentation du document. Ainsi, d'une structure logique¹, il devient possible de calculer, à l'aide de feuilles de style, plusieurs structures physiques (par exemple une présentation affichant ou n'affichant pas les liens d'un document HTML).

La deuxième conséquence, encore plus importante, est qu'il devient possible de définir des éléments de structure documentaire n'ayant plus de rapport avec la présentation, c'est à dire de mettre en place des « méta-connaissances » dans le document. La balise META de HTML en est un exemple.

9.1.2 Structures sémantiques

Les renseignements liés à la présentation d'un document et étant contenus dans sa forme d'enregistrement sont des connaissances liées à la tâche basique de lecture du document. Nous définissons dans un premier temps les structures de connaissances en opposition aux structures

1. Désormais : telle qu'elle est décrite dans les modèles documentaires classiques.

de présentation, comme l'organisation des connaissances documentaires non liées à la seule tâche de présentation.

Il est en effet possible de considérer des connaissances *sur* le document utilisables dans d'autres tâches que la présentation. Par exemple la balise META est utilisée pour l'indexation de pages Web à l'aide des mots-clé qu'elle permet de définir. Parmi les connaissances *sur* le document figure également tout ce qui peut lui être rajouté, par exemple des structures de connaissances externes au document, et ayant des liens vers certaines de ses parties. Les mots-clé d'Altavista en sont un exemple, mais aussi les graphes de connaissances tels que ceux pouvant être décrits à l'aide de RDF [RDF, 1999] ou d'autres langages. Les « structures sémantiques » [Pinon *et al.*, 1997] proposées pour redoubler la structure logique sont ainsi une tentative de définir un troisième niveau de connaissance (structures physique/logique/sémantique) à l'aide de graphes conceptuels.

De plus, il est maintenant reconnu que les balises liées à la présentation — même les plus banales — décrivent en fait des connaissances qui peuvent être utilisées dans d'autres tâches que celle de présentation. Par exemple, une balise *Titre* exprime deux choses : *titre* comme dénomination qui entraînera une certaine présentation et un arrangement de ce qui est encadré par la balise, mais aussi le fait que ce qui est encadré est un *titre*, c'est à dire une information sémantique utilisable en tant que telle, par exemple pour en extraire des mots-clé *a priori* plus pertinents que d'autres sur le document.

On constate donc que la frontière entre connaissances de présentation et méta-connaissances est en fait relativement ténue, voire s'estompe. Les structures de présentation et les structures de connaissances décrivent toutes des connaissances sur le document, et dans le même temps fondent celui-ci. Ainsi un document accompagné de sa documentation (de structures de connaissances) est encore un document, qui est désormais adapté « nativement » à des utilisations autres que la tâche classique de présentation pour l'appréhension.

Remarquons que cette constatation est également technologique : avec le développement d'XML (eXtended Markup Language, [Michard, 1999]) qui est un langage de balises permettant de définir ses propres balises et leur schéma de structuration, il devient aisé d'inclure des structures de connaissances aux documents, c'est à dire de mêler de façon naturelle structure de présentation et structures de connaissances².

Si la structure de présentation telle que nous l'avons définie pouvait s'apparenter dans un premier temps à la structure logique définie dans la terminologie documentaire classique, laquelle est bien différenciée de la structure sémantique [Pinon *et al.*, 1997], nous préférons considérer que toute structure est *sémantique* et fait partie d'une structure de connaissances documentaires³. L'approche des Strates-IA est un pas dans cette direction : annoter une unité audiovisuelle comme *Plan* revient à ne pas donner de primat à l'information *Plan* comme ce serait le cas si était définie une unité « syntaxique » ou logique *Plan*. Nous considérons donc que toute structure interne ou externe à un document est structure sémantique, laquelle prend son statut lorsqu'elle est utilisée dans le cadre d'une tâche qui *peut* être une tâche de présentation.

Le terme structure de présentation nous semble à cet égard devenir une notion dynamique et opératoire, dont la définition est finalement : « fait partie de la structure de présentation tout élément de structure de connaissance documentaire (sémantique) utilisé pour mettre en place

2. A la limite, les documents XML peuvent devenir totalement structurés, et exprimer des connaissances telles qu'on les trouve dans les bases de données les plus classiques, c'est à dire ayant perdu tout lien avec une structure de présentation documentaire.

3. Nous avons ainsi montré dans le cadre des documents audiovisuels que plusieurs organisations arborescentes pouvaient être considérées pour un même document, correspondant à diverses utilisations, sans que l'on puisse privilégier l'une ou l'autre *a priori*.

une présentation du document ». Dans ces conditions la structure logique pourrait perdre de son importance initiale pour n'être plus qu'une structure de présentation « canonique » extraite d'une structure générale de connaissances.

Il convient pour terminer cette courte étude de se poser la question des structures de connaissances documentaires minimales attachées à un contenu documentaire. Il nous semble qu'il existe minimalement une structure de présentation canonique correspondant au genre du document. Ces connaissances sont mises en place (implicitement ou explicitement) par l'auteur du document, on pourrait donc les appeler connaissances « autoriales » premières. A ces connaissances minimales il est possible d'ajouter toutes les connaissances possibles, de façon *interne*, c'est à dire dans le même langage de description (par exemple XML), ou bien *externe* dans un langage quelconque. L'ensemble de la structure de connaissances documentaires liée au document peut alors être utilisée dans une tâche quelconque (présentation, indexation, recherche, édition, *etc.*).

9.1.3 Ajouter des connaissances par l'annotation

Nous revenons un instant sur la notion d'annotation qui consiste à ajouter des connaissances à un document, représentées de façon interne ou externe, en vue d'une tâche.

De la façon la plus générale, le concept d'annotation est utilisé comme moyen de documenter des parties de documents à l'aide de texte libre. Il est par exemple possible d'annoter des parties de documents Word (sous la forme de « post-it »). La tâche envisagée est alors basique, et consiste simplement en la lecture de ces annotations, dans leur contexte, c'est à dire en regard du texte annoté.

De façon plus élaborée, un certain nombre de travaux s'intéressent aux annotations de documents pour que celles-ci soient partagées dans le cadre d'un travail collaboratif, et s'appliquent aux systèmes d'information et de communication de l'entreprise (par exemple Lotus Notes), ou bien aux documents Web⁴. Les tâches liées aux annotations sont alors associées à leur lecture, mais aussi à la recherche dans celles-ci, et plusieurs types d'annotations peuvent être distingués. Par exemple, dans leur travail sur la station de lecture de la BNF, [Virbel, 1994, Chahuneau *et al.*, 1992] s'intéressent à la lecture savante de documents, et distinguent entre autres les annotations de soulignement, le marquage pour la réutilisation, ou encore la catégorisation de parties (par exemple « Définition »). L'usage principal est alors la navigation dans les annotations, et l'utilisation des liens de références entre parties de documents sur la base d'annotations identiques.

Dans une perspective plus ambitieuse, [Chikh et Tazi, 1999] définissent un système d'annotation de parties de document prédéfinies non seulement pour l'aide à la compréhension de documents (basée sur un *modèle des actes de communication écrite*), mais aussi pour la *reconstruction* de documents à partir des annotations. Il s'agit alors de structurer les annotations elles-mêmes de façon à pouvoir les utiliser pour aider un utilisateur à construire de nouveaux documents à partir des anciens, dans le cadre d'un système de gestion de projet. Une *unité d'annotation* est par exemple définie par :

- une *action* : AttribuerPropriété ; RéférencerContexte ; ProposerAlternatives ;
- un élément du document sur lequel porte l'annotation ;

4. Voir par exemple: Ron Zohar. *Web Annotation — an Overview*. Février 1999. [Publié sur le Web <http://www.ee.technion.ac.il/~ronz/annotation/>].

- des *éléments explicatifs* : propriétés (par exemple rôle ou importance) ; ensemble de concepts [type/nom] (par exemple [Méthode/MERISE]) ; ou variantes et variations possibles du texte ;
- des explications textuelles de l’annotation.

Les annotations ne sont pas liées explicitement entre elles. Une interface permet l’annotation en fournissant des guides, ainsi que de poser des requêtes sur la structure du document et le contenu des annotations.

Les travaux que nous avons rapidement présentés se situent donc dans la perspective selon laquelle il est possible d’aider un utilisateur à exploiter des documents à l’aide des annotations, des connaissances qui leur sont rajoutées. Le dernier système est le plus intéressant au sens où il autorise un utilisateur à documenter un document déjà en place, ou bien le document même qu’il est en train de construire. Cela correspond à la création d’un document non plus uniquement en vue d’une tâche de lecture, mais également de réutilisation et d’écriture, y compris par l’auteur original du document.

Il convient donc d’interroger rapidement la notion d’auteur d’annotations et de lecteur d’annotations, c’est à dire de considérer l’acte d’annotation comme acte de communication entre un auteur et un lecteur, dans le cadre d’un document numérique structuré.

Les questions qui se posent alors sont les suivantes : qui annote, comment, pour qui et pour quelle exploitation ?

Sur la mise en place d’annotations comme connaissances internes, nous pouvons constater que l’auteur est le premier documenteur de son texte, éventuellement de façon inconsciente lorsqu’il utilise un traitement de texte, par exemple en sélectionnant un niveau de titre, ou de façon consciente par exemple s’il indique en note quelle version d’un titre il a écartée.

Concernant la mise en place d’annotations externes, il faut remarquer que la plupart du temps, l’annotation est annotation de parties de documents qui sont déjà désignées, c’est à dire que la structure « auctoriale » est utilisée comme support pour une description annotative qui lui est seconde. Ceci signifie qu’est généralement accepté le fait qu’il existe une analyse primordiale, qui est celle du concepteur du document, qui servira de guide pour l’analyse. Ceci est faux dans le cas général, où l’auteur d’un document ne maîtrise pas toutes les utilisations et exploitations qui en seront faites, à part dans des cas très particuliers. Il n’empêche que le point de vue du concepteur, bien que non définitif, n’en est pas moins très important, et est *entièrement* porté par la documentation et les prescriptions explicites du document.

La liberté de définition des annotations dépend en fait de la tâche que ces dernières doivent pouvoir supporter. Dans le cas de l’audiovisuel par exemple, on peut imaginer que dans un futur proche un auteur mette en place une structure documentant son travail créatif, à partir de laquelle il sera possible d’annoter, par exemple dans un objectif d’analyse de l’œuvre. Il est alors certain qu’il est nécessaire de pouvoir définir d’autres unités d’annotation (UAV dans les Strates-IA) que celles qu’il aura explicitement spécifiées. Dans le cas d’un travail plus précisément défini, comme celui présenté dans [Chikh et Tazi, 1999], où des parties de documents sont clairement documentées et partagées dans le cadre d’une méthode de conception, la liberté d’annotation est plus contrainte.

Nous résumerons le cas général en spécifiant que *l’auteur du document n’est pas forcément coauteur de toutes les annotations* au sens où il en définirait les zones de portée.

Les rapports entre auteur d'annotation et lecteur d'annotation se posent en terme d'exploitation. Par exemple, dans le cas du système LECAO [Stiegler, 1994], le savant qui annote est également son propre lecteur, il met alors en scène au cours de l'exploitation des annotations sa lecture du document en vue d'une relecture, et est libre d'utiliser son propre vocabulaire et ses propres méthodes. Dans le cadre d'une annotation destinée à être exploitée par d'autres (ce qui est le cas de l'indexation dans le cas le plus général) se posent par exemple des problèmes de vocabulaire ou de définitions des annotations, c'est à dire de schéma d'annotation, qui peuvent être plus ou moins partagés, et qui autoriseront une exploitation autre qu'exploratoire.

Pour conclure. Nous avons dans cette partie rapidement discuté les rapports entre documents et connaissances sous l'angle de la structure, et avons ré-interrogé la notion d'annotation dans ce contexte. Il apparaît finalement que :

- toute structure utilisable mise en place dans un document ou sur un document est une structure sémantique de connaissances documentaires, qui peut être utilisée (manipulée) en tant que telle dans le cadre d'une tâche. Toute connaissance peut également servir d'index dès qu'elle permet d'accéder au document (*cf.* 1.2.1).
- la mise en place de la structure est à la base celle d'une structure de présentation auctoriale, mais d'une part l'auteur peut annoter son propre document (structure non liée à la présentation « basique »), d'autre part une annotation peut exploiter la structure mise en place par l'auteur, mais peut tout aussi bien s'en affranchir.

Si nous revenons aux Strates-IA et à l'audiovisuel, deux remarques nous semblent nécessaires.

Ainsi, un document audiovisuel annoté dans le cadre des Strates-IA devient instrumenté par les structures mises en place. Ces structures sont mises en place en vue d'une exploitation basée sur les connaissances d'annotation, et enrichissent une structure de présentation minimale non exploitable en dehors de la visualisation.

D'autre part, dans le cas des Strates-IA, l'annotation est vue comme écriture, définitoire des parties, et prenant en compte au même niveau tous les différents types d'annotation. Tous les schémas de description s'expriment de la même manière, sont normatifs à divers degrés, et peuvent être mis en relation librement. C'est le principe de contextualisation (guidée par des schémas partagés, mais qui peut également être libre) qui permet de préciser les EA en contexte au cours de l'annotation comme à la lecture/exploitation des annotations. On peut d'ores et déjà remarquer qu'on se situe entre une approche stricte contraignant les descriptions possibles (ce qui est cependant autorisé) et une approche libre dans laquelle toute annotation est possible. Nous discuterons quelques intuitions concernant l'écriture et la lecture d'annotation sur le flux dans la partie 9.3.2.

La partie suivante de ce chapitre est consacrée à l'étude de différentes méthodes de descriptions de documents⁵ à l'aide de structures de connaissances, en vue de tâches variées.

9.2 Descriptions documentaires

Nous étudions les approches fondées sur les langages de balise, puis la notion de structure d'arbre ou de graphe pour la description de documents. Nous discutons ensuite la norme en

5. Cette partie vient compléter l'étude bibliographique exposée dans le chapitre 3.

cours de mise au point MPEG-7, avant de nous intéresser aux approches de descriptions orientées « connaissances » et liées aux formalismes de l'intelligence artificielle.

9.2.1 Langages de balises

Intégration de SGML pour la représentation audiovisuelle

[Carrer *et al.*, 1997] présentent un outil permettant de remplir une base de données d'annotation audiovisuelles. L'objectif est de segmenter d'abord le document audiovisuel en plan, avant d'annoter et de regrouper ceux-ci dans une structure arborescente. Les annotations se ramènent à deux types : les *méta-données structurelles* concernent tout à la fois ce qui est spécifique au médium (compression, durée) et la « structure cinématographique » (objets spatio-temporels, mouvements de caméra, plans, scènes, séquences, titre) ; les *méta-données de contenu* permettent de décrire les objets apparaissant dans le flux (chien, disque), et les entités conceptuelles (événements, action, objets abstraits, *etc.*). Les méta-données structurelles (surtout la décomposition hiérarchique) servent de base pour les annotations de contenu considéré comme partagé entre parties. On se trouve donc dans une approche de segmentation relativement classique, et SGML est naturellement utilisé pour représenter le document.

Des éléments de structure et des éléments de contenu sont définis dans une DTD, les éléments de structure étant composés de sous-éléments de structure (hiérarchie Document/Séquence/Scène/Plan) et d'attributs, par exemple *date* ou *mots-clé* pour *Scène*. Le niveau du document contient des éléments objets et sous-objets référencés temporellement, ce qui permet de fournir un support à la stratification. Par exemple un journal télévisé sera annoté en considérant un sujet comme *Scène*, et les objets contiendront des informations sur le *nom* du journal, la *durée*, ainsi qu'un objet *popularité* permettant d'indiquer les passages où des personnages populaires sont présents.

L'intérêt fourni ici par SGML est qu'un parseur peut valider des modèles de documents de façon simple, par exemple s'assurer que les scènes sont composées de plans. L'interface mise en place par [Carrer *et al.*, 1997] (VANE : Video Annotation Engine) s'adapte à la DTD d'un document annoté, et permet surtout de modifier dynamiquement la DTD quand l'utilisateur décide d'ajouter de nouveaux objets ou attributs. A la fin de l'annotation et en vue de l'exploitation de la base, le schéma défini par la DTD est mis en correspondance avec un schéma relationnel permettant de poser des requêtes SQL sur la base.

Formats de description et d'échange

C'est avec des objectifs différents que l'INA a développé un format appelé AEDI (Audiovisual Event Document Interface, [Auffret *et al.*, 1999]), puisqu'il s'agissait au départ de fournir un format permettant l'échange (et donc la comparaison) de résultats de traitements automatiques sur les documents audiovisuels. AEDI est à la base fondé sur le format de représentation interne de l'outil Médiascope développé à l'Inathèque de France⁶. Le format a ensuite évolué en fonction des objectifs des différents projets dans lesquels il était utilisé. Il s'agissait, comme dans l'exemple vu précédemment d'essayer d'étendre les concepts de la publication électronique (formats documentaires à base de langages de balise, ici XML) à la description de documents audiovisuels, de telle sorte qu'une DTD puisse représenter un schéma de description.

Ont alors été rajoutés des types de données à XML (différents du type original et unique PCDATA, purement textuel), puis la possibilité de mettre en place des types d'attributs, et des descripteurs comme ensemble d'attributs valués, lesquels sont situés temporellement dans

6. Une description succincte de l'interface de l'outil se trouve dans [Auffret et Prié, 1999].

des strates, qui peuvent contenir des hiérarchies de segments ou d'autres strates. Localement à une strate peuvent être définis les descripteurs qui seront utilisés dans la segmentation contenue dans cette strate. Les segments et strates du document sont décrits sur un document abstrait différent de ses réalisations physiques, ce qui permet par exemple de faire évoluer le document physique (par exemple passer d'une bande à un fichier informatique) ou d'éliminer les publicités d'un document.

La tentative de décrire des schémas de descriptions et des descriptions à l'aide des possibilités natives de XML (notamment la vérification automatique de conformité des descriptions à une DTD) s'est révélée illusoire. [Auffret et Bachimont, 1999] démontrent ainsi que la description doit être multi-axiale (axe temporel, mais aussi axes spatiaux), et qu'il y a une impossibilité théorique à décrire des documents multi-axiaux dont les contraintes de descriptions soient entièrement contenues dans une DTD. En conséquence de quoi il devient nécessaire de décrire les documents et leurs schémas de description non plus en XML, mais dans un langage de description qui soit lui-même ensuite décrit en XML.

La dernière version d'AEDI permet de définir une description comme arborescence de descripteurs, possédant des bornes sur un ou plusieurs axes (par exemple l'axe temporel) et caractérisés par des attributs qui peuvent eux-mêmes être des descripteurs. Les schémas de descriptions sont des graphes définissant des classes d'objets disponibles et permettent d'exprimer des contraintes supérieures à celles exprimées par les simple DTDs et des parseurs appropriés peuvent être construits, qui valident les descriptions au niveau du modèle de description (utilisant par exemple RDF). On trouvera un exemple d'utilisation de AEDI pour décrire une description abstraite Strates-IA dans [Auffret et Prié, 1999].

9.2.2 Strates-IA et langage de balises

Nous présentons rapidement dans cette partie le travail qui a été réalisé dans le cadre du DEA d'Elöd Egyed-Zsigmond [Egyed-Zsigmond, 1999a] sur la représentation d'un graphe Strates-IA dans un langage de balises dérivé de XML.

Représenter un graphe Strates-IA en XML

L'objectif de cette étude était de mettre en place une représentation d'un graphe de connaissances Strates-IA dans un ensemble de fichiers tels qu'elle convienne aux opérations d'exploitation des Strates-IA.

Les possibilités de construction de modèle se répartissaient suivant les critères suivants.

- La syntaxe choisie : XML en utilisant ou non des éléments avec des attributs, utilisation ou non de XML pour la mise en place de liens et de RDF comme langage déjà construit d'expression de graphes.
- La hiérarchisation des éléments XML, c'est à dire les degrés d'encapsulation directe dans la grammaire de relations du graphes (par exemple, un élément UAV peut contenir un élément EA, ou bien lui être lié par un lien explicité).
- La représentation des liens : ceux-ci peuvent être inclus dans les fichiers décrivant les nœuds, ou bien se trouver dans des fichiers à part, la description faisant appel à des mécanismes ad-hoc ou à la syntaxe normalisée XML.

Sept modèles de représentation ont ainsi été définis, associant à divers degrés les critères que nous venons de présenter. Le choix a été fait de séparer le graphe global en sous-graphes

connectés : un sous-graphe correspond à la base de connaissances, les autres sous-graphes sont liés aux flux audiovisuels étudiés. Un flux correspond ainsi à l'ensemble des unités audiovisuelles qui réfèrent à une même base temporelle, ainsi que les éléments d'annotation qui leur sont liés. Tous les modèles ont été définis et testés à l'aide de la bibliothèque XML4C de IBM.

L'évaluation des modèles se base sur quelques critères de comparaison :

- taille des fichiers C_2 ;
- complexité des requêtes d'accès (mise à jour C_3 et accès pour la recherche C_4) ;
- lisibilité par l'homme C_5 (est facilement lisible ce qui est compact et organisé de façon simple) ;
- richesse sémantique C_1 (est riche sémantiquement ce qui a peu de sous-entendus, et est donc facilement interprétable en l'absence du modèle de données sous-jacent).

	C_1 Richesse sémantique	C_2 Taille fichiers	C_3 Mise à jour	C_4 Recherche	C_5 Lisibilité
M1 : Modèle de base	7	4	3	6	4
M2 : Variante avec EA encapsulés dans les UAV	6	2	3	4	2
M3 : Modèle de base avec attributs	5	3	3	6	3
M4 : Variante attr. + EA encapsulés dans UAV	4	1	3	4	1
M5 : Liens exclus (ad-hoc)	3	6	1	2	7
M5' : Liens exclus (ad-hoc), EA encapsulés	2	5	1	1	6
M6 : Liens exclus (XLL)	1	8	2	3	5
M7 : RDF	2	7	4	5	5

TAB. 9.1 – *Comparaison et classement des modèles selon cinq critères. 1 correspond à la meilleure note et 8 à la plus mauvaise. C'est le critère correspondant aux opérations d'instanciation de graphes potentiels qui est prépondérant chez nous.*

Suivant les différentes exploitations du graphe, les critères ont une importance différente. Par exemple, représenter les liens dans les mêmes fichiers que les nœuds entraîne des opérations complexes de mise à jour des arbres XML dès qu'il s'agit d'ajouter des relations (C_3). Le tableau 9.1 résume les résultats de l'étude pour les différents modèles proposés en fonction des critères de comparaison.

Le critère principal retenu est celui de la rapidité d'exécution de requêtes simples par l'algorithme de multi-propagation : les requêtes portant sur le contenu des nœuds et la recherche de voisins dans le graphes sont donc considérées comme prioritaires (C_4).

Le modèle le plus pertinent correspond alors à une représentation utilisant des liens exclus conformes à la syntaxe XLL (ce qui correspond au modèles $M5'$ dans le tableau 9.1).

Autres connaissances d'exploitation des Strates-IA

L'étude précédente portait sur la problématique générale de représentation de graphes de connaissances dans des langages de balises, à savoir ici le graphe Strates-IA.

Il peut se révéler d'intérêt de représenter les autres connaissances liées à l'exploitation des Strates-IA dans de tels langages, par exemple les schémas de description, les graphes potentiels ou les dimensions d'analyse. Le fait de pouvoir en dernière analyse ramener toutes ces connaissances à des graphes ou des ensembles de graphes permet une représentation homogène de l'ensemble des connaissances du système, qui facilite l'exploitation d'une syntaxe unique (par exemple XML) pour décrire les formats de représentation.

L'approche que nous avons présentée au chapitre 8 de représentation homogène des cas d'utilisation expliqués par un modèle d'utilisation et des modèles de tâches et sous-tâches a également des liens intéressants avec les langages documentaires. Par exemple, un cas d'utilisation peut se concevoir comme un *document* décrivant les éléments instanciés constituant la trace de réalisation d'une tâche. La décomposition d'un modèle de tâche en modèles de sous-tâches permet sa représentation aisée sous une forme documentaire arborescente. Ainsi, dans le cas d'utilisation de la figure 8.4, page 194, le document correspondant serait constitué de la description de la structure de la tâche (notamment l'imbrication de trois niveaux pour la première sous-tâche) caractérisés par les éléments qui y sont instanciés. Des langages de description documentaire comme XML offrent la richesse d'expression suffisante pour exprimer les connaissances ainsi décrites.

9.2.3 Arbres et graphes

Les deux premiers exemples que nous avons présentés montrent une utilisation des langages de balises (SGML ou XML) pour la description et l'annotation de documents audiovisuels. L'enjeu primaire d'atteindre à des descriptions complètement décrites en XML s'est révélé illusoire sauf dans le cas d'applications limitées, basées sur une segmentation stricte du document, aux éléments représentant des segments étant ajoutés des annotations préalablement fixées. L'approche documentaire standard présente en effet des limites : une description d'un document audiovisuel n'est pas toujours un arbre, et fait appel à des contraintes plus complexes que celles exprimées dans une DTD.

Cela est par exemple visible dans AEDI et ses extensions qui se dirigent vers RDF pour la documentation de documents, sur laquelle il s'agira de se base pour calculer leur présentation à l'utilisateur⁷. C'est également le cas de MPEG-7 (voir paragraphe suivant 9.2.4) qui se base sur une simplification et une spécialisation de XML-Schéma [XML-Schema, 1999]. *A fortiori*, la représentation des Strates-IA dans un langage de balise relève du même niveau, en s'abstrayant totalement des modèles structurels liés au langages documentaires. Alors XML n'est qu'un langage permettant une description homogène des modèles de graphes de connaissances, dont toute la syntaxe et la sémantique leur restent propres⁸.

Bien entendu, le résultat final d'une description est *toujours* un arbre XML, mais ce sont les relations, et surtout les schémas de relation entre éléments (décrits hors-DTD) qui représentent la véritable organisation des connaissances sous la forme d'un graphe.

Il nous semble donc (et nous avons argumenté ailleurs dans cette direction) que l'organisation des connaissances de description ne doit pas *a priori* suivre un schéma hiérarchique, ce qui signifie

7. [Auffret et Bachimont, 1999] : « finally we came to the conclusion that content indexing was mostly a task devoted to attaching pieces of meta-information to the document following semantically expressed rules, and that the use of a knowledge representation language was certainly much better suited to this task than plain XML ».

8. Si on utilise un langage dérivé tel que RDF pour la représentation d'un graphe Strates-IA, on rapproche le langage documentaire — mais l'est-il encore — du modèle représenté : RDF sert à représenter des graphes.

que les possibilités de description doivent être suffisamment riches. Le graphe comme moyen le plus général de représentation des connaissances s'impose alors. En d'autres termes, les schémas de descriptions ne doivent pas être plaqués sur les outils de description « technologiques ».

Terminons avec une remarque : l'étude des documents audiovisuels va peut-être réussir à mettre en place ce qui a été *évit * jusque l  pour les documents   forte composante textuelle, c'est   dire remettre en cause la lin arit  de la description et l'arbre sous-jacent h rit  des m thodes d' criture sur papier. Etudier des documents n'ayant pas de structures de connaissance fix es *a priori* autorise en effet   mettre en place des descriptions qui ne leur sont pas obligatoirement redevables⁹. D'autre part et dans ce contexte, la non-obligation de mettre en place des balises dont l'organisation soit arborescente conduit   interroger la notion de balise m me, et pourrait permettre de mener des  tudes non empiriques (comme cela est le cas pour les documents textuels) sur leur statut de connaissances documentaires. Enfin, et par un retour de situation, on peut imaginer que les documents   forte composante textuelle soient  tudi s   nouveau au regard de descriptions lib r es de l'arborescence

De fa on plus prosa ique, la section suivante est consacr e   une  tude des travaux qui se d roulent autour de la norme MPEG-7.

9.2.4 MPEG-7

Le principe g n ral de MPEG-7 vise   mettre en place une norme de description autorisant toutes les descriptions possibles et imaginables pour l'audiovisuel¹⁰, en vue d'en permettre l'acc s et la manipulation. Lanc e en 1997, cette norme devrait th oriquement voir le jour vers en juillet 2001. Apr s un d marrage laborieux, certaines parties sont d'ores et d j  en place, et nous essayerons d'en donner une vision r sistante aux changements qui pourraient appara tre par la suite,   partir de documents publics, mais aussi de documents de travail auxquels nous avons acc s, mais ne pouvons citer. Il convient de garder en m moire cependant que les concepts que nous allons pr senter peuvent  voluer rapidement.

Principes g n raux de MPEG-7

Les concepteurs de MPEG-7 ont pour objectif de prendre en compte l'ensemble des applications imaginables   l'heure actuelle pour les documents audiovisuels voire multim dias : images, sons, liens doivent pouvoir  tre *d crits* de multiples mani res, et la mise en place proc de par accumulation.

D'apr s [MPEG-7-Req, 1999] : « MPEG-7 aims to create a standard for describing the multimedia content that will support these operational requirements [...] the requirements are derived from analysing a wide range of potential applications [...] MPEG-7 is not aimed at any one application in particular, rather, the elements that MPEG-7 standardizes shall support as broad a range of applications as possible ». Il s'agit donc de fournir des langages de description, et des descriptions de base permettant l'utilisation par des applications (recherche d'information,

9. Peut- tre cela pourrait-il vouloir signifier qu'un pas serait alors fait vers une logique nouvelle d'organisation des connaissances documentaires et des documents, adapt e   une vision purement multim dia des documents. Les outils d'exploitation restent alors   inventer pour permettre de « lire » des documents et d'en construire en cours de lecture des structures qu'il soit possible d'appr hender.

10. La norme d passe en fait de loin ce cadre, car ce sont rien moins que les documents multim dias qui sont vis s ! L'ensemble des propositions jusqu'ici propos es se r f rant   des flux temporels continus, nous nous limiterons aux documents audiovisuels dans cette pr sentation.

composition, navigation, *etc.*) dont ne s'occupe pas la norme, pas plus qu'elle ne s'occupe des outils permettant de mettre en place les descriptions.

Les concepts principaux de MPEG-7 sont les suivants [MPEG-7-Req, 1999] :

- données : l'information audiovisuelle ;
- caractéristiques : caractéristique distinctive des données signifiant quelque-chose à quelqu'un (peut être rapproché des objets d'intérêt des Strates-IA) ;
- descripteurs (D) : représentation d'une caractéristique qui en définit syntaxe et sémantique (histogramme de couleur, texte de titre...);
- valeurs de descripteurs : instanciation d'un descripteur ;
- schémas de description (DS) : schéma spécifiant la structure et la sémantique des relations entre ses composants, qui peuvent être des descripteurs ou des schémas de description. Un schéma de description se différencie d'un descripteur au sens où un descripteur ne contient que des types de données basiques¹¹ ;
- descriptions : association d'un schéma de description et de valeurs pour les descripteurs qu'il contient ;
- descriptions MPEG-7 codées : description encodée dans un format exploitable et conforme à la norme ;
- langage de définition de description (DDL) : langage permettant de décrire les schémas de description (et éventuellement les descripteurs structurés), mais également d'étendre des schémas de description.

MPEG-7 fournira donc principalement : un ensemble de descripteurs, un ensemble de schémas de description, un langage permettant de définir des schémas de description. Par la suite, chacun pourra définir de nouveaux schémas de description en fonction des applications visées.

Le langage de définition de description est basé sur XML et ressemble fortement à une simplification de XML-Schéma [XML-Schema, 1999], langage qui est lui-même prévu pour permettre de spécifier des schémas se rapprochant de schémas de base de données pour XML¹².

Les schémas de description (DS) se définissent récursivement à l'aide de liens de « composition »¹³. Par exemple, une définition provisoire de l'*ObjectDS* avec le DDL est la suivante :

```
<DSType name="ObjectDS" model="redefinable">
  <attrDecl name="id"> <datatypeRef name="ID"/> </attrDecl>
  <attrDecl name="href"> <datatypeRef name="uri"/> </attrDecl>
```

11. La distinction entre descripteurs et schémas de description est loin d'être claire et est régulièrement remise en cause. Il semble qu'elle fasse partie des ambiguïtés infantiles du modèle.

12. La spécification de formats documentaires fait *entre autres* partie des objectifs de XML-Schéma ! Ce fait illustre un renversement de situation : à partir d'un langage documentaire (SGML) a été mis en place un langage plus souple (XML), toujours adapté aux documents, mais également à la représentation de n'importe quels types de données (ou connaissances). Structuration documentaire et structuration de données semblent donc en voie de fusionner. Il est possible d'analyser cette évolution par l'apparition de la notion de document bien formé en XML, qui n'existait pas en SGML. Cette notion a amené la possibilité de traiter les instances indépendamment de leur modèle, et lui a ôté de son importance, l'instance devenant le lieu d'expression de toutes les données.

13. Des liens d'héritage sont également possibles afin de spécialiser des DS.

```

<DescTypeRef name="ObjectType" minOccur="1" maxOccur="1">
<DSTypeRef name="AnnotationDS" minOccur="0" maxOccur="1">
  <DSTypeRef name="ObjectDS" minOccur="0" maxOccur="*">
</DSType>

```

Ceci signifie que le DS peut être précisé, possède deux attributs qui sont un identificateur et une URI (Universal Resource Identifier), contient un descripteur de type d'objet (pour l'instant non défini), et peut être lié à une annotation (définie dans *AnnotationDS*) et un ensemble de sous-objets (définis récursivement). Cet exemple montre une partie de la syntaxe du DDL, et comment sont décrits certains liens entre descripteurs et schémas de description à l'intérieur des DS.

D'autres liens entre DS — différents de la composition — peuvent être définis dans d'autres DS, par exemple, les liens entre le *EventDS* et le *ObjectDS* sont définis dans le *EventObjectRelationDS*, spécifiant une relation par un type (qui est un descripteur) et l'identificateur de l'instance de DS point de départ, ainsi que le ou les identificateurs d'instance points d'arrivée. Les liens entre segments sont de type spatial (adjacent, à gauche, etc...), temporel (avant, adjacent, près de...) ou visuel (plus jaune que, même vitesse, même forme...).

Les schémas de descriptions sont prévus pour couvrir l'ensemble des descripteurs et des descriptions imaginables, par exemple certains DS sont consacrés aux segments pour en mettre en place une hiérarchie, d'autres aux primitives image de bas-niveau, d'autres aux personnes, d'autres encore s'intéressent aux différents temps du document et à la manière de les décrire, ou bien aux poids à attribuer aux différentes descriptions (par exemple accorder plus d'importance à une description de texture que de couleurs) ou au type de média considéré (format...). Des descripteurs de « résumé » permettent de prendre en compte des mosaïques de résumé, éventuellement hiérarchiques, ou encore des diaporamas.

Les descripteurs « syntaxiques » appartiennent à *SyntacticDS* qui peut contenir des DS dérivés de *SegmentDS* et des DS de graphes de segments (permettant de les lier, par exemple pour mettre en place une hiérarchie de segments). Les DS dérivés de *SegmentDS* sont les suivants : *VideoSegmentDS* (l'équivalent d'une strate) ; *StillRegionDS* (pour désigner des zones dans une image fixe, extraite ou non d'un flux) ; *MovingRegionDS* (pour spécifier une zone évoluant au cours du temps) ; *AudioSegment* (pour définir des strates sonores). On notera que le *SegmentDS* peut se définir récursivement, c'est à dire qu'il est possible de mettre en place des hiérarchies de segments sans passer par le graphe de relations de segments (à l'aide d'un descripteur de décomposition). De la même manière, les *StillRegionDS* peuvent se décomposer. Les *VideoSegmentDS* et *MovingRegionDS* sont décrits par des *TimeDS* permettant de les situer temporellement.

Les descripteurs « sémantiques » sont de plusieurs types. Le *SemanticDS* décrit les « notions sémantiques » apparaissant dans le flux, par exemple *voiture* et se spécialise en objets et événements. Les *EventDS* décrivent les notions ayant une extension temporelle et peuvent être décomposés (par exemple *match de tennis* se décompose en *sets*, *jeux* et *points*), tandis que les *ObjectDS* sont consacrées aux notions ayant des extensions spatio-temporelles. Les liens entre éléments sémantiques sont gérés par un graphe, de même que les liens entre notions sémantiques dans le *SemanticDS* (les liens sont repris de WordNet [Fellbaum, 1998] et décomposés en liens lexicaux, prédicatifs et d'équivalence). Les liens du *SyntacticSemanticDS* permettent de lier des segments aux éléments sémantiques (objets, événements, relations objets-événements), et signifient dans un sens des liens d'instance audiovisuelles d'éléments sémantiques, dans l'autre des liens d'interprétation sémantique d'unités syntaxiques. Ce DS représente un premier effort de liaison entre unités syntaxiques et sémantiques, et fait partie des originalités de MPEG-7 par rapport aux modèles précédents, tels que ceux que nous avons vus au chapitre 3.

A cela s'ajoute le *MetaDS*, qui permet de décrire ce qui se passe dans le document (ou même dans un segment quelconque...) selon les « 6W » (Who?, Where?, What?...), de mettre en place des annotations textuelles, des lieux, des dates, des publics visés, *etc.*, ainsi qu'un *ContentDS* liant descriptions, 6W, descriptions, objets, personnages...

La figure 9.1 contient un exemple indicatif de décomposition des schémas de descriptions à partir du schéma de description primaire *GenericAVDS*. Cet exemple, librement inspiré des travaux en cours, n'a d'autre objectif que de présenter les principaux concepts de MPEG-7, et de nombreux schémas de description ne sont pas décrits, tandis que d'autres sont simplifiés en descripteurs (par exemple les possibilités d'exprimer des liens).

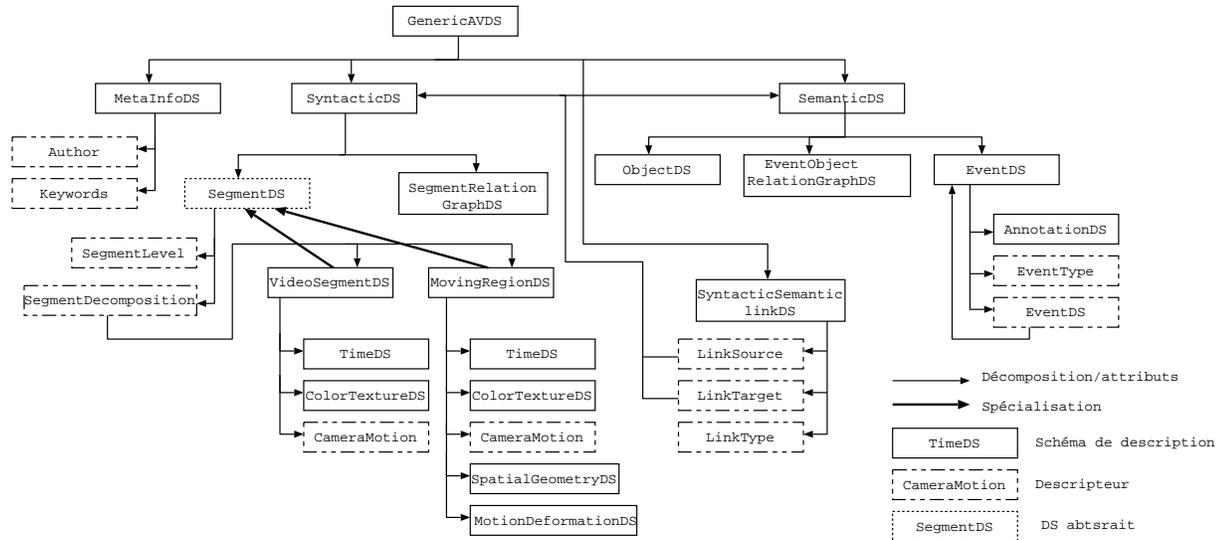


FIG. 9.1 – Un exemple simplifié de décomposition du schéma de description principal de MPEG-7 *GenericAVDS*.

Un exemple de description suivant le *GenericAVDS* de la figure 9.1 est donné figure 9.2.

Nous n'irons pas plus loin dans la description des travaux en cours de la future norme. La suite de cette partie est consacré à quelques remarques générales sur l'approche MPEG-7, et à une rapide étude des liens possibles avec les Strates-IA.

Synthèse et analyse de la situation actuelle

L'objectif à la base de MPEG-7 était de fournir des moyens de décrire des flux audiovisuels, mais aussi de préciser comment les décrire en termes de contenu. Les concepteurs se sont rendu compte que cela était impossible, et proposent donc un langage de description (DDL) et un ensemble de schémas de description fournissant des exemples de manière de décrire. La liberté n'est cependant pas totale : un schéma de description personnel doit être une spécialisation d'un schéma de description fourni, lequel inclue des « guides » de description. Par exemple, un *StillRegionDS* est composé de zéro ou plusieurs *Color/TextureDS*, ou alors le *PersonDS* spécifie de manière stricte comment décrire une personne, y compris sa « race ». Les schémas de description ne sont donc pas indépendants entre eux, il y a dans la norme une incursion dans le domaine de l'utilisation.

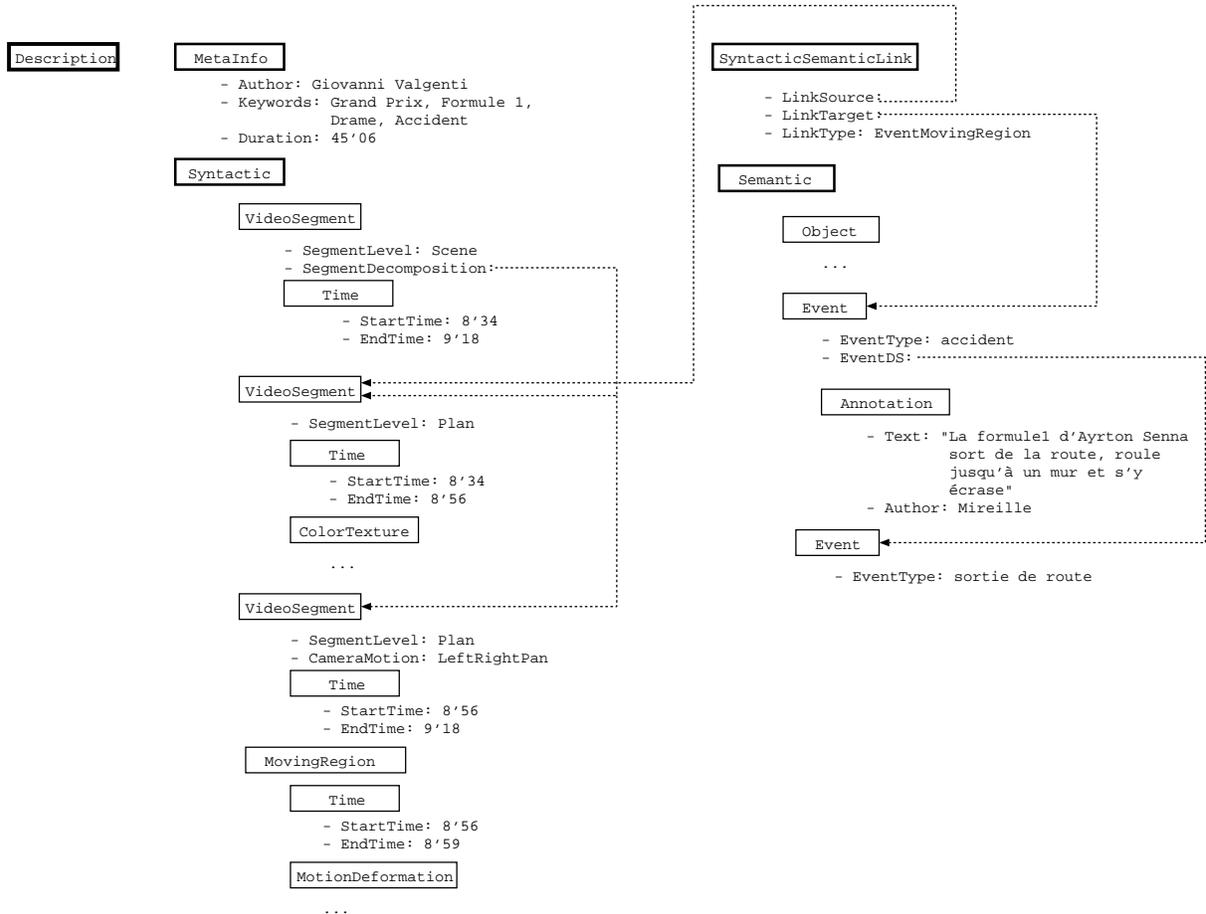


FIG. 9.2 – Un exemple de description MPEG-7. La description syntaxique contient une scène et deux plans, la description sémantique contient deux événements, dont l'un est lié au flux en utilisant une relation syntaxe/sémantique.

L'approche générale de MPEG-7 oscille donc entre une volonté prescriptive (afin d'obtenir une norme utile et effectivement utilisée) traditionnelle du groupe MPEG, et une volonté de généralité pour permettre de prendre en compte l'ensemble des applications visées. Le mode de travail, qui consiste à essayer d'affiner peu à peu des schémas de description issus d'un regroupement préalable de l'ensemble des manières de décrire proposées au début des travaux n'encourage pas vraiment à avoir une vue d'ensemble. Les descripteurs proposés dans les divers schémas de description se recoupent donc relativement souvent.

Cela est particulièrement le cas entre les schémas de description *MetaDS* et *SemanticDS*, dont les rôles respectifs sont peu différenciés. Par exemple, un méta-DS permet de définir le contenu d'un segment, quel que soit celui-ci, en utilisant par exemple des objets issus de *ObjectDS* — spécialisation de *SemanticDS* — qui peuvent à ce titre être liés avec des segments en utilisant le graphe de relations syntaxique/sémantique. La définition de ce qui se passe dans le flux peut donc être mise en place à tous les niveaux, sans que les rôles soient bien précisés.

D'autre part, les liens entre concepts utilisés en tant que descripteurs sémantiques du flux (par exemple *but*) et leurs éventuelles instances sont loin d'être totalement clarifiés. Par exemple, *but* doit n'être utilisé qu'une seule fois, et il n'est pas possible de spécifier des « instances de but » telles que *but1* et *but2*, qui pourraient être mis en relations avec des descripteurs liés à leurs auteurs. Dans l'esprit de la norme, le lien entre *but* et *but1* et *but2*, doit être réalisé en mettant en relation syntaxique/sémantique *but* comme descripteur sémantique avec ses diverses occurrences « syntaxiques » (les images) dans le flux.

En fait, les concepteurs de la norme essaient de mettre en place des liens entre d'un côté les descripteurs et schémas de description « syntaxiques », considérés comme objectifs (car issus du traitement d'image, ou bien universellement partagés, par exemple les plans), et des descripteurs sémantiques, qui décrivent ce qui se passe dans le flux audiovisuel. Les principaux problèmes découlent d'une grande difficulté à différencier ce qui est syntaxique de ce qui est sémantique, l'objectivité des uns étant la subjectivité des autres, certains descripteurs pouvant se retrouver à plusieurs niveaux.

Le mélange des genres nous semble à cet égard relativement dangereux, car s'appliquant à ce qui est le plus difficile à gérer, à savoir le niveau sémantique et symbolique des descriptions. Ce niveau risque pourtant d'être celui qui aura le plus d'importance en termes de recherche et de manipulation de documents audiovisuels indexés, au regard des descriptions par exemple issues du traitement du signal.

La volonté de normaliser autre chose qu'un langage de description minimal pour l'audiovisuel¹⁴ (c'est à dire permettant de lier des descripteurs à des morceaux de flux, quels que soient les uns et les autres) conduit les concepteurs de MPEG-7 à essayer de fournir dans un même élan normatif à la fois une syntaxe de l'audiovisuel, une sémantique, un sens commun et des genres principaux, au risque de n'arriver à rien. Ce genre de travail a par exemple été accompli par SGML puis par la TEI¹⁵, en de nombreuses années, sur un matériau bien mieux connu.

Il nous semble donc que c'est au prix d'une simplification relative de la volonté de description, c'est à dire en fournissant *a minima* un langage universel permettant de décrire des descriptions de flux audiovisuels, tenant compte de ses particularités spatio-temporelles, que les résultats

14. Ainsi que la volonté de s'attaquer également aux contenus multimédias, et de manière générale à un peu toute la connaissance imaginable : le *MultimediaDS* devrait contenir un *HumanBodyDS* pour décrire les parties du corps et leurs relations ! (lesquelles devront en toute logique être connectées à des actions, à des commentaires...)

15. Qui était à la base conçue pour un usage uniquement universitaire, sans ambition universelle, et est reprise dans d'autres domaines.

pourront avoir une utilité. Alors leur utilisation dans un certain nombre d'applications permettront de faire émerger des pratiques et des genres liés à l'audiovisuel « totalement indexé ».

Notons pour terminer que les retombées de l'utilisation de MPEG-7 pourraient se faire sentir au niveau du traitement du signal :

- la mise en place d'un format d'échange de résultats¹⁶ permettra de diffuser et de comparer facilement les algorithmes et d'en évaluer la pertinence ;
- l'utilisation de primitives de bas-niveau en même temps que de caractéristiques de haut-niveau devrait rapidement permettre de mesurer leurs utilités respectives, et fournir un cadre de test pour la mise en place de caractéristiques interprétables.

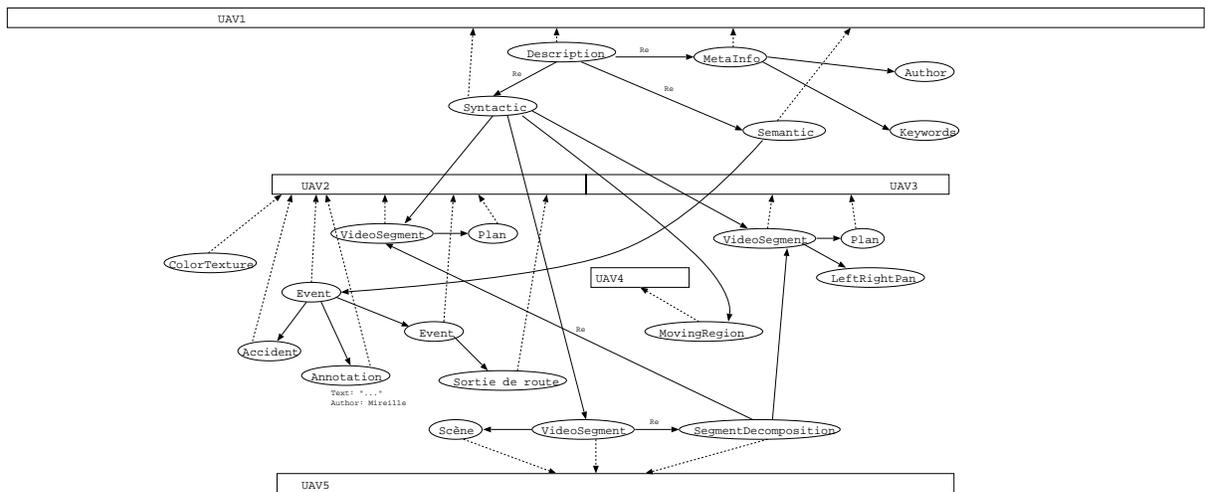


FIG. 9.3 – Exemple de transposition d'une description MPEG-7 dans les Strates-IA.

MPEG-7 et Strates-IA

Nous discutons les rapports entre Strates-IA à MPEG-7 sur quelques points.

Les Strates-IA fournissent tout d'abord un mécanisme d'opérationnalisation des descriptions MPEG-7. En effet, toute description MPEG-7 d'un document audiovisuel, toute instance de schéma de description ou de descripteur est liée, d'une manière ou d'une autre à une partie — éventuellement spatialisée — d'un flux temporel. Cela est bien entendu le cas pour les segments de façon directe (lorsque ceux-ci sont liés à une instance de *TimeDS*), mais aussi de façon implicite, pour toutes les descriptions qui leur sont liées. Par exemple, une description globale est liée au flux comme tout ; une description sémantique liée à un événement peut-être considérée comme globale, et liée à une description locale, ou bien être elle-même locale, *etc.*

Il apparaît donc qu'il est possible de considérer tout lien effectif au flux comme lien vers une unité audiovisuelle étendue¹⁷, et de définir les instances de schémas de description comme éléments d'annotation, les diverses relations entre éléments de MPEG-7 étant explicitement décrits à l'aide d'EA de relation. Il s'agit donc de « mettre à plat » les descriptions MPEG-7, en exprimant toute description d'annotation sémantique (au sens où nous l'avons défini, de façon très

16. Ce que voulait être AEDI à ses débuts.

17. Aux objets à l'intérieur des images, ou aux objets en mouvement, cf. 5.2.1.

large), à l'aide d'un ou de plusieurs éléments d'annotation, situés dans le flux au moyen des unités syntaxiques minimales que sont les UAV étendues. Les relations de composition entre schémas de description MPEG-7 sont représentées explicitement comme relations Strates-IA, de même que les relations explicitement décrites dans les graphes de relations MPEG-7.

La base de connaissances Strates-IA correspondant à une telle opérationnalisation devrait être construite à partir des schémas de description MPEG-7 effectivement utilisés.

La figure 9.3 présente un exemple de transposition de la succincte description présentée figure 9.2 dans les Strates-IA. Cinq unités audiovisuelles sont mises en place, correspondant au flux dans son ensemble, aux trois *VideoSegment* et au *MovingRegion*. Les instances de schémas de description sont linéarisées, c'est à dire représentées à l'aide d'éléments d'annotation et de relations (toutes ne sont pas décomposées totalement : ainsi, la relation entre $\langle Event \rangle$ et $\langle Accident \rangle$ devrait par exemple être explicitée à l'aide d'un EA $\langle EventType \rangle$, comme c'est le cas pour la décomposition de $\langle VideoSegment \rangle$). On remarquera que divers choix de transposition sont également possibles en termes de descripteurs, par exemple il est possible de les passer en EA (cf. *EventType*), ou bien de les garder comme attributs d'EA (cas des attributs *Text* et *Author* de $\langle Annotation \rangle$ par exemple). Les liens syntaxe/sémantique sont considérés comme toujours explicités à l'aide de relation d'annotation entre EA (forcément sémantiques) et UAV étendues.

Les Strates-IA offrent donc un moyen de transposer les descriptions MPEG-7 en liant de façon explicite tous les descripteurs au flux. Tous les éléments de description sont également remis à un même niveau, qui est celui des éléments d'annotation. Dans la figure 9.3 par exemple, les descripteurs image et sémantiques annotant une même UAV se trouvent contextualisés directement et temporellement, ce qui n'était pas le cas dans la description MPEG-7. Les mécanismes d'exploitation contextuelle des Strates-IA peuvent alors être utilisés pour mettre en liaison descriptions de haut et de bas-niveau conceptuel, pour rechercher des morceaux de flux ou des descripteurs spécifiques, *etc.*

Les Strates-IA fournissent un moyen de contrôler l'annotation sous la forme de schémas de description, lesquels existent également dans MPEG-7. Les principes en sont cependant différents : les Strates-IA s'appuient d'une part sur une base de connaissances décrivant les éléments d'annotation et leurs attributs, d'autre part sur les schémas de description comme spécification d'utilisation d'éléments d'annotation et de leurs relations. Les schémas de description MPEG-7 organisent toute la connaissance d'annotation (comme descripteurs ou schéma de description) et ses contraintes de relation à la manière de schéma XML, c'est à dire en faisant un usage appuyé de la notion de composition, laquelle notion peut exister dans les Strates-IA, mais doit explicitement être décrite à l'aide d'EA.

En tant que format d'échange, MPEG-7 pourrait également permettre de décrire les Strates-IA. On pourrait par exemple imaginer que chaque EAA soit exprimé comme un schéma de description MPEG-7, la base de connaissances étant un DS MPEG-7 mettant en relation les EAA. Alors les UAV pourraient être décrites comme unités syntaxiques minimales dans un DS MPEG-7 adéquat, les EA étant les instances de DS EAA. Dimensions d'analyse et schémas de description pourraient également être décrits comme DS MPEG-7 liés à des descriptions particulières.

Il apparaît au final que les points communs entre MPEG-7 et Strates-IA sont les suivants : dans les deux modèles, l'annotateur est autorisé à définir lui-même la manière dont il va décrire les documents ; les prescriptions proviennent des schémas de description ; l'objectif est de décrire le maximum de connaissances de tout niveau, de façon extensible.

Au chapitre des différences, hormis la différence d'échelle, notons qu'en tant que future norme, MPEG-7 vise à autoriser tous les usages dans un format de description unique, les schémas de description MPEG-7 faisant office de points de départ et de briques de construction d'autres schémas de description, ceux-ci jouant le rôle de contraintes sur la description, à l'image des DTD de XML. Dans le cas des Strates-IA, les schémas de descriptions sont beaucoup plus proches des annotations elles-mêmes, au sens où ils peuvent en être extraits aisément, en fonction des usages avérés. Les Strates-IA définissent un usage d'annotation générique (EA, UAV, relations entre EA), toutes les annotations étant mises au même niveau en vue de pouvoir modifier et compléter aisément annotations et schémas.

Bien que ce ne soit pas un des objectifs majeurs de l'approche des Strates-IA, qui se base sur une relative évolutivité du système en fonction de son utilisation, et la réutilisation de l'expérience à tous les niveaux, il est certain qu'en fonction d'une utilisation avérée, des schémas de description stables *différents des schémas de description initiaux* pourront émerger et éventuellement être réutilisés de façon plus ou moins figée dans d'autres systèmes.

Nous présentons dans la section suivante quelques approches fondées sur les connaissances pour la représentation de document, c'est à dire faisant plus ou moins explicitement appel à des techniques d'Intelligence Artificielle (IA), notamment au concept fondamental d'inférence.

9.2.5 Approches orientées « connaissances »

Nous allons dans cette partie nous intéresser aux aspects liés aux connaissances documentaires, en tant qu'elles sont connaissances, c'est à dire impliquées dans la réalisation d'une tâche de raisonnement ou d'inférence automatiques permettant l'exploitation des documents.

Les quelques approches que nous allons présenter représentent des connaissances documentaires à plusieurs niveaux, que nous divisons grossièrement en deux groupes.

Dans le premier cas, il s'agit de connaissances hors-documents, qui représentent une réalité hors-documents (mais qui en est éventuellement tirée) sous la forme de graphes de connaissances à partir desquels il est possible d'accéder aux documents. Ainsi, [Agius et Angelides, 1999] présentent une organisation de connaissances zoologiques comme objets ayant diverses facettes, permettant d'accéder à leurs réalisations documentaires, tandis que [Nanard et Nanard, 1995] décrivent un réseau de connaissances techniques permettant de s'orienter dans une documentation.

Dans le deuxième cas, l'organisation de connaissances hors-documents se limite à des thésaurus ou bases de connaissances décrivant les termes et la manière de décrire les documents à l'aide d'organisation de ces termes en réseaux sémantiques « instances ». Ces réseaux sont alors créés comme descriptions de documents particuliers [Genest, 1999], ou bien intégrés [Chiarabella, 1997] dans la description documentaire complète.

Graphes de connaissances hors-documents

Connaissances pour l'enseignement. [Agius et Angelides, 1999] présentent une approche de description de connaissances audiovisuelles utilisées dans un système multimédia d'apprentissage. L'objectif est donc de décrire des connaissances factuelles, et de les illustrer à l'aide d'extraits audiovisuels.

Les auteurs définissent des unités syntaxiques au niveau des flux audiovisuels, qui serviront de points d'ancrage dans le document. Ainsi, les unités *syntaxiques* sont des séquences (strates quelconques), dont on peut considérer les composantes vidéo et audio, et des informations concernant

les objets visibles (voire invisibles mais présents) dans une image (ou dans un plan si la situation ne change pas) et les arrangements spatiaux entre objets exprimées à l'aide d'un ensemble de relations primitives (sur, sous, à droite, devant, *etc.*). On aura par exemple *objets: chat (30,30,50,70), télévision (0,0,30,80)* exprimant les positions à l'image de deux objets, ainsi que des relations spatiales telles que *(chat) [sur] (télévision), (perso1) [sous] (télévision)* ou encore *(chapeau) [sur] (perso1), etc.* Les unités syntaxiques serviront à faire le lien entre le flux et les unités sémantiques. La description des objets des images et de leurs relations spatiales ne semble pas utilisée, les unités syntaxiques se limitent dans la suite de l'article à des strates temporelles.

Les *entités d'intérêt* représentent les connaissances du domaine du système d'enseignement. Une entité d'intérêt regroupe trois type d'unités sémantiques : une entité *description*, une entité *événement* et une entité *action*. Les unités sémantiques contiennent des éléments pouvant être liés aux flux audiovisuels, *via* les éléments syntaxiques.

Si on considère par exemple l'entité d'intérêt *Léopard* (nous sommes dans un système d'apprentissage consacré aux animaux), alors celle-ci regroupera les unités sémantiques suivantes :

- une unité de description décrivant l'entité d'intérêt *Léopard*, spécialisation de *Félin*, ayant 4 pattes et vivant dans la savane ;
- une unité sémantique liée aux événements associée à l'entité d'intérêt qui par exemple une facette *Capable de*, avec les valeurs *Chasser* et *Tuer*.
- une unité d'action regroupera les actions par événements, par exemple la facette *Chasser* aura pour valeurs *Observer*, *Attraper*, *Etouffer*, et la facette *Tuer* sera associée à *Attraper* et *Etouffer*.

Les attributs de toutes les unités sémantiques peuvent être liés au flux, c'est à dire à des ensembles d'unités syntaxiques (par exemple *Etouffer* peut être lié à tous les plans contenant des scènes d'étouffement¹⁸). D'autres attributs, tel *Savane* peuvent n'avoir pas de liens aux flux, ou avoir des liens de connaissances *hors-flux* tel le lien de spécialisation entre l'unité de description de nom *Léopard* et l'unité de description de nom *Mammifère*.

A partir du modèle de représentation de connaissances hors-documents que nous venons de présenter, les auteurs utilisent des réseaux de représentation conceptuelle du domaine visé dans un système d'apprentissage pour jeunes enfants (consacré au animaux), que nous ne détaillerons pas. Ce qui nous intéresse ici est en effet le modèle de représentation, pour lequel l'enjeu principal n'est pas de mettre en place une structure documentaire, mais bien de lier des connaissances conceptuelles abstraites à des « illustrations » contenues dans le flux. Ce sont d'une certaine manière les connaissances encyclopédiques (objet de l'apprentissage) qui sont illustrées à l'aide de connaissances contenues dans le flux par l'intermédiaire des unités syntaxiques de description.

Connaissances pour la recherche de documentation. [Nanard et Nanard, 1995] mettent en place un réseau sémantique de connaissances factuelles liées à des documents sur la maintenance d'installations nucléaires. La base de documents est alors transformée en hypertexte, dans lequel se retrouvent plusieurs types de connaissances : connaissances sur le contenu des documents, connaissances générales sur le domaine, et connaissances sur la tâche (offrant une utilisation du réseau sous la forme de scripts). La structure hypertexte est alors un graphe

18. Les auteurs notent très justement qu'alors les plans sont mis en relation par les attributs d'unités sémantiques qui les concernent, ce qui est l'équivalent de ce que note Stiegler en considérant le travail d'annotation aidé du chercheur.

étiqueté orienté qui peut être parcouru automatiquement ou manuellement afin d'accéder aux documents.

Le réseau sémantique de description est obtenu en définissant un certain nombre de classes organisées en hiérarchies et connectées par des relations. Par exemple, la relation *est-né-dans* pourra être utilisée entre la classe *Ville* et la classe *Humain*). Les concepts (instances de classes) peuvent être définis et mis en relation, ce qui permet de décrire un graphe de connaissances factuelles (par exemple « Mickaël Jackson est né à Chicago »). L'ensemble de ces connaissances générales est mis en place à partir de l'ensemble des documents et d'un expert.

Les documents eux-mêmes sont décrits dans un formalisme documentaire tel que SGML (+ HyTime), qui permet de définir des ancres (des termes de documents) et leur contexte (une partie de document contenant l'ancre et considérée comme minimale pour permettre de l'expliquer). Les ancres sont liées aux concepts à l'aide de liens typés, exprimant une qualification du contexte utilisé, par exemple *Définition*, *Exemple*, *Exception*, etc. Un lien entre concept et ancre met donc en jeu un contexte documentaire dont le rôle au regard du terme-ancre est qualifié par le type du lien. Par exemple, une ancre peut être *cette valve*, liée au concept *Valve VZ48*, dans le contexte d'une phrase, et le lien peut être qualifié par *Règle générale* explicitant comme il convient de considérer le lien au document du concept, ici comme règle générale qui lui est attachée.

Les utilisations les plus basiques du système concernent la navigation dans les documents en utilisant les liens explicites hypertextes, mais aussi la navigation prenant avantage du réseau de connaissances (dans les connaissances, des connaissances vers les documents, des documents vers les connaissances).

Un autre point développé dans [Nanard et Nanard, 1995] est la possibilité de mettre en place (sous la forme de scripts associés aux concepts) des *expressions de chemins sémantiques* spécifiant des chemins dans le graphe liant concepts entre eux et concepts et ancres. Ainsi la navigation à partir d'un concept peut consister à lancer le script associé et à naviguer jusqu'à l'extrémité du lien¹⁹. Le « modèle de tâche » du système (ici la tâche de lire des documents dans un contexte professionnel de recherche de documentation technique), est lié à la base de connaissance et à l'ensemble des scripts mis en place dans les concepts. Un script peut par exemple être lancé après qu'un utilisateur ait navigué d'une partie de texte vers un concept qui lui est lié (*e.g.* valve). S'il utilise alors un script « donne moi les informations procédurales concernant ce concept », le système lui proposera un ensemble de liens vers des parties de documents calculées à l'aide du script, comme s'il avait utilisé un lien « informations procédurales concernant cette valve ». L'avantage ici par rapport à des liens pré-calculées est que les scripts peuvent être hérités entre concepts d'une part, d'autre part s'appliquent sur l'état actuel du graphe de connaissances, donc prendra éventuellement en compte des ajout à celui-ci.

De façon plus générale, les scripts permettent de définir des documents virtuels, en tant que connaissances dynamiques de construction (au contraire des DTD) : un document est construit en associant des connaissances structurelles (éventuellement conditionnelles et évaluables) et des contenus pré-déterminés ou résultats d'application de chemins sémantiques.

Graphes conceptuels

Nous présentons ici quelques approches de description utilisant explicitement les graphes conceptuels [Sowa, 1984, Sowa, 1991, Mugnier et Chein, 1996] pour la représentation de connais-

19. Cela pourrait s'apparenter dans les Strates-IA à l'utilisation d'un graphe potentiel à partir d'un élément d'annotation abstrait. Les graphes potentiels associés aux valences peuvent alors être candidats à être utilisés comme graphe potentiels de navigation.

sances sur les documents.

Rappelons rapidement ce qu'est un graphe conceptuel. Un système de représentation en graphe conceptuel est composé d'un support, qui est un ensemble de types de concepts organisés en treillis, et d'un ensemble de types de relation, également organisés en treillis. Les relations permettent de lier n instances de concepts (qui peuvent être générales, par exemple $[Humain: *]$ ou particulières à un réfèrent $[Humain: Jean]$), par exemple on pourra avoir $[Humain: Jean] \rightarrow (Agent) \rightarrow [Action: Marcher]$. Un graphe conceptuel est un ensemble de concepts connectés à l'aide de relations. Il permet d'exprimer des connaissances de façon facilement lisible par l'homme, mais aussi de raisonner, au moyen d'opérations telles que la spécialisation, les joints internes et externes, la généralisation, qui s'expriment dans les termes de l'opération de *projection*. Toutes les connaissances et les manipulations de graphes s'expriment également dans les termes de la logique du premier ordre²⁰, ce qui permet d'assurer la validité des raisonnements en terme de logique standard. La relation *est-un* qui sert à organiser les treillis du support est à cet égard fondamentale et strictement définie selon les inférences qu'elle permet de réaliser : tout ce qui est valable pour un concept le sera également pour tous ses concepts fils (au niveau des types).

Le support définit donc une base de connaissances définissant des intensions d'objets du monde et les manières de les mettre en relation, et les faits et connaissances, utilisant des extensions de concepts et leurs relations sont exprimés dans les graphes conceptuels. Les inférences se ramènent à des vérifications que certains graphes dérivent logiquement d'autres (subsumption), et à des manipulations de graphes par jonction afin d'en déduire de nouvelles connaissances.

Les trois travaux que nous présentons dans la suite sont des exemples d'utilisation des graphes conceptuels pour l'exploitation de documents.

Structuration de mots-clé L'objectif de [Genest, 1999] est d'exprimer des connaissances d'indexation en organisant des termes de description explicitement. Ces termes correspondent à ceux utilisés dans le langage documentaire Rameau, qui définit un thésaurus ayant des renvois d'équivalence (tel terme est utilisé pour tel autre), des relations hiérarchiques et des relations d'association. L'utilisation du langage est codifiée (pré-coordonnée) : l'ordre des termes utilisés pour décrire un document n'est pas neutre. Un terme principal pourra être précisé par des *subdivisions de sujet* en objets, lieu et temps, par exemple :

*médecine ** aspect social ** France ** 19ème siècle*

Le principe proposé dans [Genest, 1999] consiste à représenter les descriptions à l'aide de graphes conceptuels, ce qui permettra d'une part d'autoriser des relations plus explicites entre termes et subdivisions, mais aussi d'organiser les termes de description en réseau, afin d'augmenter les capacités d'expressivité du langage. Un support est alors mis en place, avec des types de relations *Objet*, *Lieu*, *Temps*, *Agent*, *Influence*, etc. Les types de concepts sont extraits de Rameau, et il est possible de mettre en place une transcription automatique des descriptions actuelles en description en graphes conceptuels, en utilisant simplement les trois types de subdivisions comme relations.

La recherche consiste alors à déterminer un graphe requête comme graphe conceptuel, et à chercher dans la base des graphes qui en soient des projections (ce qui permet de prendre en compte la relation de subsumption portée par la relation *est-un* entre types de concepts). Sur la constatation qu'une telle recherche est source de silence (on ne cherche pas les graphes répondant partiellement au graphe requête), l'auteur propose d'adapter le modèle logique de

20. En considérant grossièrement les nœuds comme des prédicats et les liens comme des opérateurs logiques.

recherche d'information et son principe de transformation²¹ aux graphes conceptuels. Certaines opérations sont alors proposées qui permettent de transformer le graphe requête, principalement par joint, changement de types de sommets concepts ou relations, *etc.* La manière de mener efficacement ces recherches n'est pas encore étudiée.

Lier des graphes conceptuels à des parties de documents Dans le cadre d'un outil d'acquisition de connaissances (sous la forme de graphes conceptuels) à partir de textes, [Martin et Alpay, 1996] étudient les moyens de représenter des graphes conceptuels dans un langage de balise, mais aussi et surtout de décrire des parties de documents structurés à l'aide de graphes conceptuels représentant des connaissances²². Les graphes conceptuels de description étant décrits dans le même langage que les documents, ils peuvent y être directement inclus (dans un arbre annexe). Deux objectifs guident cette description/indexation : permettre de retrouver ou ré-assembler des morceaux de documents sur des critères sémantiques et en utilisant des techniques fondées sur les connaissances ; mais aussi documenter les éléments de connaissances par des liens aux documents.

Les éléments de documents balisés par SGML sont décrits par des graphes conceptuels, qui leur sont liés par une relation *Représentation*. Un « principe d'inclusion » est utilisé qui spécifie que la description d'un élément peut soit être un concept unique (un graphe à un seul nœud), soit un graphe liant des graphes conceptuels décrivant des sous-éléments de l'élément considéré. Il est possible d'utiliser des graphes comme référents de concepts, par exemple le concept *Proposition* peut avoir pour référent un autre graphe *cg* décrivant une phrase : [*Proposition:cg*], ce qui permet d'exprimer des connaissances « emboîtées ». L'utilisation des descriptions de textes mime les phrases de la langue²³, en utilisant un principe de compositionnalité ascendante : par exemple, les descriptions du type *symbole* sont liées à des mots du texte, et comprennent des graphes concepts qui doivent exprimer des référents (telle personne, tel objet), tandis que les description du type *description* expriment des assertions permettant de lier les référents (telle personne utilise tel objet). La construction d'un graphe conceptuel de description est ainsi fortement contrainte par la linéarité du texte et l'organisation arborescente des éléments mis en place à l'aide de SGML.

La base de type de concepts proposée est constituée de l'ensemble des termes de Wordnet [Fellbaum, 1998], et d'un ensemble de 200 relations (thématiques, mathématiques, temporelles, *etc.*).

La navigation est possible à tous les niveaux du système : niveau des graphes conceptuels, des graphes conceptuels vers les parties de document décrites, ainsi que dans les liens documentaires classiques. La recherche de graphes conceptuels ou des documents décrits passe soit par une utilisation d'index (trouver les graphes conceptuel contenant tels concept, par exemple [*Chat*], ou bien en essayant de trouver des descriptions qui soient spécialisations d'un graphe conceptuel requête (suivant la relation de spécialisation du treillis des types de concepts).

Décrire les documents avec des graphes conceptuels, propager des attributs. Le système PRIME, développé au CLIPS-IMAG de Grenoble se base sur une approche de des-

21. Si d est une représentation de document, q une représentation de requête et K l'ensemble des connaissances, alors la mesure de pertinence $r(d \rightarrow q)$ relativement à K est déterminée par transformation minimale appliquée à d pour obtenir d' tel que $d' \rightarrow q$. Bref, c'est la transformation de la requête telle qu'elle implique (dans le modèle logique) le document.

22. Dans un ordre d'idée proche, mais sans spécialement utiliser les possibilités d'inférence des graphes conceptuels, citons les travaux de [Pouillet *et al.*, 1997] repris par [Roussey *et al.*, 1998] dans le cadre des bases de documents multilingues.

23. Et se rapproche des objectifs initiaux de Sowa [Sowa, 1984].

cription de documents à l'aide de graphes conceptuels pour la recherche d'informations. L'indexation d'images est considérée [Mechkour, 1995] ainsi que l'indexation de documents structurés [Fourel *et al.*, 1998, Chiaramella, 1997].

L'objectif est ici de décrire précisément des connaissances documentaires afin de pouvoir mener des recherches de documents précises, en utilisant à la fois navigation et requêtes. Les graphes conceptuels, au contraire de l'approche précédente sont utilisés tout à la fois pour représenter des connaissances structurelles (structure logique + liens de navigation), mais aussi des connaissances de contenu de documents de façon unifiée. Toutes ces connaissances sont alors disponibles aussi bien pour la navigation que pour les requêtes.

Un *hyperindex* contient toutes les connaissances de structure et de connaissances qui peuvent être utilisées pour décrire les documents. Il joue le rôle du thésaurus de tous les concepts utilisés dans les descriptions. Une *hyperbase* contient les hyperdocuments ainsi que les liens implantant leur structure logique et les liens de navigation. Différents niveaux d'abstraction dans la description logique des documents existent qui permettent d'en considérer les parties de façon autonome. Les requêtes au système portent alors sur toutes les connaissances disponibles. [Chiaramella, 1997] remarque alors que la classique notion de document en recherche d'information change pour signifier n'importe quelle unité structurelle (et avec elle la notion de corpus : le corpus est l'ensemble des unités structurelles de tous les documents), tandis que la non moins classique notion d'index contient maintenant tout type de connaissances.

Une description est un graphe conceptuel utilisant par exemple les relations *contains* (structurelle) entre différents concepts à différents niveaux d'abstraction et *is-about* permettant de lier une unité structurelle à un graphe conceptuel de description (par le biais d'un graphe conceptuel emboîté dans un concept). Aux nœuds du graphe de représentation peuvent être ajoutés des attributs, mais il est aussi possible de compléter leurs descriptions par des liens internes entre concepts, ou de liens vers d'autres documents.

Des *unités d'indexation* sont les unités structurelles indexées, c'est à dire ayant un graphe de représentation associé. Le langage de description est contrôlé par les types de concepts des graphes conceptuels. Les unités d'indexation sont limitées volontairement à certains niveaux d'abstraction utiles (par exemple tous les niveaux entre *Chapitre* et *Sous-section*). La stratégie d'indexation est ascendante : on considère pour une unité d'indexation qu'elle est décrite par l'aggrégation des index des unités avec lesquelles elle est liée par la relation *contains*, c'est à dire qu'il y a une compositionnalité explicite des index suivant les relations de structure. L'aggrégation de graphes conceptuels correspond ici au calcul de leur joint maximal, exprimant leur spécialisation maximale commune : le graphe conceptuel résultant décrit tous les graphes conceptuels dont il est aggrégation, ceci faisant appel une fois de plus à la relation *is-a* du treillis des types de concepts.

La notion d'indexation dynamique précédente est ensuite généralisée à la notion de *propagation d'attributs* entre unités d'indexation qui permet de gérer la manière dont sont composées les descriptions. Sont alors définis des attributs statiques, qui restent attachés aux unités d'indexation qu'ils décrivent (par exemple un titre n'est pas propagé aux composants de l'unité dont il est attribut), ainsi que les attributs dynamiques, qui peuvent se propager le long de la structure du document, soit de façon ascendante (par exemple des mots-clé), soit de façon descendante (par exemple une date de publication).

Une requête consiste en un graphe conceptuel d'indexation à partir duquel il va falloir trouver des graphes conceptuels solutions (décrivant des unités d'indexation) qui permettent de le déduire en terme de logique du premier ordre (on est toujours dans le cadre du modèle logique de [van Rijsbergen, 1986]). La recherche se fait de telle sorte qu'on retrouve une unité du plus petit niveau d'abstraction possible telle que son graphe indexation implique le graphe requête.

9.2.6 Analyse des approches

Nous présentons dans cette partie quelques conclusions sur l'approche de description de documents orientée connaissances.

Un schéma général de représentation

Remarquons tout d'abord que l'on se trouve dans tous les cas que nous avons présentés dans le cadre d'un schéma général mettant en jeu un ensemble de connaissances de description (*base de connaissances*), permettant la mise en place de *descriptions* associées à des *parties* de documents.

Base de connaissances. La base de connaissances contient le plus souvent les termes à partir desquels les descriptions pourront être générées. Dans le cas des descriptions par graphes conceptuels, on y trouve ainsi le treillis des types de concepts et le treillis des types de relations, organisés suivant la relation *est-un*, c'est à dire la relation de spécialisation. Les possibilités de relations entre concepts sont spécifiées dans la définition des relations. Dans le cadre de l'approche de [Nanard et Nanard, 1995], les concepts sont définis comme classes dans un langage orienté objets, et peuvent être spécialisés suivant une relation d'héritage (par exemple humain/célébrité/auteur). Les relations définissent leur portée sur des classes et n'ont pas d'instances.

La couverture de la base de connaissances peut varier du tout au tout : elle peut ainsi contenir des connaissances techniques dans un domaine limité [Nanard et Nanard, 1995], associer descripteurs de contenu et descripteurs structurels [Chiaramella, 1997] ou encore l'ensemble des mots du réseau Wordnet²⁴ et un ensemble de relations supposé quasi-exhaustif [Martin, 1996].

Parties de documents. Les parties de documents décrites le sont le plus souvent à l'aide de balises SGML/XML qui en donnent la portée. L'approche de [Nanard et Nanard, 1995] permet de définir un contexte de compréhension minimal plus étendu et significatif que la simple ancre.

Descriptions. Les parties de documents sont intégrées dans la description globale proposée dans [Chiaramella, 1997], c'est à dire qu'elles font partie de la description. Les descripteurs structurels permettent ainsi de structurer et de créer le document, qui n'a pas d'existence en dehors d'eux. Dans l'ensemble des approches utilisant des graphes conceptuels, ceux-ci sont utilisés pour décrire des parties de documents, et permettent éventuellement de lier celles-ci (*cf.* [Martin et Alpay, 1996]).

Si l'approche de description par graphes conceptuels vise à générer un graphe qui annotera par une relation simple une partie de document, la notion de qualificateur de contexte définie dans [Nanard et Nanard, 1995] apporte une connaissance supplémentaire. Un concept général décrit une partie de document, mais son acception, c'est à dire le sens dans lequel il convient de l'interpréter est porté par le qualificateur de la relation au document.

Un ensemble structuré de connaissances, organisé en hiérarchie de spécialisation est donc utilisé pour décrire des parties de documents, voire pour décrire les documents dans leur structure. Ces descriptions structurées peuvent ensuite être utilisées dans le cadre d'inférences.

24. Voir <http://www.cogsci.princeton.edu/~jn/>.

Exploitation de la représentation : l'inférence

Inférence. Nous définissons de façon générale l'inférence comme la déduction de nouvelles connaissances à partir de connaissances connues (valables selon une interprétation particulière), dans le cadre d'une tâche explicite. Il s'agit par exemple en appliquant un certain nombre de règles de répondre à des questions, en mettant explicitement en relation des faits, ou bien en déduisant de nouveaux éléments de connaissances. Toutes ces nouvelles connaissances sont bien entendu connues *en puissance* dans le système, mais c'est leur mise en évidence pour une certaine utilisation qui se révèle intéressante.

Dans un système à base de connaissances considéré de la façon la plus standard (un ensemble de faits, un ensemble de règles permettant de déduire des nouveaux faits), il importe que toutes les inférences soient valides dans le domaine de réalité modélisé par le système. Bachimont a montré que ceci suppose alors une organisation stricte des connaissances en fonction des calculs possibles, de telle sorte qu'elles puissent conduire à des inférences « justes », dont les résultats constitueront de nouvelles connaissances, c'est à dire « des actions possibles (contenu gnoséologique) » dans le cadre d'une tâche explicite [Bachimont, 1996]. Ceci suppose notamment que les connaissances décrites en langue par des termes soient définies de façon non ambiguë et puissent ainsi devenir des primitives de modélisation, afin que l'expression en langue des résultats de calculs soient interprétables de façon juste²⁵.

Inférences documentaires. Les différentes représentations que nous avons vues précédemment proposent plusieurs types d'inférences. Ainsi, on distinguera d'une part les inférences d'*extension de requêtes*, d'autre part les *inférences fondées sur les chemins*²⁶.

A la première catégorie appartiennent les approches de représentation fondées sur les graphes conceptuels. Par exemple [Martin et Alpay, 1996] proposent d'utiliser l'opération de projection pour retrouver des graphes conceptuels qui correspondent à des spécialisations du graphe conceptuel requête. [Genest, 1999], sur un constat d'insuffisance de cette approche, définit des opérations de transformation plus élaborées (par exemple le remplacement de concept ou de relation). S'il utilise l'opération de projection comme inférence finale, [Chiarabella, 1997] étend un peu le calcul général, en utilisant une construction d'index basée sur une utilisation des index des parties contenues structurellement dans une partie de document, qui correspond à de la propagation d'attribut. Dans tous les cas, on se retrouve peu ou prou avec une résolution fondée sur le principe qu'un graphe requête doit pouvoir être inféré des graphes de description de documents ou de parties de documents, cette inférence étant fondée sur la relation de spécialisation de la base de connaissances.

Ce type d'extension de requête correspond à ce qui est par exemple proposé en recherche d'information : à partir des informations données par l'utilisateur sur ses besoins (le plus souvent sous la forme de mots-clé), on essaye d'étendre la requête en remplaçant certains de ses éléments par d'autres en vertu d'une proximité sémantique se ramenant souvent à de la généralisation ou de la spécialisation. Doit alors se poser la question de la pertinence de telles extensions : inférer que je cherche un document sur les animaux parce que j'ai posé une requête sur les chats (et vice-versa) peut très bien se révéler faux de mon point de vue. Ce sont en fait les relations selon

25. Un processus de normalisation sémantique de la base de connaissance terminologique doit ainsi permettre que les sémantiques formelles (le concept comme prédicat) et opérationnelles (les manipulations sur les prédicats) aboutissent à des résultats signifiant quelque-chose lorsqu'ils seront interprétés. Ceci n'est réalisable que dans un domaine limité, avec des primitives parfaitement fixées, permettant d'exprimer des concepts non ambigus sur lesquels on fera des calculs.

26. Inférences contextuelles dirons-nous.

lesquelles se calcule la proximité sémantique qui importent, et leur signification pour l'utilisateur dans le cadre de sa recherche.

Dans la deuxième catégorie d'inférences, on trouvera les expressions de chemins comme règles d'inférence visant à mettre en relation des nœuds d'un graphe, pour la recherche ou la génération de documents [Nanard et Nanard, 1995]. Les expressions de chemins sont stockées dans les nœuds concept du graphe de description, et correspondent à des utilisations particulières du système d'information documentaire composé des documents et de leur graphe de description. Les expressions de chemins peuvent donc être ajoutées au besoin en fonction des usages, et sont *a priori* indépendantes des relations d'héritage liant les concepts qu'elles utilisent.

Nous nous retrouvons donc en face de deux catégories d'inférences. Les premières peuvent être considérées comme globales au système, logiques et bien formées, et s'appliquant sur des éléments quelconques de celui-ci. Dans le deuxième cas, les inférences contextuelles, locales aux éléments qu'elles concernent s'appliquent au besoin et peuvent être modifiées.

Nous concluons notre courte discussion sur les approches basées sur les connaissances et les inférences pour la description documentaire en évoquant le problème de la « réalité » représentée par les descriptions, et son lien avec l'utilisation qui est faite des description.

Réalité et tâches

Les inférences réalisées dans le cadre d'une utilisation de descriptions fondées explicitement sur les connaissances le sont en effet dans le cadre d'une tâche. Les connaissances inférées se révèlent utiles car elles permettent à un utilisateur d'*exploiter* des documents.

On peut alors se poser rapidement le problème une réalité à laquelle se confronte un système de description à base de connaissances. Soit celui-ci fonde le document comme réalité, qui va être objet d'analyse, de description et d'inférences. Soit on s'intéresse à de la réalité hors-document, et on exprime celle-ci en faisant un lien vers une représentation de la réalité qui se trouve dans le flux (c'est par exemple explicitement le cas dans l'approche de [Agius et Angelides, 1999]). Si nous préférons la première approche, dans laquelle toute connaissance est censée provenir des documents, la démarcation n'est bien sûr pas si tranchée. L'utilisation du sens commun, de connaissances encyclopédiques est de toute façon naturelle et le plus souvent inconsciente, et celles-ci sont encasulées, tant dans l'organisation de la base de connaissances que des descriptions.

Nous considérons que toute la réalité est dans les textes, dans les documents, ici audiovisuels, et dans leur descriptions. Ce qu'on capte de la réalité extérieure, du sens commun, est fatalement capté dans la description²⁷ que l'on fait du document.

On représente donc une « réalité », on la fonde en la décrivant, ainsi que les manipulations qu'il est possible d'en faire, et celles-ci doivent n'avoir de pertinence qu'au regard des documents et de leur exploitation.

En fait, la question sous-jacente au (faux) problème de la réalité est la suivante : en quoi la mise en place ou le repérage et la documentation de connaissances dans un document a t'il valeur dans le cadre d'une exploitation ? Quelle que soit la valeur de vérité d'un signe repéré

27. Le plus souvent, la mise en place d'ontologies de domaines résulte d'un traitement automatique de documents techniques associés au domaine, l'organisation des descripteurs dans l'ontologie relevant d'une démarche de rationalisation la plus limitée possible au niveau des ajouts de termes [Biébow et Szulman, 1998].

et instrumenté dans le document, celui-ci n'est signifiant que par l'action réelle qu'il engendre chez son récepteur, action qui dépend forcément du contexte du message (les signes n'ont pas de valeur « ontologique »).

A la description, par l'expression imparfaite du signe en un graphe de descripteurs, on entraîne celui-ci dans le système. A l'utilisation, on essayera d'utiliser la documentation de la tâche, du contexte, bref tout ce qui est dans le système, ainsi que son propre sens commun pour donner une utilité à la description.

C'est pourquoi, de façon générale, une description, et les connaissances la permettant sont mises en place en considérant *a priori* les tâches qui pourront être accomplies en l'utilisant, les inférences qui pourront être réalisées, leur donnant un véritable statut de connaissances (description XML d'un document pourra être utilisée par une feuille de style pour sa présentation ; une description d'un document à l'aide de mots-clé liés à des parties de document sera utilisée pour rechercher celui-ci). Mais, et c'est heureux, elles peuvent également devenir connaissances dans le cadre d'autres tâches *à condition qu'en existe la possibilité*, c'est à dire qu'il soit possible d'utiliser des inférences, des déductions non prévues au départ.

Dans le cadre d'un système de connaissances figées, à inférences globales, il faudra alors changer de système d'inférence, c'est à dire remettre en cause les règles en vertu desquelles le système a été construit, y compris au niveau de la base de connaissances (cela revient par exemple à un changement de modèle d'interprétation logique). Dans le cadre d'inférences plus localisées, c'est à dire n'étant pas consubstanciées au système de description, cette flexibilité par rapport à la tâche et à l'utilisation est naturelle. Il va alors de soi que cette dernière approche a notre faveur, et correspond à ce qu'il est possible de réaliser dans le cadre général des Strates-IA.

9.3 Strates-IA et connaissances

Nous situons dans cette partie les Strates-IA et les mécanismes d'exploitation associés comme approche originale de description de documents fondée sur les connaissances. Nous proposons ensuite quelques intuitions sur l'annotation comme écriture sur un flux audiovisuel, et discutons enfin la gestion et l'évolution des connaissances liées aux Strates-IA.

9.3.1 Strates-IA et approches centrées sur les connaissances

Nous passons d'abord rapidement en revue les trois parties du schéma général de description de système fondé sur les connaissances que nous avons proposé en 9.2.6.

Base de connaissances. Notre base de connaissances contient l'ensemble des éléments d'annotation abstraits qui permettent de décrire un document, organisés en un graphe connexe. La relation de spécialisation permet de décrire une hiérarchie de concepts, mais cette relation n'a pas *a priori* de signification calculatoire, c'est à dire que l'héritage d'attributs par exemple n'est pas une propriété de la relation²⁸. L'organisation de la base de connaissances est plutôt à la base celle d'un thésaurus décrivant l'ensemble des termes utiles et leurs relations, lesquelles lui permettent de dépasser le niveau de la simple liste de termes.

Les possibilités de relation entre éléments d'annotation « instances » de concepts sont définies dans les valences et les schémas de description. Les valences peuvent, toutes proportions gardées, s'apparenter à des rôles tels que ceux que l'on peut trouver dans les logiques de description.

28. Quoiqu'on puisse l'envisager dans le cadre de certaines dimensions d'analyse, voir plus loin.

Ancrage des connaissances dans les documents. Les unités audiovisuelles telles que nous les avons définies ne portent aucune sémantique autre que celle liée à leur existence. Ce sont de pures unités d’ancrage dans le flux, fondées sur sa temporalité. En un sens, ces unités pourraient être qualifiées de « pré-syntaxiques » et correspondraient dans un texte à une suite quelconque de caractères²⁹.

Descriptions. Les descriptions Strates-IA sont des réseaux d’éléments d’annotation instances d’éléments d’annotation abstraits et annotant des unités audiovisuelles. Une unité audiovisuelle est définie à la base par un EA primitif. Le sens des unités audiovisuelles est entièrement et contextuellement porté par les éléments d’annotations auxquels ils sont liés. Les graphes sont mis en place en utilisant des dimensions d’analyse, des schémas de description et des valences, qui fournissent des guides à l’annotation. Les relations entre éléments d’annotation sont d’un unique type, et leur sémantique est éventuellement portée par des éléments d’annotation intermédiaires. La conséquence en est que la liberté d’annotation fournie par le système est totale, et n’accorde pas de primat à une éventuelle décomposition ou description.

Toute description Strates-IA est autorisée par la base de connaissances. Ce sont les schémas de description et les valences qui fournissent la connaissance supplémentaire permettant de contraindre une description. On se trouve ici dans une approche différente de l’approche de description par graphes conceptuels où les règles d’organisation sont portées par la base de connaissances.

Inférence et tâches

Les inférences dans un système Strates-IA passent obligatoirement par l’instanciation d’un graphe potentiel, comme expression de chemin, c’est à dire de relations entre éléments du graphe général. L’instanciation permet de trouver des éléments du graphe qui sont dans les rapports qu’on souhaite leur trouver avec d’autres éléments.

Notre inférence se rapproche de l’inférence fondée sur les chemins de [Nanard et Nanard, 1995], elle est cependant généralisée par rapport à elle, car un graphe potentiel caractérisé est une combinaison de chemins, de rapports contextuels entre éléments, y compris les unités audiovisuelles comme représentants des flux. De plus, un graphe potentiel est appliqué à un nombre de points de départ quelconques, c’est à dire que l’on cherche des chemins à partir de plusieurs éléments du graphe.

29. Si on considère des unités audiovisuelles « étendues » (pour désigner par exemple un objet se déplaçant ou bien une piste d’un flux sonore), il convient d’être plus prudent. Car alors il devient possible de charger l’unité d’ancrage d’une signification, par exemple « ceci est un objet déformable » ou « ceci est la piste numéro 1 destinée aux effets spéciaux ». Ces unités d’inspiration technologique peuvent alors être rapidement intronisées unités syntaxiques d’un éventuel « langage audiovisuel », de la même manière que l’on distingue des mots et des phrases dans les textes. Si cela est tout à fait possible, il n’en reste pas moins qu’il convient d’être conscient des limitations et des *a priori* que l’on introduit par une telle approche. Les mots de textes n’en sont pas les unités ultimes, et les discours se ramènent en dernier lieu à une suite de phonèmes dépourvus de toute signification.

Les documents audiovisuels, comme superpositions extrêmement complexes de flux audio et vidéo peuvent être étudiés suivant des décompositions syntaxiques à caractère technologique (par exemple les objets vidéo MPEG-4), pourvu qu’on n’affirme pas qu’il ne peut en être autrement. Les unités audiovisuelles des Strates-IA se fondent sur l’axe minimal qu’est l’axe temporel. L’extension des unités audiovisuelles de la façon la plus neutre possible consisterait par exemple à considérer une zone image (et non un objet) et à considérer un unique flux sonore. Il y aurait alors des strates audiovisuelles (avec une partie vidéo étant une zone d’image évolutive), des strates vidéo, et des strates audio.

On se rapproche donc, à partir de l'inférence contextuelle basée sur les chemins, de la recherche de motifs, de graphes requêtes dans le graphe général. A la différence des approches basées sur les graphes conceptuels, nulle inférence de spécialisation n'est utilisée dans une opération de projection, l'opération unique étant l'instanciation.

Cependant, il va de soi qu'indiquer dans un graphe potentiel qu'un EA générique doit être inscription dans le flux d'un EAA qui soit en relation de spécialisation avec un autre implique une inférence de spécialisation explicitement posée par le concepteur du graphe potentiel. Ainsi, spécifier qu'un EA générique devra provenir d'un EAA qui soit spécialisation de (*EAA:Homme politique*) par exemple signifie qu'on pose explicitement les inférences qui pourront être faites par le système.

A fortiori, utiliser un graphe requête (généralisation de graphe potentiel caractérisé) découle du même principe. Là où les dimensions d'analyse utilisées correspondent à des contextes — ensembles d'EAA possibles — dans la base de connaissances, exprimés avec autant de filtres de désignation que nécessaires, et pouvant utiliser *a priori* toutes les relations de la base de connaissances sans se limiter à la relation de spécialisation³⁰.

L'utilisation du contexte dans la base de connaissances est donc à la base des opérations de « déduction » qu'il est possible de réaliser pour une recherche. D'une certaine manière, cela permet de parer, si on le désire, au surplus de lâcheté introduit par l'approche résolument « orientée thésaurus » que nous avons adoptée. La perte d'une sémantique formelle précise pour tout le système autorise la mise en place d'inférences ayant une sémantique (formelle si l'on veut) localisée.

Les inférences, définies par les graphes potentiels, peuvent par conséquent être adaptées à la tâche en cours. Les schémas de description, contraignant localement la mise en place des annotations fournissent des graphes potentiels, des inférences possibles et plausibles correspondant à celles prévues lorsque l'annotation a eu lieu.

[Chiaromella, 1997] définit la notion de propagation d'attributs le long d'une arborescence structurelle de documents, et précise qu'il serait utile que cette notion soit étendue à d'autres types de structures. Ceci est en partie réalisé par [Fourel *et al.*, 1998], qui définit pour tous les attributs correspondant à un attribut structurel de document leur portée en termes de relation syntaxiques (structure, hyperlien, enchainement) ainsi que les conditions dans lesquelles la propagation peut se réaliser.

L'approche contextuelle que nous adoptons, et la mise au même niveau de toutes les annotations, quelles qu'elles soient nous semble fournir une autre solution à cette extension, se basant sur n'importe quelle relation. Rechercher le contexte d'un élément du graphe (par exemple une unité audiovisuelle) comme ensemble d'éléments d'annotation à l'aide d'un graphe potentiel permet d'une certaine manière d'en considérer la transmission (comme attribut contextuel) vers l'UAV. L'intérêt est ici qu'il n'y a *pas de différenciation entre les éléments à propager et le médium de transmission*: tout élément d'annotation peut être considéré comme annotation contextuelle (attribut propagé), et servir de support au chemin permettant la propagation.

Entre connaissances strictes et mots-clé

L'approche des Strates-IA et les mécanismes d'exploitation qui sont associés nous semblent se situer entre deux grandes classes de systèmes de représentation.

30. Cela pourrait être rapproché de certaines inférences dans les réseaux sémantiques, telles que les « path-based inferences » proposées par [Shapiro, 1991].

Les premiers sont lâches dans leurs pré-supposés : les documents sont décrits par des mots-clé, éventuellement extraits d'un thésaurus, et les requêtes se limitent à des ensembles de mots-clé, dont on recherche les occurrences combinées dans les descriptions de documents. Les connaissances d'annotation sont donc peu contrôlées, peu explicites, et la mise en place de relations entre mots-clé ne peut découler que de leur juxtaposition, l'annotateur ou l'utilisateur posant une requête inférant celles-ci de façon implicite. Par exemple, poser *jacques AND chirac* sur un moteur de recherche implique que je cherche des pages concernant la tête de l'Etat français, et que les deux termes forment un nom complet, mais cela pourrait à la limite indiquer plutôt que je cherche des renseignements sur les révoltes paysannes dans un certain village de la Creuse.

Sur le constat du manque d'expressivité explicite des systèmes à mots-clé, l'idée de complexifier les descriptions vise à obtenir des descriptions qui dépassent ces systèmes. Dans la seconde classe de systèmes, il s'agit alors de mettre en place des possibilités d'expression de connaissances précises équivalentes à l'expression des connaissances telle qu'elle se retrouve dans les systèmes à base de connaissances. Les réseaux sémantiques deviennent des candidats idéaux, qui doivent également inclure des connaissances d'ordre structurel permettant une exploitation des documents dépassant le niveau de la simple recherche. L'exploitation se fonde alors sur des inférences logiques et supposent une organisation non ambiguë des connaissances de description si l'on veut que le formalisme puisse conduire à des résultats justes au regard de leur interprétation.

Il nous semble que notre approche se situe entre ces deux classes de systèmes. Fondamentalement, nous proposons d'organiser des mots-clé, (des termes disponibles dans un thésaurus) dans les descriptions de documents. Les mots-clé de relation permettent d'exprimer le sens des relations mises en place. Il est donc possible d'exprimer aussi bien des relations implicites entre termes que des relations précises, y compris par exemple des relations structurelles.

Ce sont les schémas de description qui permettent de contraindre les termes d'annotation utilisés et les relations qu'ils doivent entretenir dans les graphes d'annotation. L'organisation des connaissances de description, la base de connaissances sert alors de thésaurus, comprenant *a minima* une relation d'organisation par spécialisation, mais aussi éventuellement d'autres relations.

Les termes utilisés sont expliqués dans les annotations par les relations qu'ils entretiennent, par les contextes dans lesquels ils trouvent. Il n'y a donc pas de sens strict et unique pour un terme, tout dépendra de la contextualisation en cours : il est possible alors de multiplier les points de vue sur un même terme, c'est à dire d'adapter sa signification à la tâche en cours. Les inférences sont contextuelles, adaptées à la tâche en cours, et plus ou moins spécifiées par les utilisateurs. Certains graphes potentiels sont ainsi fournis par les schémas de description, d'autres peuvent être construits au besoin, par annotation d'unités audiovisuelles virtuelles.

L'organisation des connaissances dans la base de connaissances pilote les descriptions par le biais des dimensions d'analyse qui y sont induites par la relation de spécialisation. Cependant, d'autres dimensions d'analyse, construites de diverses façons et en fonction des besoins sont également mises en place. En d'autres termes, l'organisation et la signification des connaissances dans la base de connaissances est contrainte par les relations de la base de connaissances, mais aussi par les dimensions d'analyse et les schémas de description qui les utilisent³¹. La base de connaissances n'est pas une ontologie d'essences, figée pour toute utilisation, mais dépend de son utilisation. L'organisation des connaissances dans la base de connaissances (avec les dimensions d'analyse et les schémas de description) ne découle donc pas uniquement d'un questionnement

31. Ainsi bien entendu que par les mots utilisés pour les termes de description, dont l'interprétation a été suspendue par le passage au terme, mais qui sont, quoiqu'il arrive et quelque soit le contexte, interprétables [Rastier, 1995b].

différentiel sur les termes, mais également et surtout d'un questionnement sur leur utilisation en contexte, laquelle peut remettre en cause toute volonté de figement de sens hors contexte.

Si la base de connaissances est ontologie, elle ne l'est que dans le cadre des tâches qui permettent de la considérer comme telle. Ainsi, rien n'interdit de créer une base de connaissances et des graphes potentiels (inférences) stricts, et de considérer le système comme un système à base de connaissances standards, auquel cas la base de connaissances sera ontologie. Dans un autre registre, il est éventuellement possible que certaines parties de la base de connaissances, certains termes puissent être partagés par tous les utilisateurs dans l'utilisation (la contextualisation) qui en est faite, et conduire à des ontologies « locales ».

Au final l'approche des Strates-IA est moins stricte que celle des systèmes à base de connaissances, mais permet de complexifier des approches fondées sur des description par ensembles de termes, de mots-clé, dont le fonctionnement se révèle pour l'instant le seul à peu près satisfaisant en recherche d'informations.

Il nous semble qu'il s'agit alors d'essayer de bénéficier de ces résultats pour permettre d'écrire sur les documents audiovisuels avec des termes que l'on contextualise, cette écriture pouvant atteindre à la rigueur d'une représentation logique, mais également s'en écarter totalement ; il s'ensuit que la « vérité » des inférences n'est pour nous qu'une notion relative, utile dans certains cas, et non pertinente dans d'autres. La liberté qu'on laisse par rapport à une annotation stricte peut alors se justifier comme des incursions du fait linguistique dans le domaine du calcul, de l'inférence et des connaissances figées. Par rapport aux approches privilégiant les mots-clé, nous essayons de rester dans le fait de langue, parce que l'interrogation est fait de langue, et que la mise en place de l'annotation est également fait de langue.

Le formalisme n'est alors que manipulation, et correspond aux mécanismes d'instanciation. Pour permettre de plus d'avoir de l'effectivité — du fait formel qui prend sens comme fait de langue — il convient de contraindre la manière de faire des inférences (par exemple à l'aide de graphes potentiels déjà formés), ainsi que le vocabulaire utilisé par les inférence, ce qui correspond à des schémas de description stricts.

Dans la partie suivante, nous développons cette intuition de l'annotation comme écriture, sans prétendre à une étude exhaustive et définitive.

9.3.2 Écriture sur le flux

Nous prenons ici le parti de considérer l'annotation comme écriture sur le flux, à partir d'un vocabulaire donné (les termes de la base de connaissances) et d'un règle de syntaxe limitée correspondant à la possibilité de mise en relation élémentaire, ainsi que des contraintes portées par les flux audiovisuels et leur appréhension.

Une écriture libre ou contrôlée

Le fait de mettre en place une unité audiovisuelle en suivant une dimension d'analyse découle du repérage d'un signe dans le flux, correspondant aux objectifs de description de la dimension d'analyse. Dans le cas de l'annotation libre, la dimension d'analyse n'est pas contrainte et décrit l'ensemble de la base de connaissances.

Après la mise en place d'un EA primitif et d'une unité audiovisuelle, soit on passe à l'annotation suivante (continuer à suivre la dimension d'analyse, ou changer d'idée de description), soit on peut compléter l'annotation de l'unité audiovisuelle, c'est à dire l'annoter par d'autres

éléments d'annotations qui auront alors même portée temporelle que l'EA primitif. La mise en relation d'éléments d'annotation est également possible, qui permet de contextualiser les termes utilisés³² (*cf.* le diagramme de décomposition de tâches de la figure 8.1). L'annotation peut alors se concevoir comme une suite de créations d'EA et de mises en relation, guidées par le matériau audiovisuel, mais aussi par les associations d'idées de l'utilisateur. Remarquons également qu'écrire une relation explicitée de la même manière que l'on annoter permet d'homogénéiser la manière de décrire, mais aussi de prendre en compte le fait que la verbalisation d'une relation sémantique quelconque est fait de langue, et n'entraîne pas obligatoirement l'utilisation d'une primitive hors-langue comme c'est le cas dans les graphes conceptuels.

Si l'écriture libre permet de décrire de façon quelconque un flux audiovisuel, à un degré de précision quelconque, et suivant un cheminement d'idées quelconque, le texte final risque de n'être que peu utilisable (en terme de requêtes), si ce n'est par l'auteur de l'annotation (il est cependant toujours possible d'explorer un graphe d'annotation). Les dimensions d'analyse permettent ainsi de contrôler l'écriture sur le flux, en proposant à l'utilisateur des regroupements de termes jouant le même rôle au regard de la description. Associées, mises en relation dans des schémas de description, les dimensions d'analyse prennent un rôle plus précis puisqu'elles spécifient alors des classes paradigmatiques au regard d'un arrangement syntagmatique (c'est à dire que les termes décrits jouent le même rôle relationnel par rapport à d'autres classes de termes³³).

Les schémas de description peuvent alors être considérés comme des expressions de règles de grammaire du langage des Strates-IA, règles plus ou moins strictes, qui peuvent être combinées, étendues, bref manipulées. De la même manière, un graphe requête propose de retrouver des instances d'arrangements syntagmatiques de classes paradigmatiques de termes, et les valences fournissent pour chaque terme les possibilités de relation, c'est à dire de participation à des phrases de termes.

Au final, il apparaît bien que l'annotation peut être considérée comme écriture sur le flux par un utilisateur, aidée par le système, d'un « texte ». Ce « texte » de termes doit alors être lu par un utilisateur qui en retrouvera ou non le sens initial.

Lectures

Remarquons tout d'abord que l'interprétation d'un unique terme d'annotation découle d'une description et d'une contextualisation au sens où le lecteur interprète le terme en projetant ses propres visées de contextualisation³⁴. Le degré d'ambiguïté des termes peut varier, par exemple, un EA *⟨LienHypermedia⟩* est non ambigu, tandis qu'un EA *⟨Chirac⟩* l'est beaucoup plus.

Dans le cas des « textes », qui sont des ensembles de *phrases multi-axiales*³⁵ de termes, la lecture et l'interprétation découlent également de descriptions et mises en contextes. Lire un texte revient alors à être auteur de sa propre lecture, c'est à dire à suivre des cheminements de description, qui ne correspondront pas forcément à celui qui a été utilisé à l'écriture, et qui seront médiés par les contextes mobilisés et l'affichage proposé (sous la forme de phrases linéaires de termes, de graphes, d'arbres, de lignes de temps, *etc.*).

Les relations entre termes, suivant qu'elles sont verbalisées ou non, font plus ou moins appel aux facultés du lecteur pour interpréter les ellipses qui lui sont proposées. Par exemple *Senna*,

32. Il est alors intéressant de constater que d'une certaine manière, on « termine la phrase » qui a été commencée par la mise en place du premier élément d'annotation.

33. Ces classes peuvent être déclarées d'entrée de jeu dans la base de connaissances en utilisant les relations de spécialisation.

34. Cependant, retrouver le contexte dans lequel un EA a été utilisé est l'assurance d'une lecture juste.

35. Au sens où à la fois l'axe temporel et les axes de relations des graphes (un graphe n'est pas *a priori* un chemin) sont mobilisés.

Mur, Italie (qui annotent une même unité audiovisuelle) est une phrase de termes interprétable, de la même manière que *Chirac Agent Toucher Objet Vaches*; ou encore *Veaux Vaches Cochons* répartis sur une ligne de temps.

Le fait de considérer l'annotation comme une écriture dans un langage semi-contrôlé peut se révéler intéressant notamment en ce qui concerne la *dynamique* de l'annotation. En effet l'annotateur dispose d'un vocabulaire, est aidé par des règles de grammaire, ainsi que des phrases toutes faites qu'il peut adapter et utiliser. Engagé à la fois dans l'annotation comme écriture et dans l'appréhension du flux, l'annotateur doit être capable d'annoter à la manière d'une prise de notes sur une fiche de lecture³⁶.

De la même manière, la description d'un graphe requête passe par une écriture dans le même langage, guidée par les mêmes règles.

La question de la lecture de l'annotation, associant l'utilisation des règles d'écriture comme filtres et la présentation des termes d'annotation et des UAV liées reste bien évidemment complexe. Cette complexité est cependant le lot commun de tout système proposant une annotation (semi-)structurée pour la recherche et l'exploitation d'information : un graphe conceptuel par exemple, dès qu'il dépasse quelques nœuds, devient illisible, et doit être linéarisé, voire transformé en phrases pour pouvoir être lu et compris.

9.3.3 Gestion et évolution des connaissances Strates-IA

Nous donnons dans cette dernière partie du mémoire quelques pistes sur les possibilités de gestion et d'évolution de connaissances permettant l'annotation liées aux Strates-IA telles que la base de connaissances, les valences, les dimensions d'analyses et les schémas de description.

Manipuler les connaissances de description

Vers une conception dynamique des connaissances. Nous avons déjà plusieurs fois évoqué le fait que certaines dimensions d'analyse étaient « naturelles » à la base de connaissances, car explicitement basées sur la relation de spécialisation (section 6.3.4, page 166). La base de connaissances est donc « naturellement » organisée en classes, éventuellement paradigmatiques si elles sont utilisées comme telles dans des schémas de description.

D'autres dimensions d'analyse, mises en place au cours de l'exploitation d'un graphe Strates-IA, peuvent prendre un caractère général si elles sont réutilisées et attestées³⁷. En d'autres termes, certaines connaissances d'utilisation, certains schémas de description existent à l'initialisation du système, tandis que d'autres sont mis en place au cours de l'utilisation. Si les éléments d'annotation abstraits de la base de connaissances sont relativement figés, la manière de les utiliser, les dimensions d'analyse et les schémas de description varient³⁸. La connaissance de description disponible varie alors pour chaque nouvelle session d'annotation, puisqu'elle comprend la base de connaissances et les schémas de description utilisés.

36. Tout l'enjeu des interfaces d'annotation est alors de d'aider l'utilisateur à fournir ce travail, c'est à dire à gérer en même temps la détemporalisation de l'annotation et l'engagement dans le flux temporel qu'il visionne. L'interface d'annotation que nous proposons (*cf.* figure 7.2) est un exemple de réponse à cet enjeu.

37. Par exemple en étudiant des cas d'utilisation rationalisés par les mêmes modèles de tâches.

38. Certains valences peuvent également être proposées « dynamiquement » pour des EAA en fonction des schémas de description auxquels ils participent par l'intermédiaire de dimensions d'analyse.

Gérer les connaissances à partir des dimensions d'analyse. Les dimensions d'analyse désignant en dernier lieu, une fois qu'elles ont été totalement résolues, un ensemble d'éléments d'annotation abstraits, il est possible de les organiser dans une hiérarchie de subsomption ensembliste stricte et bien définie. Un ordre partiel peut ainsi être mis en place basé simplement sur une inclusion ensembliste des termes désignés.

Alors il devient possible d'utiliser des outils de manipulation et d'exploitation de connaissances tels que les logiques de description, qui permettent de mettre en place et de gérer des graphes de connaissances, avec pour sommets des « instances » de ces dimensions d'analyse (qui sont alors des types de concepts).

De tels graphes se trouvent être des schémas de description. On peut alors considérer qu'il est possible de manipuler des schémas de description comme assertions de logiques terminologiques. Ceci signifie que si l'on ne se donne pas le droit de figer la connaissance dans la base de connaissances, on se donne celui de manipuler les archétypes de contraintes de description reconnus dans un cadre totalement formel.

Les manipulations et la classification des schémas de description dans un tel cadre pourraient alors se révéler d'intérêt, par exemple pour en détecter les régularités.

Apprentissages

Schémas de description. Ainsi si on imagine qu'un système est mis en place avec un certain nombre d'EAA³⁹ et de dimensions d'analyse « naturelles » et peu ou aucun schémas de description de départ, l'apprentissage des régularités d'annotation au travers des schémas de description mis en place permettra d'apprendre des règles de description. Les schémas de description les plus spécifiques (qui correspondent en fait directement à des sous-graphes d'annotation) peuvent alors être généralisés en schémas de description plus généraux, simplement en remplaçant les dimensions d'analyse qui les composent par des dimensions d'analyse subsumantes.

L'objectif est donc bien de réussir à décrire la manière dont décrivent les annotateurs,

- pour fournir à d'autres annotateurs des règles de description, afin d'homogénéiser les descriptions de la base,
- mais aussi afin de proposer aux utilisateurs exploitant une base Strates-IA des moyens de décrire de la même manière que celle qu'ont utilisée les annotateurs.

Inversement, si apparaissent des régularités dans la manière dont les utilisateurs décrivent les documents qu'ils cherchent, ces régularités pourraient servir de guides d'annotation pour les annotateurs.

Remarquons que toutes ces manipulations sont rendues possibles du fait de l'homogénéité des concepts d'exploitation des Strates-IA, par exemple entre graphes potentiels, graphes requêtes, schémas de description et graphes d'annotation.

Relations dans la base de connaissances. La base de connaissances, une fois qu'elle a été mise en place, ne peut qu'être complétée, c'est à dire qu'on peut lui ajouter des éléments d'annotation abstraits et des relations.

Les dimensions d'analyses utilisées permettent de mettre en place une organisation virtuelle, des vues de la base de connaissances, c'est à dire une sorte de réorganisation. Cette réorganisation virtuelle pourrait devenir réelle et aller jusqu'à la création de nouveaux éléments d'annotation

³⁹. Le thésaurus pourrait par exemple être mis en place par un traitement automatique sur des textes de description, ou bien en utilisant un dictionnaire tel que Wordnet.

abstraites (correspondant aux noms des dimensions d'analyse par exemple) mis en relation avec les EAA qu'elles désignent.

Il est également possible de compléter la base de connaissances par apprentissage. Par exemple, l'utilisation quasi-systématique d'un schéma de description peut conduire à inscrire dans les EAA qui en composent les dimensions d'analyse des valences qui en sont extraites. Ces valences pourront alors être par la suite systématiquement proposées à l'utilisateur qui emploiera les EAA, même en dehors du schéma de description.

D'une autre manière, par apprentissage des relations élémentaires mises en place au cours des diverses utilisations d'un EAA, il est possible d'apprendre des possibilités de relation qui seront ensuite proposées aux utilisateurs.

Enfin des valences systématiquement utilisées devraient permettre d'une part d'en inférer des extensions de schémas de description, d'autre part d'apprendre des relations dans la base de connaissances. Par exemple, cela pourrait être un moyen d'aggréger dans la base de connaissances des connaissances de sens commun ou encyclopédiques, *de par leur utilisation réelle en situation de description*. Si *Winnie Mandela* est souvent mis en relation avec *Nelson Mandela* alors il peut se révéler utile de créer une relation du type *Epouse* entre les deux EAA correspondants dans la base de connaissances.

Unités syntaxiques et sémantiques

Terminons en évoquant le statut des unités audiovisuelles, auxquelles nous refusons *a priori* tout statut syntaxique dans un éventuel langage de l'audiovisuel. Suivant les éléments d'annotation primitifs qui permettent de les mettre en place, qui président à leur détection, il pourrait convenir, en cas d'utilisation régulière de distinguer un « continuum d'acceptation ».

On pourrait par exemple placer l'ensemble des UAV dans un continuum partant des unités les plus fortes, les plus acceptées et les plus repérables (les documents, les plans) aux unités les moins consensuelles (UAV annotées sur une dimension esthétique par exemple), en passant par des régularités « de fait » sur lesquelles peuvent s'accorder un grand nombre de gens (voir par exemple des scènes récurrentes telle que des « transitions de JT » liées à des types d'émission particuliers, ou des unités correspondant systématiquement à des apparitions de personnalités).

Cela reviendrait à transmettre les propriétés dues à des consensus d'utilisation des éléments d'annotation aux unités audiovisuelles qu'ils définissent, et d'envisager alors un figement syntaxique des unités audiovisuelles au sein d'un langage de description plus spécialisé que les Strates-IA. Alors il deviendrait possible de faire émerger, pour l'audiovisuel, et pour certains genres, des descriptions acceptées et correspondant à une pratique avérée, ce qu'on pourrait assimiler aux descriptions de genres telles que celles mises en place au sein de la TEI.

9.4 Conclusion

Nous avons, au cours de ce dernier chapitre discuté les rapports entre documents et connaissances, en étudiant tout d'abord ceux-ci au travers de la structuration des documents. Nous avons ensuite présenté quelques tentatives d'enrégimenter les descriptions de documents dans des formats documentaires, et avons constaté qu'il convenait en fait de mettre en place des approches de description qui en soient indépendantes. Nous avons ensuite présenté quelques caractéristiques de la norme en cours de développement MPEG-7, et avons discuté les relations possibles avec les Strates-IA.

Nous nous sommes ensuite intéressés aux approches de description documentaires liées explicitement aux systèmes à base de connaissances, et avons dans ce cadre étudié les mécanismes

d'inférence proposés. Nous avons ensuite comparé les Strates-IA et les mécanismes d'exploitation associés à ces approches, ainsi que la notion d'inférence contextuelle, avant de présenter une vision à notre sens féconde des Strates-IA comme écriture sur le flux, dans un langage dont il s'agit d'utiliser et de découvrir les règles. Nous avons enfin évoqué quelques intuitions sur la manipulation de schémas de description dans un cadre formel en permettant la classification, et quelques pistes d'apprentissage de nouvelles connaissances de description.

Conclusion générale

La dernière partie de ce mémoire est pour nous l'occasion de revenir sur les principales étapes de la recherche que nous avons présentée, d'en discuter quelques points et perspectives importantes, et de décrire les développements en cours.

1 Bilan de la recherche

Nécessité d'une indexation intelligente. Nous constatons dans un premier temps que les documents « simplement numérisés » évoluent vers des documents multimédias, et que les systèmes d'information documentaire prennent en compte cette dimension, intégrant recherche par requêtes et exploration, gestion des connaissances documentaires et aide à l'utilisateur fondée sur ces connaissances. L'enjeu est donc de définir une indexation intelligente, qui permette de manipuler les descripteurs à la fois comme index et connaissances.

Instrumenter des documents audiovisuels. Les documents audiovisuels n'échappent pas à la règle, et leur exploitation dépassant la simple visualisation doit passer par leur indexation, sous une forme structurée qui en permettra un accès direct. La seule structure disponible actuellement pour les documents audiovisuels étant une structure de présentation associant simplement flux vidéo (suites d'images) et flux audio, on se trouve (à la différence des documents textuels) dans une situation dans laquelle toute structuration doit être mise en place à partir de rien.

L'annotation comme support d'analyse de la représentation audiovisuelle. Nous définissons l'annotation, c'est à dire l'attachement d'une description à un morceau de flux comme le support d'analyse fondamental de la représentation audiovisuelle, et nous constatons, en étudiant les diverses propositions de modélisation de la littérature que : (1) différentes caractéristiques d'annotation existent et dépendent de la manière dont elle sont mises en place (automatiquement, manuellement) et interprétées par un utilisateur ; (2) différents types de structuration des annotations sont possibles. Nous constatons également que les différentes fonctionnalités d'un système d'information audiovisuelle dépendent totalement de la modélisation choisie pour les documents.

Nécessité d'une écriture contextuelle sur le flux. Nous discutons les nécessités de modélisation induites par les exploitations possibles, mais aussi la prise en compte des contextes internes aux documents et ceux liés à la tâche de l'utilisateur, et argumentons sur le fait que les contextes documentaires dépendent de la tâche de l'utilisateur, et doivent être pris en compte dans une modélisation. Nous justifions le fait qu'une modélisation doit pour l'instant être basée sur des caractéristiques d'annotation interprétables telles que les textes ou les termes, comme unités sémiotiquement pertinentes pour l'homme et manipulables par la machine, l'utilisation

des primitives venant en complément. La modélisation doit également être suffisamment générale pour s'adapter à toute activité de description de document audiovisuel, quelle que soit la tâche que cette activité supporte. En bref, il s'agit d'écrire sur le flux pour contextualiser les annotations, et de lire en (re-)contextualisant celles-ci.

Un modèle très riche adapté à l'annotation. Nous proposons en conséquence le modèle des Strates Interconnectées par les Annotations pour la représentation de documents audiovisuels, qui se caractérise par plusieurs propriétés fondamentales : (1) toutes les caractéristiques d'annotation sont exprimées dans des objets ayant un nom qui est un terme d'annotation (éléments d'annotation) ; (2) les parties de flux annotées (unités audiovisuelles) sont dépourvues de toute sémantique dans le modèle, et représentent un pur découpage du flux audiovisuel ; (3) l'annotation est structurée à l'aide d'un unique type de relation, à la manière de l'annotation (mettre en relation est également annoter) ; (4) les connaissances de description (éléments d'annotation abstraits) sont organisées dans une base de connaissances qui est au minimum un thésaurus. Au final, l'ensemble des connaissances d'un système Strates-IA est contenu dans un graphe orienté étiqueté par des objets Strates-IA.

Un contexte précisément défini. Nous définissons opératoirement le contexte d'un élément du graphe comme les extrémités de chemins partant de cet élément. Nous justifions cette notion par rapport aux différents cas de figures des Strates-IA, en considérant que chaque élément est une connaissance locale qui doit être éclairée contextuellement par un entour privilégié au cours d'une tâche (par exemple l'annotation contextuelle d'une UAV par un EA), et proposons la notion de graphe potentiel pour permettre la gestion de contextes. Les graphes potentiels s'instancient dans le graphe Strates-IA, et nous considérons que toute recherche d'instance est une recherche d'isomorphismes de sous-graphes partiels à partir de correspondances connues, ce qui nous permet de proposer un algorithme efficace de multi-propagation ayant le triple avantage d'être pilotable par une heuristique simple, de rendre des solutions dès qu'elles sont trouvées, et d'avoir de meilleures performances que les algorithmes auxquels nous l'avons comparé.

Divers degrés de contrôle et d'exploitation du modèle. Sur la base des graphes potentiels, nous définissons plusieurs outils d'exploitation contextuelle des Strates-IA. Les graphes potentiels caractérisés permettent de définir les sommets utiles de graphes potentiels et de les manipuler. Les relations contextuelles potentielles sont particulièrement adaptées à l'expression de chemins et les filtres de désignation permettent de désigner des éléments de la base de connaissances. Les dimensions d'analyse permettent d'exprimer une visée de description de l'annotateur en regroupant des EAA ayant un rapport à la tâche en cours. Plus précis, les schémas de description permettent de contraindre l'annotation, en exprimant des relations syntagmatiques entre ensembles paradigmatiques de termes d'annotation. Nous présentons comment ces différents concepts et outils peuvent être utilisés dans les diverses tâches d'exploitation d'un système Strates-IA, et remarquons notamment qu'il est possible d'intégrer requêtes et navigations dans un schéma unique. Le principe qui régit l'ensemble des utilisations repose sur le fait que l'expression d'un graphe potentiel est l'expression d'une visée de contextualisation (de description) liée à la tâche.

Modéliser l'utilisation pour stocker de l'expérience utile. Nous nous intéressons à la mise en place d'un modèle de description de systèmes d'information permettant de stocker l'expérience d'utilisation sous la forme de cas d'utilisation expliqués par les modèles d'utilisation et

de tâches. Nous montrons que ce genre de considération s'adapte aux Strates-IA comme système permettant de représenter des connaissances dans un format semi-structuré, adaptés à un grand nombre de tâches. Nous proposons quelques pistes d'exploitation des cas pour l'apprentissage et l'aide à l'utilisateur fondée sur l'expérience.

Structurer et exploiter des documents avec des connaissances. Nous terminons enfin en opérant un retour sur l'utilisation de représentation de documents à la fois comme index et connaissances dans le cadre de tâches diverses. Nous proposons de considérer toute description documentaire comme structure de connaissances, à partir de laquelle il est possible de construire des structures de présentation pour des tâches de présentation, mais aussi d'inférer de nouvelles connaissances si besoin. Nous confrontons les Strates-IA à différents modèles de description de documents : les premiers sont basés sur des langages de balises, le suivant est la norme actuellement en discussion de description de documents audiovisuels MPEG-7, les autres modèles ont une approche orientée connaissances et inférence documentaire.

Nous situons alors les Strates-IA dans le cadre des connaissances pour l'inférence, et discutons l'inférence contextuelle comme inférence unique liée au formalisme du modèle. La réalisation des tâches est pilotée par l'expression des graphes potentiels, et l'indexation se révèle bien « intelligente », au sens où il est nativement possible de l'utiliser à la fois comme index et comme connaissance mobilisée dans une tâche.

Ecrire et lire sur un flux suivant des règles de langage adaptées à la tâche. Nous discutons à nouveau les notions d'écriture et de lecture contextuelles sur le flux, avant de proposer quelques pistes d'apprentissage et d'évolution des connaissances de description Strates-IA. Nous présentons notamment la notion de connaissances « dynamiques », c'est à dire liées à la reconfiguration de la base de connaissances en fonction de l'usage qui en est fait, exprimé au travers des dimensions d'analyse utilisées.

Notre approche se situe fondamentalement entre une approche libre utilisant des mots-clés et une approche strictement basée sur les connaissances, dans laquelle les concepts sont non ambigus et les inférences strictement et globalement définies.

2 Discussion

Nous nous intéressons successivement à quelques sujets d'intérêt qui nous semblent importants au regard de la recherche que nous avons menée. Nous discutons ainsi rapidement la transmission de connaissances entre utilisateurs des Strates-IA ; l'évolution des utilisations des documents audiovisuels documentés ; la probable émergence de langages audiovisuels et ses conséquences possibles ; et enfin quelques points dont il nous semble que nous sommes redevables à la théorie linguistique de la Sémantique Interprétative.

Transmission de connaissances. Un système minimal de Strates-IA comprend un ensemble de connaissances d'annotation, qui est un ensemble d'éléments d'annotation abstraits (un vocabulaire d'annotation), à partir desquels il est possible de mettre en place des graphes de descriptions. Ces connaissances minimales sont complétées par des connaissances « supplémentaires », qui sont les dimensions d'analyses, les schémas de description, les modèles de tâches, *etc.*

L'utilisation minimale des Strates-IA consiste en la présentation des différentes unités audiovisuelles annotées, en l'exploration des graphes d'annotation, en des recherches faisant appel à des graphes potentiels et à des contextes simples et génériques. Par exemple une recherche

peut s'apparenter à une recherche par mots-clés, ou bien il est possible de définir une annotation contextuelle par une longueur de chemin. Les connaissances partagées sont donc la base de connaissances¹ d'une part et d'autre part la langue (par exemple le français). La mise à disposition des connaissances supplémentaires d'annotation peut fournir des indications supplémentaires sur la manière de poser des requêtes plus complexes, sur des contextualisations conduisant à des inférences contextuelles plus complexes ou plus strictes.

Les Strates-IA permettent donc de gérer de plusieurs façons des mêmes annotations, suivant la manière dont les connaissances supplémentaires — si elles existent — sont partagées entre utilisateurs. Cette caractéristique du modèle nous semble importante, car elle permet toujours une utilisation minimale d'un document annoté, sur lequel toutes les visées de contextualisation peuvent toujours être appliquées, mais ne correspondront pas obligatoirement aux contextualisations utilisées à l'annotation. L'important problème sous-jacent que nous évoquons ici est celui de « l'inter-opérabilité sémantique » entre utilisateurs d'un même système d'information, laquelle se révèle impérative pour que des connaissances de toutes formes puissent être partagées et réutilisées. Dépasser le niveau du simple échange de documents ou de tuples (alors supposés auto-suffisants) nécessite un partage de connaissances à des niveaux variés et pose des questions non triviales. En ce qui concerne les documents audiovisuels, les Strates-IA, les mécanismes de contextualisation associés et les différentes connaissances supplémentaires échangeables permettent de gérer à divers degrés le partage de connaissances.

Intégration des documents audiovisuels dans les systèmes d'information multimédias. Le sujet d'intérêt suivant concerne l'intégration et la documentation (l'enrichissement) des documents audiovisuels dans les systèmes d'information, qui leur permet de devenir eux-mêmes pleinement multimédia au sens que nous avons défini dans notre chapitre introductif. Ainsi, l'accès direct au flux et sa documentation en permettront une manipulation qui changera notre manière d'appréhender les documents audiovisuels. Par exemple, les hyperliens, la possibilité d'arrêt sur image (et surtout de retour sur image), ou de présentation simultanée de plusieurs images permettront de délinéariser et de détemporaliser le flux. La possibilité de navigation et d'interaction avec celui-ci changeront les modes de lecture et d'engagement des spectateurs (ou utilisateurs) de flux.

D'autre part, les documents audiovisuels documentés peuvent être réellement intégrés à des ensembles de documents multimédias tels que le Web (ou d'une certaine manière les cédéroms), non plus en tant que documents/fichiers indépendants inclus en tant que tels dans des documents plus vastes, mais comme parties de documents à part entière permettant d'accéder à d'autres parties. Passer d'un reportage télévisé ou d'une émission à ses rushes, aux articles de presse qui l'accompagnent, aux dépêches de presse qui y sont liés et inversement devient possible par la documentation et la description intégrées de documents audiovisuels. L'INA propose par exemple une démonstration illustrant la documentation d'une émission littéraire : « Roger Caillois interactif »², et participe au projet EuroDelphes³ de mise en place d'un environnement pédagogique multimédia pour l'enseignement de l'histoire.

Il semble donc que la documentation, l'instrumentation des documents audiovisuels permette d'utiliser ceux-ci de manière différente de ce pour quoi ils avaient été prévus, et de les intégrer dans des systèmes multimédias. Alors change également l'appréhension des documents. Par exemple,

1. Quoiqu'à la limite il soit possible d'imaginer un document et ses annotations sans base de connaissances : une base de connaissance minimale, plate, constituée des EAA utilisés dans l'annotation peut être mise en place et permettre l'exploitation du document.

2. Voir <http://www.ina.fr/Production/Studio/index.fr.html>.

3. Voir <http://www.ina.fr/Recherche/Delphes/index.fr.html>.

Groensteen [1997] remarque que l'esthétique et le plaisir de la bande dessinée proviennent du fait qu'il est possible de revenir sur une case du récit, ce qui n'est pas le cas du roman (où la phrase est noyée), ni du film (où le plan disparaît). L'instrumentation d'un document audiovisuel, permettant d'en gérer une détemporalisation, nous semble réduire cette limitation, et conduire à une appréhension nouvelle des documents audiovisuels.

Les Strates-IA permettent d'opérationnaliser des descriptions documentaires riches autorisant des tâches de lecture et d'exploitation variées telles que celles que nous avons présentées.

Emergence de langages audiovisuels et conséquences. Nous avons déjà souligné que le fait que les structures de présentation actuelles des documents audiovisuels soient minimales et brutes (simplement images + sons) se révèle extrêmement intéressant pour la Recherche sur l'étude de la modélisation de ceux-ci. Ainsi, les documents audiovisuels peuvent être étudiés suivant une grande richesse de points de vue non contraints par les structures informatiques sous-jacentes, et de nombreux modèles en sont proposés, qui dépassent largement la simple réutilisation de modèles textuels⁴.

Il apparaît en conséquence qu'il est utile de proposer des modèles de description se basant sur le concept minimal de l'annotation comme attachement de description à une partie de flux, afin d'être à même de tester différents modèles de façon aisée. Les Strates-IA nous semblent un bon candidat, car elles fournissent un cadre riche et des possibilités de contraintes étendues permettant de mener ce type d'expérimentation, et de faire émerger — à partir des descriptions — des modèles de description adaptés à des tâches diverses. L'analyse assistée d'un flux peut permettre de mettre en place de véritables langages audiovisuels suivant les différents types de documents et les pratiques visées par la description.

Le niveau syntaxique minimal (les unités significantes) choisi jouera alors un rôle : choisir de décrire des unités audiovisuelles comme simples strates temporelles revient à se limiter à une approche « classique » de l'analyse (type Metz) correspondant à celle qui pouvait être menée avec un magnétoscope. De façon plus poussée, s'intéresser à des strates « étendues » (video, video+audio, audio), y compris par la désignation de formes dans les images, entraînera d'autres types d'analyses et de langages, plus novateurs, peut être plus liés au « sens commun » de l'audiovisuel (mouvements de caméra liés aux mouvement d'objets, aux types d'actions, *etc.*).

Ces derniers langages audiovisuels, dont le niveau syntaxique descend jusqu'aux parties d'image pourraient, de descriptifs, devenir prescriptifs, c'est à dire être utilisés pour construire des documents autrement que comme regroupements de strates⁵.

Une telle approche est d'ores et déjà utilisée pour la construction de documents audiovisuels. Ainsi, le format MPEG-4 permet de définir des documents comme composés d'objets vidéo temporels, c'est à dire de spécifier et de construire au besoin la représentation des documents à l'écran. Si les résultats sont pour l'instant peu concluants pour les scènes naturelles, un premier pas est cependant fait vers une description à un niveau de décomposition inférieur à celui de l'image des documents. Les jeux vidéos sont un autre exemple de construction de documents audiovisuels au besoin à partir de descriptions variées.

Les langages audiovisuels de description pourraient donc permettre de définir des prescriptions documentaires, avec l'avantage d'avoir été conçus par l'analyse de documents réels dans le cadre de pratiques d'utilisation avérées ayant permis l'émergence d'unités syntaxiques significantes utiles, et non construites *ad-hoc* (par exemple, le présentateur du journal télévisé comme

4. Même s'il est certain que des modèles textuels simples se révèlent utiles d'entrée de jeu, par exemple la possibilité d'accéder à une table des matières des chapitres dans les DVD.

5. Nous nous sommes au long de ce mémoire limité à ce dernier degré.

unité visuelle signifiante est utile si on construit un journal télévisé, et il n'a pas de jambes, le bureau n'est pas forcément utile et peut faire partie du fond de la scène).

Strates-IA et Sémantique Interprétative Après ces prévisions d'utilisation sans doute ambitieuses, mais qui correspondent sans conteste à l'importance que va prendre la génération automatique de documents audiovisuels dans les prochaines années ou décennies, nous disons un mot sur les rapport entre notre approche et la théorie sémantique linguistique de la Sémantique Interprétative (SI).

La SI a été mise au point par Rastier [1987] et nous avons étudié au cours de notre DEA ses rapports avec l'informatique [Prié, 1995]. La SI pose qu'il n'existe pas de sens définitif pour les mots et que toute analyse de texte doit faire appel à des descriptions sémantiques de ceux-ci à l'aide de *sèmes* (ou traits sémantiques) permettant de les comparer (sèmes génériques communs et sèmes spécifiques de différenciation). De plus tous les sèmes de mots peuvent toujours être virtualisés ou transmis dans le contexte d'autres mots, par exemple l'énumération *pomme, poire, scoubidou* peut permettre de transmettre le sème /fruit/ à *scoubidou*. Les *isotopies* sont constituées par des récurrences de sèmes équivalents (mettre *pomme* et *poire* ensemble permet de définir une isotopie liée à /fruit/).

Il nous semble *a posteriori* que la Sémantique Interprétative nous a inspiré à certains niveaux de notre approche, et surtout que les correspondances qu'il est possible de pointer entre SI et Strates-IA pourraient peut-être servir de source future d'inspiration. Nous en donnons ci-après quelques exemples.

Tout d'abord, comme le présente Cavazza [1998], le modèle interprétatif de la SI peut être décrit comme une sorte d'annotation sémantique. L'interprétation d'un texte consiste alors dans l'activation de primitives sémantiques (sèmes) et de leur association avec des entrées lexicales (mots exprimant les sèmes). Cela correspondrait alors à notre annotation comme interprétation d'un document audiovisuel.

D'autre part, dans la SI, les sèmes mis en place sont virtualisés ou transmis en fonction de contextes. Nous considérons également que les contextes jouent une importance capitale dans la compréhension par exemple de l'annotation d'unités audiovisuelles, en fonction de relations aussi bien temporelles que sémantiques. De la même manière, le sens des mots dans un document varie par rapport au sens hors-document du fait par exemple de la transmission de sèmes. Le principe est le même chez nous, où nous considérons que tout EA est unique, et peut posséder un sens différent de l'EAA duquel il provient, en fonction des EA auxquels il est lié.

Nos dimensions d'analyse, qui *forcent* une analyse cohérente d'un flux nous semblent avoir des résonances avec la notion d'isotopie sémantique. De la même manière, l'appartenance à des dimensions d'analyse différentes d'un EA, en fonction de la tâche de description et d'interprétation en cours est à mettre en liaison avec les taxèmes de la SI dans lesquels sont regroupés les mots en fonction des besoins de description (par exemple *chat, léopard, panthère* si l'on s'intéresse à la zoologie, mais *chat, chien, canari* si ce sont les animaux familiers qui sont visés). Considérer la base de connaissances différemment en fonction des visées de la tâche en cours exprimées par des dimensions d'analyse participe de ce principe d'adaptation.

Au chapitre des parallèles qui pourraient se révéler féconds, citons la possible adaptation de la notion de molécule sémique, c'est à dire d'arrangement de sèmes en graphes pour exprimer des « personnages » d'une interprétation (par exemple /jaune/+sale/ est un personnage récurrent de *l'Assommoir* de Zola). Adapté aux Strates-IA, cela reviendrait à rechercher des régularités de description sous la forme de sous-graphes récurrents).

Enfin, remarquons qu'il pourrait être intéressant d'utiliser les Strates-IA pour des documents

textuels, dont la séquentialité linéaire peut être mise en parallèle avec la séquentialité temporelle des documents audiovisuels. L'adaptation des Strates-IA à des documents non séquentiels (multimédias) est possible, mais les conséquences induites sur le modèle restent à étudier. Remarquons par exemple que la notion de contexte non sémantique doit être étendue d'un axe linéaire à des axes multiples (par exemple de position); et qu'il ne s'agirait sans doute plus seulement de décrire des documents existants, mais d'en assumer également la définition et les instructions de construction...

3 Développements en cours

Nous quittons dans cette dernière section les grandes perspectives que nous avons ouvertes et envisageons plus prosaïquement les perspectives immédiates liées au développement des Strates-IA.

Nous considérons en effet que nous avons posé les bases d'une approche de description de documents audiovisuels originale, permettant de dépasser les descriptions par mots-clé pour pouvoir atteindre (ou non) à la rigueur de systèmes à base de connaissances. Il nous semble important d'accorder une grande liberté à l'annotation et de permettre d'exprimer des inférences contextuelles liées à la tâche de l'utilisateur, qui ne peut être définie une fois pour toutes. Le principe des Strates-IA et les outils contextuels qui leur sont liés nous apparaissent donc comme pertinents et validés, mais ils ne représentent qu'un début dans la jeune histoire des documents audiovisuels numériques.

Beaucoup de travail reste donc à accomplir, et nous listons ci-après quelques-unes des perspectives à court terme des Strates-IA :

- Etude en vraie grandeur de l'annotation sur un jeu d'essai de taille moyenne (10 heures).
- Etude de la mise en place de la base de connaissances, soit incrémentalement, soit par exemple par extraction de terminologie d'un ensemble de textes liés aux documents.
- Etude précise des tâches d'exploitation des Strates-IA.
- Gestion des connaissances de description (schémas de description et dimensions d'analyse) de façon formelle.
- Formalisation plus précise des modèles d'utilisation et de tâches et implantation de ces modèles pour permettre la gestion des modèles de tâches, leur utilisation. La mise en place de cas d'utilisation complet et leur stockage, ainsi que de cas partiels est nécessaire.
- Etude de l'apprentissage dans les Strates-IA : apprentissage de schémas de description, recherche de métriques pour les cas, adaptation de cas d'utilisation retrouvés pour l'aide à l'utilisateur.
- Evolutions de l'algorithme en liaison avec le stockage des description dans un graphe composé de multiples documents XML.
- Etude et évolution en liaison avec les développements de la norme MPEG-7.

Ces perspectives trouvent leur concrétisation dans le cadre d'un projet issu de SESAME. Le projet RECIS⁶ est un projet RNRT et est pour Elöd Egyed-Zsigmond l'occasion de poursuivre l'effort de recherche ici présenté. Une prolongation du projet SESAME est également en cours de discussion.

6. Recherche et Exploration de Contenus Image et Son.

Glossaire

Ce glossaire regroupe quelques termes liés à l'audiovisuel en général, à notre travail sur les Strates-IA, sur l'algorithme de multipropagation (instanciation) et sur le stockage de l'expérience d'utilisation.

Annotation : Action d'annoter un flux ou ensemble des éléments d'annotation liés à une unité audiovisuelle.

Annotation contextuelle : Éléments d'annotation dans le contexte d'une UAV.

Assistant : Agent intelligent aidant l'utilisateur dans l'exploitation des Strates-IA, peut posséder des connaissances qui lui sont propres.

Base de connaissances (BC) : Graphe d'éléments d'annotation abstraits correspondant *a minima* à un thésaurus basé sur une relation de spécialisation.

Base des documents annotés (BDA) : Ensemble des UAV, des EA qui les annotent et de leurs relations.

Caractéristique calculée de haut niveau d'abstraction : Caractéristique calculée verbalisée par un terme : Plan, Chirac, PanoGaucheDroite.

Caractéristique d'annotation : Codage d'un objet d'intérêt audiovisuel (voir primitive, caractéristique calculée de haut niveau d'abstraction, caractéristique interprétée).

Caractéristique interprétée : Caractéristique verbalisée, mise en place par une personne lors d'un acte d'interprétation du flux.

Cas d'utilisation : Description des instances des objets du modèle d'utilisation utilisés dans le cadre d'une tâche dont un modèle de tâche est disponible.

Contexte d'un élément des Strates-IA : Autre élément du graphe global en relation contextuelle avec lui.

Correspondance (instanciation) : Couple associant un sommet d'un graphe potentiel et un sommet du graphe global qui lui est similaire au cours d'une instanciation.

Dimension d'analyse : Expression de la visée de description d'un annotateur comme regroupement d'EEA. Ensemble nommé de méthodes de désignation qui se résoud en un ensemble d'EAA.

Document audiovisuel (DAV) : Généralement lié à des critères de production et de réalisation : superposition de flux vidéo et audio dont le visionnage est possible indépendamment de tout contexte.

Ecriture sur le flux : Action de poser et de mettre en relation des termes sous la forme d'éléments d'annotation.

Élément d'annotation (EA) : Objet des Strates-IA, constitué d'un nom et d'attributs annexes, il annote une UAV, est inscription dans le flux d'un EAA, et est en relation élémentaire avec d'autres EA.

- Élément d'annotation abstrait (EAA)** : Objet des Strates-IA, abstraction d'élément d'annotation en spécifiant le nom et les attributs, organisé dans la base de connaissances.
- Étape (instanciation)** : Ensemble d'états. Le passage à l'étape suivante se fait par extension d'un des états.
- État (instanciation)** : Ensemble de correspondances décrivant une solution partielle de l'instanciation
- Explication (d'un cas d'utilisation)** : Ensemble des relations que celui-ci noue avec le modèle d'utilisation, avec les modèles de tâches dont il découle et ensemble des relations internes non prévues dans le modèle de tâche utilisé.
- Extension d'état (instanciation)** : Ensemble des correspondances permettant d'étendre la solution partielle représentée par l'état.
- Filtre de désignation** : Graphe potentiel caractérisé composé uniquement d'EAA et servant à désigner des éléments de la base de connaissances.
- Flux audiovisuel** : Fichier informatique correspondant à un document audiovisuel temporellement continu.
- Fusion d'états (instanciation)** : Jonction de deux états de graines disjointes, dont l'un s'étend dans l'autre.
- Graine (instanciation)** : Correspondance de départ d'une instanciation.
- Graphe global** : Graphe composé des UAV, EA et EAA d'un système Strates-IA et de leurs relations. Formellement, graphe orienté étiqueté par des objets des Strates-IA.
- Graphe potentiel** : Expression d'un ou plusieurs chemins dans un graphe global. Formellement, graphe d'objets Strates-IA munis de fonctions de similarité permettant de rendre des sommets « génériques » et de relations étendues aux relations temporelles.
- Graphe potentiel caractérisé** : Graphe potentiel nommé dont certains sommets sont nommés et permettent de le manipuler.
- Graphe requête** : Extension de graphe potentiel caractérisé où les EAA sont éventuellement représentés par des dimensions d'analyse.
- Hypothèse de la correspondance initiale (instanciation)** : Hypothèse selon laquelle un sommet au moins d'un graphe potentiel a un correspondant non ambigu dans le graphe global : toute recherche se fait à partir d'un ou de plusieurs sommets connus du graphe.
- Instanciation** : Recherche de sous-graphe du graphe global partiellement isomorphes à un graphe potentiel, aux contraintes de similarités près.
- Méthode de désignation** : Ensemble de filtres de désignation et/ou de méthodes de désignation. Une méthode de désignation se résoud en un ensemble d'EAA.
- Modèle d'utilisation** : Description des objets de base d'un système informatique et des relations qu'ils entretiennent.
- Modèle de tâche** : Description des objets du modèle d'utilisation mobilisés pour une tâche particulière avec les relations qu'ils entretiennent dans le cadre de cette tâche. Un modèle de tâche peut se décomposer en modèles de sous-tâches.
- Montage** : Association temporelle de morceaux de flux audio et vidéo de façon à obtenir des effets audiovisuels contextuels d'ensemble.
- Multipropagation (algorithme de)** : Instanciation de graphe potentiel basée sur une énumération de l'espace des solutions pilotée par une heuristique et des jonctions entre états.
- Objet d'intérêt** : Objet (au sens général du terme) repéré durant l'appréhension du flux.

Objet des Strates-IA : Ensemble de couples attribut/valeur possédant au minimum un Type et un Nom, exprimant une UAV, un EA ou un EAA. Etiquettes des sommets d'un graphe global.

Plan : Suite d'images filmées sans interruption de la caméra (ou entre deux interruptions de la caméra).

Primitive : Caractéristique calculée non verbalisée (histogramme de couleur, texture,...).

Relation élémentaire : Type unique de relation entre éléments d'annotation.

Relation conceptuelle : Relation de la base de connaissance entre éléments d'annotation abstraits.

Relation contextuelle potentielle (RCP) : Graphe potentiel caractérisé décrivant un chemin dans le graphe entre un point de départ et d'arrivée.

Relation d'annotation : Relation entre un EA et l'UAV qu'il annote.

Relation d'inscription dans le flux : Relation entre un EAA et un EA exprimant l'« instantiation » temporellement située d'un concept dans le flux.

Relation de décontextualisation : Relation entre un EA et un EAA exprimant l'abstraction de l'élément d'annotation hors du flux où il se trouvait contextualisé.

Relations temporelles : Relations entre EA déduites des relations temporelles entre les UAV qu'ils annotent (après, pendant, ...). Les relations temporelles sont utilisées explicitement dans les graphes potentiels et les schémas de description.

Scène : Suite de plans consécutifs constituant un tout sous le rapport d'une action dramatique déterminée, obéissant à des règles locales d'agencement de plans, à des techniques de montage.

Schéma de description : Graphe dont les sommets sont étiquetés par des dimensions d'analyse (spécifiant ou non la création d'UAV) et les arêtes par des relations élémentaires $R_e[m, n]$, des relations d'annotation et des relations temporelles.

Strate : Partie de document audiovisuel caractérisée par un instant de début et un instant de fin.

Unité Audiovisuelle (UAV) : Objet des Strates-IA représentant dans le graphe une strate d'un flux.

Unité audiovisuelle étendue : Extension des UAV permettant de prendre en compte des objets audiovisuels à un niveau de décomposition en deça de la simple strate (par ex. pour gérer des objets vidéo MPEG-4).

Valence : Attribut d'un EAA exprimant une possibilité de mise en relation d'un EA avec un EA situé dans son contexte décrit par une relation contextuelle potentielle. La relation peut être directe ou expliquée par un EA.

Vue d'un élément du graphe (interface) : Fenêtre graphique permettant l'appréhension d'une UAV, d'un EA ou d'un EAA dans un contexte particulier.

Bibliographie

- [AAModt et Plaza, 1994] **AAMODT, A. et PLAZA, E.** Case based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 1994, vol. 7, n° 1, pp. 39–59.
- [Adali *et al.*, 1996] **ADALI, S., CANDAN, K., CHEN, S. et al.** Advanced video information system: Data structures and query processing. *Multimedia Systems*, 1996, vol. 4, pp. 172–186.
- [Agius et Angelides, 1999] **AGIUS, H. et ANGELIDES, M.** Developing knowledge-based intelligent multimedia tutoring systems using semantic content-based modelling. *Artificial Intelligence Review*, 1999, vol. 13, pp. 55–83.
- [Aguierre Smith, 1991] **AGUIERRE SMITH, T. G.** *Stratification : Toward a computer representation of the moving image*, Boston, MA : The Media Lab, MIT, 1991, 15 p., Rapport technique.
- [Aguierre-Smith, 1992] **AGUIERRE-SMITH, T. G.** *If you could see what I mean... Descriptions of video in an Anthropologist's video notebook*. Masters' thesis, University of California, Berkeley, 1992. 108 p.
- [Aguierre Smith et Davenport, 1992] **AGUIERRE SMITH, T. G. et DAVENPORT, G.** The stratification system, a design environment for random access video. *In : Proc. Network and Operating System Support for Digital Audio and Video - 3rd International Workshop*, 1992, San Diego, CA, pp. 250–261.
- [Ahanger et Little, 1998] **AHANGER, G. et LITTLE, T.** Automatic composition techniques for video production. *IEEE Transactions on Knowledge and Data Engineering*, 1998, vol. 10, n° 6, pp. 967–987.
- [Aigrain *et al.*, 1995a] **AIGRAIN, P., JOLY, P., LEPAIN, P. et al.** Representation-based user interfaces for the audiovisual library of year 2000. *In : Proc. IS&T SPIE'95 Multimedia Computing and Networking*, Février 1995, San Jose, pp. 35–45.
- [Aigrain *et al.*, 1995b] **AIGRAIN, P., JOLY, P. et LONGUEVILLE, V.** Medium knowledge-based macro-segmentation of video into sequences. *In : IJCAI'95, Workshop on Intelligent Multimedia Information Retrieval*, éd. par Maybury M., Août 1995, Montreal. 14 p.

- [Aigrain *et al.*, 1996] **AIGRAIN, P., PETKOVIC, D. et ZHANG, H.** Content-based representation and retrieval of visual media: A state-of-the-art review. *Multimedia Tools and Applications special issue on Representation and Retrieval of Visual Media*, 1996, vol. 3, n° 3, pp. 179–202.
- [Alaya *et al.*, 1995] **ALAYA, N., HENRY, J. et PINON, J.-M.** Using a corporate memory to support computer assisted authoring. *In: Proceedings of BIWIT'95*, 1995, San Sebastian, Spain, pp. 29–36.
- [Allen, 1983] **ALLEN, J.** Maintaining temporal knowledge about temporal intervals. *Communications of the ACM*, 1983, vol. 26, n° 11, pp. 832–843.
- [Alty, 1991] **ALTY, J.** Multimedia — what is it and how do we exploit it? *In: Human Computer Interaction '91*, Août 1991, Edinburgh, pp. 31–44.
- [Amann *et al.*, 1994] **AMANN, B., SCHOLL, M. et RIZK, A.** Querying typed hypertexts in multiscard/o2. *In: ECHT '94. Proceedings of the 1994 ACM European conference on Hypermedia Technology*, Sep 1994, Edinburgh, pp. 198–205.
- [Ardizzone et La Cascia, 1997] **ARDIZZONE, E. et LA CASCIA, M.** Automatic video database indexing and retrieval. *Multimedia Tools and Applications*, 1997, vol. 4, n° 1, pp. 29–56.
- [Ariki *et al.*, 1998] **ARIKI, Y., OGATA, J. et NISHIDA, M.** News dictation and article classification using automatically extracted announcer utterance. *In: Int. Advanced Multimedia Content Processing Conf.*, Novembre 1998, Osaka, Lecture Notes in Computer Science, vol. 1554, pp. 78–89.
- [Arman *et al.*, 1994] **ARMAN, F., DEPOMMIER, R., HSU, A. et al.** Content-based browsing of video sequences. *In: Second ACM Multimedia Conference*, Août 1994, Anaheim, pp. 97–103.
- [Auffret et Bachimont, 1999] **AUFFRET, G. et BACHIMONT, B.** Audiovisual cultural heritage: from tv and radio archiving to hypermedia publishing. *In: Third European Conference on Research and Advanced Technology for Digital Libraries*, Septembre 1999, Paris, Lecture Notes in Computer Science, vol. 1696, pp. 58–75.
- [Auffret *et al.*, 1999] **AUFFRET, G., CARRIVE, J., CHEVET, O. et al.** Audiovisual-based hypermedia authoring: using structured representations for efficient access to av documents. *In: ACM Hypertext '99*, éd. par Tochtermann K., Westbomke J., Wiil U. K. et al., 1999, Darmstadt, pp. 169–178.
- [Auffret et Prié, 1999] **AUFFRET, G. et PRIÉ, Y.** Managing full-indexed audiovisual documents: a new perspective for the humanities. *Computer and the Humanities, special issue on Digital Images*, 1999. A paraître. 26 p.
- [Auffret *et al.*, 1999b] **AUFFRET, G., PRIÉ, Y. et BACHIMONT, B.** Digitizing tv and radio archives: supporting scholarship by providing new means of access to audiovisual documents. *In: Digital Ressources in the Humanities DRH'99*, Septembre 1999, Londres, Angleterre, pp. 95–99.

-
- [Bachimont, 1996] **BACHIMONT, B.** *Herméneutique matérielle et Artéfacture: des machines qui pensent aux machines qui donnent à penser.* Thèse d'épistémologie, Ecole Polytechnique, France, 1996. 340 p.
- [Bachimont, 1999] **BACHIMONT, B.** Bibliothèques numériques audiovisuelles: des enjeux scientifiques et techniques. *Document Numérique, Numéro Spécial "Les Bibliothèques Numériques"*, 1999, vol. 2, n° 3-4, pp. 219-242.
- [Biébow et Szulman, 1998] **BIÉBOW, B. et SZULMAN, S.** Une approche terminologique pour catégoriser les concepts d'une ontologie. In: *Ingénierie des Connaissances 98*, Mai 1998, Pont-à-Mousson, pp. 51-58.
- [Blackburn et DeRoure, 1998] **BLACKLBURN, S. et DEROURE, D.** A tool for content-based navigation of music. In: *ACM Multimedia 98 Proc.*, Septembre 1998, Bristol, pp. 361-368.
- [Bolle et al., 1998] **BOLLE, R., YEO, B.-L. et YEUNG, M.** Video query: Research directions. *IBM Journal of Research and Development*, 1998, vol. 42, n° 2, pp. 233-252.
- [Bonnet et al., 1989] **BONNET, C., GHIGLIONE, R. et RICHARD, J.-F.** *Traité de psychologie cognitive. Tome 1: Perception, Action, Langage*, Paris: Dunod, 1989. 280 p.
- [Bordwell, 1993] **BORDWELL, D.** *The cinema of Eisenstein*, Cambridge, MA: Harvard university press, 1993. 336 p.
- [Bouet et Djeraba, 1998] **BOUET, M. et DJERABA, C.** Powerful image organization in visual retrieval systems. In: *ACM Multimedia 98 Proc.*, Septembre 1998, Bristol, pp. 315-322.
- [Bourguet, 1992] **BOURGUET, M.-L.** *Conception et réalisation d'une interface de dialogue personne-machine multimodale.* Thèse de doctorat informatique, INPG, Grenoble, 1992. 206 p.
- [Bradtke et Lehnert, 1988] **BRADTKE, S. et LEHNERT, W.** Some experiments with case based search. In: *AAAI-88: The Seventh National Conference on Artificial Intelligence*, Aug 1988, vol. 1, Saint Paul, Minnesota, pp. 133-138.
- [Bres et Jolion, 1999] **BRES, S. et JOLION, J.-M.** Detection of interest points for image indexation. In: *3rd Int. Conf. on Visual Information Systems, Visual'99*, Juin 1999, Amsterdam, The Netherlands, Lecture Notes in Computer Science, n° 1614, pp. 427-434.
- [Brézillon et Cavalcanti, 1997] **BRÉZILLON, P. et CAVALCANTI, M.** Modeling and using context: Report on the first international and interdisciplinary conference context-97. *The Knowledge Engineering Review*, 1997, vol. 12, n° 4, pp. 1-10.
- [Brézillon et Saker, 1999] **BRÉZILLON, P. et SAKER, I.** *Exploring the Contexts of Information Behaviour*, London: Taylor Graham Publishing, 1999, chap. Modeling context in information seeking, pp. 479-493.

- [Brown *et al.*, 1995] **BROWN, M., J.T., F., G.T.F., J. et al.** Automatic content-based retrieval of broadcast news. *In: ACM Multimedia '95*, Novembre 1995, San Fransisco, pp. 35–43.
- [Bruza, 1990] **BRUZA, P.** Hyperindices: a novel aid for searching in hypermedia. *In: ACM European Conference on Hypermedia Technology*, 1990, pp. 109–122.
- [Buijs et Lew, 1999] **BUIJS, J.-M. et LEW, M.** Visual learning of simple semantics in imagescape. *In: Proc. of VISUAL'99*, 1999, Amsterdam, The Netherlands, Lecture Notes in Computer Science, n° 1614, pp. 131–138.
- [Burnard et Sperberg-McQueen, 1994] **BURNARD, L. et SPERBERG-MCQUEEN, C.-M.** (édité par). *TEI P3: Guidelines for Electronic Text Encoding for Interchange*, Chicago and Oxford: Burnard, Sperberg-McQueen, [En ligne], 1994, [24.10.99]. Disponible sur Internet <URL: <http://etext.virginia.edu/TEI.html>>.
- [Bush, 1945] **BUSH, V.** As we may think. *Atlantic Monthly*, 1945, vol. 176, pp. 101–108.
- [Calderan *et al.*, 1982] **CALDERAN, L., HIDOINE, B. et RAMBERT, P.** *Guide pratique pour la description des audiovisuels scientifiques en techniques*, Le Chesnay: INRIA, 1982. 32 pages.
- [Card *et al.*, 1991] **CARD, S., ROBERTSON, G. et MACLINLAY, J.** The information vizualizer: An information workspace. *In: Proceedings of ACM CHI'91*, 1991, New Orleans, LA, pp. 181–188.
- [Carlos et Uehara, 1998] **CARLOS, R. et UEHARA, K.** Video summarization based on semantic representation. *In: Int. Advanced Multimedia Content Processing Conf.*, Novembre 1998, Osaka, Lecture Notes in Computer Science, vol. 1554, pp. 1–16.
- [Carpineto et Romano, 1996] **CARPINETO, C. et ROMANO, G.** A lattice conceptual clustering system and its application to browsing retrieval. *Machine Learning*, 1996, vol. 24, n° 2, pp. 95–122.
- [Carrer *et al.*, 1997] **CARRER, M., LIGRESTI, L., AHANGER, G. et al.** An annotation engine for supporting video database population. *Journal of Multimedia Tools and Applications*, 1997, vol. 5, n° 3, pp. 233–258.
- [Carrive *et al.*, 1998] **CARRIVE, J., PACHET, F. et RONFARD, R.** Using description logics for indexing audiovisual documents. *In: Proceedings of the International Workshop on Description Logics - DL-98*, 1998, Trento, Italy, pp. 116–120.
- [Cavazza, 1998] **CAVAZZA, M.** Textual semantics and corpus specific lexicons. *In: First International Conference on Language Resources and Evaluation, Workshop on Adapting Lexical and Corpus Resources to Sublanguages and Application*, Mai 1998, Grenada, Spain. Non numéroté, 8 p.
- [Cavazza *et al.*, 1998] **CAVAZZA, M., GREEN, R. et PALMER, I.** Multimedia semantic features and image content description. *In: International Conference on Multimedia Modeling, MMM '98*, Octobre 1998, Lausanne, pp. 29–46.

-
- [Chahuneau *et al.*, 1992] **CHAHUNEAU, F., LÉCLUSE, C., STIEGLER, B. et al.** Prototyping the ultimate tool for scholarly qualitative research on texts. *In: 8th Annual conference of the UW Centre for the New Oxford English Dictionary and Text Research*, Oct 1992, Waterloo, Ontario, Canada. 8 p.
- [Chandrasekaran *et al.*, 1998] **CHANDRASEKARAN, B., JOSEPHSON, J. et Richard Benjamins, V.** Ontology of tasks and methods. *In: Proceedings of the Workshop on Applications of Ontologies and Problem-Solving Methods, ECAI'98*, 1998, Brighton, UK, pp. 31–43.
- [Chang *et al.*, 1997b] **CHANG, S., SMITH, J., BEIGI, M. et al.** Visual information retrieval from large distributed online repositories. *Communications of the ACM*, 1997, vol. 40, n° 12, pp. 63–71.
- [Chang *et al.*, 1997a] **CHANG, S.-F., SMITH, J. R., MENG, H. J. et al.** Finding images/video in large archives. *CNRI Digital Library Magazine*, 1997. Disponible sur Internet <URL: <http://www.dlib.org/dlib/february97/columbia/02chang.html>>.
- [Chen *et al.*, 1998a] **CHEN, J.-Y., TASKIRAN, C., DELP, E. et al.** Vibe: A new paradigm for video database browsing and search. *In: IEEE Workshop on Content-Based Access of Image & Video Libraries (CBAIVL'98)*, Juin 1998, Santa Barbara, CA, pp. 96–100.
- [Chen *et al.*, 1998b] **CHEN, L., FONTAINE, D. et HAMMOUD, R.** La segmentation sémantique de la vidéo basée sur les indices spatio-temporels. *In: 4èmes Journées d'Etudes et d'Echanges «Compression et Représentation des Signaux Audiovisuels»*, CNET — France Télécom, Juin 1998, Lannion, pp. 67–75.
- [Cherfaoui et Bertin, 1994] **CHERFAOUI, M. et BERTIN, C.** Two-stage strategy for indexing and presenting video. *In: Proc. of the SPIE Conference on Storage and Retrieval for Image and Video Databases II*, 1994, San Jose, CA, SPIE, vol. 2185, pp. 174–184.
- [Chevrier, 1995] **CHEVRIER, H.** *Le langage du cinéma narratif*, Laval, Québec : Les 400 coups, 1995, *Cinéma*. 172 p.
- [Chiararella, 1997] **CHIARAMELLA, Y.** Browsing and querying: two complementary approaches for multimedia information retrieval. *In: Hypertext — Information Retrieval — Multimedia HIM'97*, [En ligne], Septembre 1997, [24.10.99], Dortmund. Conférence invitée. Disponible sur Internet : <URL: <http://lrb.cs.uni-dortmund.de/HIM97/Konferenzband/Chiararella/Paper0.htm>>.
- [Chikh et Tazi, 1999] **CHIKH, A. et TAZI, S.** Annotations structurées d'aide à la reconstruction de documents, le modèle asard. *In: Conférence Internationale sur le Document Electronique CIDE'99*, Juillet 1999, Damas, pp. 205–218.
- [Chion, 1994] **CHION, M.** *L'audiovision : son et image au cinéma (2^e édition)*, Paris : Nathan Universités, 1994. 186 p.
- [Chiueh, 1994] **CHIUEH, T.-C.** Content-based image indexing. *In: 20th VLDB Conference*, 1994, Santiago, Chile, pp. 582–593.

- [Cho et Kim, 1992] **CHO, C. et KIM, J.** Recognizing 3-d objects by forward checking constrained tree search. *Pattern Recognition Letters*, 1992, vol. 13, n° 8, pp. 587–597.
- [Christel, 1995] **CHRISTEL, M.** Addressing the contents of video in a digital library. In: *Electronic Proceedings of the ACM Workshop on Effective Abstraction in Multimedia: Layout, Presentation, and Interaction. Held in conjunction with ACM MM '95*, [En ligne], Novembre 1995, [24.10.99], San Francisco, California. Disponible sur Internet <URL: <http://www.cs.tufts.edu/isabel/christel/christel.html>>.
- [Christel et Martin, 1998] **CHRISTEL, M. et MARTIN, D.** Information visualization within a digital video library. *Journal of Intelligent Information Systems, Special issue on Information Visualization*, 1998, vol. 11, n° 3, pp. 235–257.
- [Chua et Ruan, 1995] **CHUA, T.-S. et RUAN, L.-Q.** A video retrieval and sequencing system. *ACM Transactions on Information Systems*, 1995, vol. 13, n° 4, pp. 372–407.
- [Cluzeau-Ciry, 1988] **CLUZEAU-CIRY, M.** Typologie des utilisateurs et des utilisations d'une banque d'images. *Le documentaliste*, 1988, vol. 25, n° 3, pp. 115–120.
- [Colin, 1992] **COLIN, M.** *Cinéma, télévision, cognition*, Nancy : Presses Universitaires de Nancy, 1992, *Processus discursifs, langage et cognition*. 152 p.
- [Cook et Holder, 1994] **COOK, D. et HOLDER, L.** Substructure discovery using minimum description length and background knowledge. *Journal of Artificial Intelligence Research*, 1994, pp. 231–255.
- [Cordella et al., 1998] **CORDELLA, L., FOGGIA, P., SANSONE, C. et al.** Subgraph transformations for inexact matching of attributed relational graphs. *Computing supplement*, 1998, vol. 12, pp. 43–52.
- [Corridoni et al., 1996] **CORRIDONI, J. M., DEL BIMBO, A., LUCARELLA, D. et al.** Multi-perspective navigation of movies. *Journal of Visual Languages and Computing*, 1996, vol. 7, pp. 445–466.
- [Corvaisier et al., 1998] **CORVAISIER, F., MILLE, A. et PINON, J.-M.** Radix 2, assistance à la recherche d'information documentaire sur le web. In: *IC'98, Ingénierie des Connaissances, Pont-à-Mousson, France*, Mai 1998, pp. 153–163.
- [Cunliffe et al., 1997] **CUNLIFFE, D., TAYLOR, C. et TUDHOPE, D.** Query-based navigation in semantically indexed hypermedia. In: *Hypertext'97*, 1997, Southampton, pp. 87–95.
- [Dalamagas, 1998] **DALAMAGAS, T.** Nhs: A tool for automatic construction of news hypertext. In: *Electronic Workshops in Computing: Proceedings of IRSG'98*, éd. par Dunlop M.-D., [En ligne], 1998, [24.10.99], Grenoble, France. Disponible sur Internet (payant) <URL: <http://www.ewic.org.uk/ewic/workshop/view.cfm/IRSG-98>>.

-
- [Davenport, 1996] **DAVENPORT, G.** Indexes are out, models are in. *IEEE Multimedia*, 1996, vol. 3, n° 3, pp. 10–15.
- [Davenport *et al.*, 1991] **DAVENPORT, G., AGUIERRE SMITH, T. et PINCEVER, N.** Cinematic primitives for multimedia. *IEEE Computer Graphics and Applications*, 1991, vol. 11, n° 4, pp. 67–75.
- [Davenport et Murtaugh, 1995] **DAVENPORT, G. et MURTAUGH, M.** Context : toward the evolving documentary. *In : ACM Multimedia '95*, Novembre 1995, San Fransisco, pp. 381–389.
- [Davis, 1993] **DAVIS, M.** Media streams: An iconic visual language for video annotation. *In : Proceedings of the 1993 IEEE Symposium on Visual Languages*, Août 1993, Bergen, Norway, pp. 196–203.
- [Decleir *et al.*, 1999] **DECLAIR, C., HACID, M.-S. et KOULOUMDJIAN, J.** A database approach for modeling and querying video data. *In : Fifteenth International Conference on Data Engineering*, éd. par Kitsuregawa M., Maciaszek L., Papazoglou M. et al., IEEE, Mars 1999, Sydney, Australia, pp. 6–13.
- [Del Bimbo *et al.*, 1995] **DEL BIMBO, A., VICARIO, E. et ZINGONI, D.** Symbolic description and visual querying of image sequences using spatio-temporal logic. *IEEE Transactions on Knowledge and Data Engineering*, 1995, vol. 7, n° 4, pp. 609–621.
- [Delanoy, 1996] **DELANOY, R.** Supervised learning of tools for content-based search of image databases. *Proceedings of the SPIE - The Internaional Society for Optical Engineering Co.*, 1996, vol. 2670, pp. 194–205.
- [Denel *et al.*, 1994] **DENEL, F., PIÉJUT, G. et RODES, J.-M.** Le dépôt légal de la radio et de la télévision. *Les dossiers de l'audiovisuel, Numéro spécial*, La documentation française, 1994, n° 54. 118 p.
- [Desclés *et al.*, 1997] **DESCLÉS, J., CARTIER, E., JACKIEWICZ, A. et al.** Textual processing and contextual exploration method. *In : Proc. of the International and Interdisciplinary Conference on Modeling and Using Context*, Février 1997, Rio de Janeiro, Brazil, pp. 189–197.
- [Dimitrova et Golshani, 1995] **DIMITROVA, N. et GOLSHANI, F.** Motion recovery for video content classification. *ACM Transactions on Information Systems*, 1995, vol. 13, n° 4, pp. 409–439.
- [Domeshek *et al.*, 1996] **DOMESHEK, E., KEDAR, S. et GORDON, A.** Interactive information retrieval systems with minimalist representation. *In : AAAI-96:Thirteenth National Conference On Artificial Intelligence*, 1996, Portland, OR., pp. 419–446.
- [Düffing, 1999] **DÜFFING, G.** Analyse et organisation de corpus pour une recherche thématique-visuelle d'images. *In : Actes Inforsid'99*, 1999, Toulon, pp. 427–446.

- [Dunlop et Rijsbergen, 1993] **DUNLOP, M. et RIJSBERGEN, C.** Hypermedia and free text retrieval. *Information Processing and Management*, 1993, vol. 29, n° 3, pp. 287–298.
- [Edmondson, 1998] **EDMONDSON, R.** *A philosophy of audiovisual archiving*, Paris: Unesco, 1998, Rapport technique. 60 p.
- [Egyed-Zsigmond, 1999b] **EGYED-ZSIGMOND, E.** *Représentation et exploitation de graphes de description de documents multimédias dans un langage structuré*. INSA-Lyon, Rapport de projet de fin d'études, Département Informatique, INSA-Lyon, Septembre 1999. 71 p.
- [Egyed-Zsigmond, 1999a] **EGYED-ZSIGMOND, E.** *Représentation et exploitation de graphes de description de documents multimédias dans un langage structuré*. INSA-Lyon, Rapport de dea, DEA d'informatique de Lyon, Juin 1999. 53 p.
- [Ehrig, 1992] **EHRIG, H.** Introduction to graph grammars with applications to semantic networks. *Computer and Mathematics with applications*, 1992, vol. 23, pp. 557–572.
- [Etievent et al., 1999] **ETIEVENT, A., LEBOURGEOIS, F. et JOLION, J.-M.** Assisted video sequences indexing: shot detection and motion analysis based on interest points. *In: 10th Int. Conf. on Image Analysis and Processing, ICIAP99*, Septembre 1999, Venice, pp. 1059–1062.
- [Falkenhainer et al., 1989] **FALKENHAINER, B., FORBUS, K. et GENTNER, D.** The structure mapping engine: Algorithms and examples. *Artificial Intelligence*, 1989, vol. 41, pp. 1–63.
- [Faron et al., 1996] **FARON, C., AKOULCHINA, I. et GANASCIA, J.** Integrating knowledge in hypermedia: the satelit system. *In: Symposium on Robotics and Cybernetics. CESA '96 IMACS Multiconference. Computational Engineering in Systems Applications*, Juillet 1996, Lille, pp. 94–102.
- [Faudemay et al., 1997] **FAUDEMAY, P., CHEN, L., MONTACIÉ, C. et al.** Segmentation multi-canaux de vidéos en séquences. *In: CORESA 97*, [En ligne], 1997, [24.10.99], Issy-les-Moulineaux. Disponible sur Internet <URL: <http://www-poleia.lip6.fr/PAROLE/articles/coresa97fp.ps.gz>>.
- [Faudemay et al., 1996] **FAUDEMAY, P., CHEN, L., MONTACIÉ, C. et al.** Multi-channel video segmentation. *In: Int. Conf. on Multimedia Storage and Archiving Systems, SPIE Symposium*, éd. par Kuo C.-C., Novembre 1996, Boston, MA, SPIE, vol. 2916, pp. 252–264.
- [Fellbaum, 1998] **FELLBAUM, C.** (édité par). *WordNet — An Electronic Lexical Database*, Cambridge, MA: The MIT Press, 1998, *Language, Speech, and Communication*. 423 p.
- [Ferro et Planchais, 1997] **FERRO, M. et PLANCHAIS, J.** *Les médias et l'histoire*, Paris: CFPJ ed., 1997. 166 p.
- [Fillmore, 1968] **FILLMORE, C.** *Universals in linguistic theory*, New-York: Holt, Rinehart and Winston, 1968, chap. The case for case, pp. 1–90.

-
- [Fischer *et al.*, 1995] **FISCHER, S., R., L. et EFFELBERG, W.** Automatic recognition of film genres. *In : ACM Multimedia '95*, Novembre 1995, San Fransisco, pp. 295–301.
- [Fisher, 1990] **FISHER, D.** *Readings in Machine Learning*, San Francisco, CA : Morgan Kaufmann, 1990, chap. Knowledge acquisition via incremental conceptual clustering, pp. 267–283.
- [Flickner *et al.*, 1995] **FLICKNER, M., SAWHNEY, H., NIBLACK, W. et al.** Query by image and video content : the QBIC system. *IEEE Computer*, 1995, vol. 28, n° 9, pp. 23–32.
- [Foote, 1999] **FOOTE, J.** An overview of audio information systems. *Multimedia Systems*, 1999, n° 7, pp. 2–10.
- [Foote *et al.*, 1998] **FOOTE, J., BORECZKY, J., GIRGENSOHN, A. et al.** An intelligent media browser using automatic multimodal analysis. *In : ACM Multimedia 98 Proc.*, Septembre 1998, Bristol, pp. 375–380.
- [Fourel *et al.*, 1998] **FOUREL, F., MULHEM, P. et BRUANDET, M.-F.** A generic framework for structured document access. *In : 9th International Conference on Database and Expert Systems Applications (DEXA '98)*, Août 1998, Vienna, pp. 521–530.
- [Foutain et Tan, 1998] **FOUTAIN, S. et TAN, T.** Content based annotation and retrieval in raid. *In : Electronic Workshops in Computing: Proceedings of IRSG'98*, éd. par Dunlop M.-D., [En ligne], 1998, [24.10.99], Grenoble, France. Disponible sur Internet (payant) <URL: <http://www.ewic.org.uk/ewic/workshop/view.cfm/IRSG-98>>.
- [Furuta, 1997] **FURUTA, R.** What can digital libraries teach us about hypertext? *ACM SIGLINK Newsletter*, 1997, vol. 6, n° 3, pp. 7–9.
- [Gabrielli, 1998] **GABRIELLI, S.** Social and cognitive factors in design and evaluation of multimedia systems. *In : Electronic Workshops in Computing: Proceedings of IRSG'98*, éd. par Dunlop M.-D., [En ligne], 1998, [24.10.99], Grenoble, France. Disponible sur Internet (payant) <URL: <http://www.ewic.org.uk/ewic/workshop/view.cfm/IRSG-98>>.
- [Gardies, 1993] **GARDIES, A.** *Le récit filmique*, Paris : Hachette Supérieur, 1993. 155 p.
- [Geißler, 1995] **GEISSLER, J.** Surfing the movie space — advanced navigation in movie-only hypermedia. *In : ACM Multimedia '95*, Novembre 1995, San Fransisco, pp. 391–400.
- [Genest, 1999] **GENEST, D.** Vers un système d'information documentaire basé sur les graphes conceptuels. *In : Actes de INFORSID'99*, Juin 1999, Toulon, pp. 115–131.
- [Goldfarb, 1990] **GOLDFARB, C. F.** *The SGML Handbook*, Oxford : Clarendon Press, 1990. 663 p.
- [Groensteen, 1997] **GROENSTEEN, T.** Plaisir de la bande dessinée. *9ème Art*, 1997, n° 2, pp. 14–21.

- [Grosky, 1997] **GROSKY, W.** Managing multimedia information in database systems. *Communications of the ACM*, 1997, vol. 40, n° 12, pp. 73–80.
- [Groupe Mu, 1992] **GROUPE MU.** *Traité du signe visuel. Pour une rhétorique de l'image*, Paris : Seuil, 1992. 504 p.
- [Gupta et al., 1997] **GUPTA, A., SANTINI, S. et JAIN, R.** In search of information in visual media. *Communications of the ACM*, 1997, vol. 40, n° 12, pp. 35–42.
- [Haddad et al., 1996] **HADDAD, H.** *Indexation et Modélisation de Documents Vidéo : un Etat de l'Art*, Grenoble: MRIM, CLIPS-IMAG, 1996, 86 p., Rapport technique.
- [Haddad, 1996] **HADDAD, H., BERRUT, C. et BRUANDET, M.-L.** Un modèle vectoriel de recherche d'informations adapté aux documents vidéo. In: *CORESA 96, Compression et Représentation des Signaux Audiovisuels*, Février 1996, Grenoble, France, pp. 281–289.
- [Harman, 1992] **HARMAN, D.** Relevance feedback revisited. In: *ACM SIGIR International Conference on Research en Development in Information Retrieval*, 1992, Copenhagen, pp. 1–10.
- [Heinrichs et al., 1999] **HEINRICHS, A., KOUBAROULIS, D., LEVIENAISE-OBADIA, B. et al.** *Recherche Robuste d'Images par Similarité Pré-attentive*, Villeurbanne: RFV, Insa-Lyon, 1999, 10 p., Rapport de recherche n° RR 04.99. Soumis a RIAO'2000.
- [Hibino et Rundensteiner, 1996] **HIBINO, S. et RUNDENSTEINER, E.** *Multimedia Database Systems: Design and Implementation Strategies*, Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996, chap. A Visual Multimedia Query Language for Temporal Analysis of Video Data, pp. 123–159.
- [Hibino et Rundensteiner, 1997] **HIBINO, S. et RUNDENSTEINER, E.** *Intelligent Multimedia Information Retrieval*, Boston, MA : MIT Press, 1997, chap. Interactive Visualizations for Temporal Analysis: Application to CSCW Multimedia Data, pp. 313–335.
- [Hibino et Rundensteiner, 1998] **HIBINO, S. et RUNDENSTEINER, E.** Comparing MMVIS to a timeline for temporal trend analysis of video data. In: *Advanced Visual Interfaces 1998 (AVI'98) Conference Proceedings*, 1998, Acquila, Italy, pp. 195–204.
- [Hjelsvold et al., 1995a] **HJELSVOLD, R., LANGORGEN, S., MIDTSTRAUM, R. et al.** Integrated video archive tools. In: *ACM Multimedia '95*, ACM, 1995, San Francisco, CA, pp. 283–293.
- [Hjelsvold et al., 1995b] **HJELSVOLD, R., MIDTSTRAUM, R. et SANDSTA, O.** Searching and browsing a shared video database. In: *Proc. of the Int. Workshop on Multimedia Database Management Systems*, 1995, New-York, pp. 90–98.
- [Hjelsvold et al., 1995c] **HJELSVOLD, R., MIDTSTRAUM, R. et SANDSTA, O.** A temporal foundation of video databases. In: *First International Workshop on Temporal Databases*, Septembre 1995, Zürich, pp. 295–314.

-
- [Hjesvold et Midtstraum, 1994] **HJESVOLD, R. et MIDTSTRAUM, R.** Modelling and querying video data. *In: 20th VLDB Conference, 1994, Santiago, Chile*, pp. 686–694.
- [Holm Nelson, 1965] **HOLM NELSON, T.** A file structure for the complex, the changing and the indeterminate. *In: ACM Proceedings of the 20th National Conference*, éd. par ACM, 1965, New-York, pp. 84–100.
- [Horaud et Skordas, 1989] **HORAUD, R. et SKORDAS, T.** Stereo correspondence through feature grouping and maximal cliques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, vol. 11, n° 11, pp. 1168–1180.
- [Irani et al., 1996] **IRANI, M., ANANDAN, P., BERGEN, J. et al.** Mosaic based representations of video sequences and their applications. *Image Communication, Special issue on Image and Video Semantics: Processing, Analysis and Application*, 1996, vol. 8, n° 4, pp. 327–351.
- [Jacquinot, 1977] **JACQUINOT, G.** *Image et pédagogie: analyse sémiologique du film à intention didactique*, Paris: Presses universitaires de France, 1977. 200 p.
- [Jacquinot, 1985] **JACQUINOT, G.** *L'école devant les écrans*, Paris: E.S.F., 1985. 135 p.
- [Jain, 1996] **JAIN, R.** *Multimedia Systems and Techniques*, Norwell, MA, USA: Kluwer Academic Publishers, 1996, chap. InfoScopes: Multimedia information systems, pp. 217–254.
- [Jain et Hampapur, 1994] **JAIN, R. et HAMPAPUR, A.** Metadata in video databases. *SIGMOD RECORD*, 1994, vol. 23, n° 4, pp. 27–41.
- [Jost, 1992] **JOST, F.** *Un monde à notre image: énonciation, cinéma, télévision*, Paris: Meridiens Klincksiek, 1992. 140 p.
- [Kelly et al., 1995] **KELLY, P. H., CANNON, M. et HUSH, D.-R.** Query by image example: the candid approach. *In: Proceedings of SPIE Storage and Retrieval of Image and Video Databases III*, 1995, pp. 238–248.
- [Kim et Kak, 1991] **KIM, W. et KAK, A.** 3-D object recognition using bipartite matchnig embedded in discrete relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, vol. 13, pp. 224–251.
- [Kimber, 1995] **KIMBER, E.-W.** *Practical Hypermedia: An Introduction to HyTime*, New York: Prentice-Hall Professional Technical Reference, 1995. 250p.
- [Kitchen et Rosenfeld, 1979] **KITCHEN, L. et ROSENFELD, A.** Discrete relaxation for matching relational structures. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, vol. 9, n° 12, pp. 869–874.

- [Kobla *et al.*, 1997] **KOBLA, V., DOERMANN, D. et FALOUTSOS, C.** Videotrails: Representing and visualizing structure in video sequences. *In: ACM Multimedia 97*, Novembre 1997, Seattle, pp. 335–346.
- [Koenen, 1999] **KOENEN, R.** Mpeg-4: Multimedia for our time. *IEEE Spectrum*, 1999, vol. 36, n° 2, pp. 26–33.
- [Kuo et Chen, 1996] **KUO, T. C. et CHEN, A. L.** Indexing, query interface and query processing for venus: A video database system. *In: CODAS'96*, Novembre 1996, Kyoto, Japan. 8 p.
- [La Cascia et Ardizzone, 1996] **LA CASCIA, M. et ARDIZZONE, E.** JACOB: Just a content-based query system for video databases. *In: Proceedings ICASSP-96*, 1996, Atlanta, GA, pp. 56–71.
- [Li *et al.*, 1996] **LI, W., GAUCH, S., GAUCH, J. et al.** Vision: A digital video library. *In: ACM Digital Libraries 96*, 1996, Bethesda, Maryland, pp. 19–27.
- [Lieberman, 1994] **LIEBERMAN, H.** A user interface for knowledge acquisition from video. *In: Proc. of AAAI-94 Conference*, Août 1994, Seattle, WA, pp. 527–534.
- [Lienhart *et al.*, 1997] **LIENHART, R., PFEIFFER, S. et EFFELSBERG, W.** Video abstracting. *Communications of the ACM*, 1997, vol. 40, n° 12, pp. 55–62.
- [Lindley et Vercoustre, 1998] **LINDLEY, C. et VERCOUSTRE, A.** Intelligent video synthesis using virtual video prescriptions. *In: Proc. of the International Conference on Computational Intelligence and Multimedia Applications*, Février 1998, Melbourne, Australia, pp. 661–666.
- [Martin, 1994] **MARTIN, P.** *La méthodologie d'acquisition des connaissances KADS et les explications*, Sophia-Antipolis: INRIA, Janvier 1994, 107 p., Rapport de recherche n° 2179.
- [Martin, 1996] **MARTIN, P.** *Exploitation de graphes conceptuels et de documents structurés et hypertextes pour l'acquisition de connaissances et la recherche d'informations*. Thèse de doctorat, Université de Nice — Sophia Antipolis, 1996. 300 p.
- [Martin et Alpay, 1996] **MARTIN, P. et ALPAY, L.** Conceptual structures and structured documents. *In: Proceedings of ICCS'96, 4th International Conference on Conceptual Structures*, Août 1996, Sydney, Australia, Lecture Notes in Artificial Intelligence, vol. 1115, pp. 145–159.
- [Massey et Bender, 1996] **MASSEY, M. et BENDER, W.** Salient stills: Process and practice. *IBM Systems Journal*, 1996, vol. 35, n° 3–4, pp. 557–573.
- [Mc Carthy, 1993] **MC CARTHY, J.** Notes on formalizing context. *In: Proc. of the 13th IJCAI*, 1993, Chambéry, France, pp. 555–560.
- [McHugh *et al.*, 1997] **MCHUGH, J., ABITEBOUL, S., GOLDMAN, R. et al.** Lore: a database management system for semistructured data. *SIGMOD Record*, 1997, vol. 3, pp. 54–66.

-
- [Mechkour, 1995] **MECHKOUR, M.** Emir2. an extended model for image representation and retrieval. *In: DEXA '95. Database and EXpert system Applications*, Septembre 1995, London, pp. 395–404.
- [Mehlhorn et Näher, 1992] **MEHLHORN, K. et NÄHER, S.** Algorithm design and software libraries: Recents developments in leda project-algorithms. *In: IFIP'92*, 1992, Madrid, Spain, pp. 493–505.
- [Mehlhorn et al., 1998] **MEHLHORN, K., NÄHER, S., SEEL, M. et al.** *The LEDA User Manual, Version 3.7.1*, Saarbrücken, Germany: Max-Planck-Institut für Informatik, 1998. 415 p.
- [Merialdo, 1998] **MERIALDO, B.** Utilisation de corpus pour l'indexation vidéo de journaux télévisés. *In: RFIA '98*, 1998, Clermont-Ferrand, pp. 37–46.
- [Merlino et al., 1997] **MERLINO, A., MOREY, D. et MAYBURY, M.** Broadcast news navigation using story segmentation. *In: ACM Multimedia 97*, Novembre 1997, Seattle, pp. 381–391.
- [Messmer, 1996] **MESSMER, B.** *Efficient Graph Matching Algorithms for preprocessed Model Graph*. Thèse de Doctorat, Institute of Computer Science and Applied Mathematics, University of Bern, 1996. 163 p.
- [Metz, 1968] **METZ, C.** *Essais sur la signification au cinéma I*, Paris: Klincksieck, 1968. 227 p.
- [Metz, 1973] **METZ, C.** *Essais sur la signification au cinéma II*, Paris: Klincksieck, 1973. 219 p.
- [Metz, 1977] **METZ, C.** *Langage et cinéma*, Paris: Larousse, 1977. 238 p.
- [Michard, 1999] **MICHARD, A.** *XML Langage et applications*, Paris: Eyrolles, 1999. 361 p.
- [Michel, 1997] **MICHEL, J.** L'apport de l'av (analyse de la valeur) à une réflexion sur l'avenir de la tradition av (audiovisuel). *In: Colloque de Monte Verita, L'avenir de la tradition audiovisuelle: stratégies de collaboration, de coordination et d'intégration*, Memoriav, [En ligne], Mai 1997, [24.10.99], Bern. Disponible sur Internet <URL: http://www.memoriav.ch/d_doc_aktiv/d-jmichel.htm>.
- [Mille et al., 1999] **MILLE, A., FUCHS, B. et CHIRON, B.** Raisonnement fondé sur l'expérience: un nouveau paradigme en supervision industrielle? *Revue d'intelligence artificielle*, 1999, vol. 13, pp. 97–128.
- [Mille et al., 1998] **MILLE, A., PRIÉ, Y., ETIÉVENT, E. et al.** *Le démonstrateur SESAME (niveau 1)*, Lyon: CPE-Lyon, INSA-Lyon, Dec 1998, 14 p., Rapport contractuel SESAME n° 4–6.
- [Minka, 1996] **MINKA, T.** *An Image Database System that Learns From User Interaction*. Cambridge, MA, Masters' thesis, MIT, Février 1996. 55 p.
- [Minka et Picard, 1997] **MINKA, T.-P. et PICARD, R.-W.** Interactive learning using a "society of models". *Pattern Recognition, special issue on Image Databases*, 1997, vol. 30, n° 4, pp. 565–581.

- [Mostefaoui *et al.*, 1999] **MOSTEFAOUI, A., PEREZ, C. et BRUNIE, L.** Serveur de séquences audiovisuelles parallèle sur réseau haut débit : concepts et expérimentations. *In: 11th French Meeting on Architecture & Systems Parallelism*, Juin 1999, pp. 127–132.
- [MPEG-7-Req, 1999] **MPEG.** Mpeg-7 requirements v.9, iso/iec jtc1/sc29/wg11, [En ligne], Juillet 1999, [24.10.99]. Disponible Internet :<URL: http://drogo.cselt.stet.it/mpeg/public/mpeg-7_requirements.zip>.
- [Mugnier et Chein, 1996] **MUGNIER, M.-L. et CHEIN, M.** Représenter des connaissances et raisonner avec des graphes. *Revue d'intelligence artificielle*, Hermès, 1996, vol. 10, n° 1, pp. 7–56.
- [Myaeng et Lopez-Lopez, 1997] **MYAENG, S. et LOPEZ-LOPEZ, A.** Conceptual graph matching : a flexible algorithm and experiments. *Journal of Experimental and Theoretical Artificial Intelligence*, 1997, vol. 4, pp. 107–126.
- [Nabeshima *et al.*, 1998] **NABESHIMA, S., OKAMURA, K., KAKIUCHI, T. et al.** Extended digital video broadcasting with time-lined hypermedia. *In: Int. Advanced Multimedia Content Processing Conf.*, Novembre 1998, Osaka, Lecture Notes in Computer Science, vol. 1554, pp. 240–255.
- [Nack et Parkes, 1997] **NACK, F. et PARKES, A.** Towards the automated editing of theme-oriented video sequences. *Applied Artificial Intelligence, Special Issue on Entertainment and AI/Artificial Life*, 1997, n° 11, pp. 331–366.
- [Nack et Steinmetz, 1998] **NACK, F. et STEINMETZ, A.** Approaches on intelligent video production. *In: Proceedings of ECAI-98 Workshop on AI/Alife and Entertainment*, ECAI, Août 1998, Brighton, UK, pp. 48–56.
- [Nakamura et Kanade, 1997] **NAKAMURA, Y. et KANADE, T.** Semantic analysis for video contents extraction — spotting by association in news video. *In: ACM Multimedia 97*, Novembre 1997, Seattle, pp. 393–401.
- [Nanard et Nanard, 1991] **NANARD, J. et NANARD, M.** Using structured types to incorporate knowledge in hypertext. *In: ACM Hypertext '91*, Décembre 1991, San Antonio, pp. 329–343.
- [Nanard et Nanard, 1995] **NANARD, J. et NANARD, M.** Adding macroscopic semantics to anchors in knowledge-based hypertext. *Int. J. Human-Computer Studies*, 1995, vol. 43, pp. 363–382.
- [Nastar *et al.*, 1998] **NASTAR, C., MITSCHKE, M., MEILHAC, C. et al.** Surfimage: a flexible content-based image retrieval system. *In: ACM Multimedia 98*, Septembre 1998, Bristol, pp. 339–344.
- [Nigay et Coutaz, 1996] **NIGAY, L. et COUTAZ, J.** Espaces conceptuels pour l'interaction multi-média et multimodale. *Techniques et sciences informatiques*, 1996, vol. 15, n° 9, pp. 1195–1225.

-
- [Oomoto et Tanaka, 1993] **OOMOTO, E. et TANAKA, K.** Ovid: Design and implementation of a video-object database system. *IEEE Transactions on Knowledge and Data Engineering*, 1993, vol. 5, n° 4, pp. 629–643.
- [Picard, 1996] **PICARD, R. W.** A society of models for video and image libraries. *IBM Systems Journal*, 1996, vol. 35, n° 3–4, pp. 292–312.
- [Picard *et al.*, 1996] **PICARD, R. W., MINKA, T. et SZUMMER, M.** Modeling user subjectivity in image libraries. *In: Proc. International Conference on Image Processing, vol.2*, Septembre 1996, Lausanne, Switzerland, pp. 777–780.
- [Pinon *et al.*, 1997] **PINON, J.-M., CALABRETTO, S. et POULLET, L.** Document semantic model: An experiment with patient medical record. *In: Electronic Publishing '97 – New Models and Opportunities*, éd. par Rowland F. et Meadows J., Avril 1997, Canterbury, UK, pp. 262–272.
- [Ponceleon *et al.*, 1998] **PONCELEON, D., SRINIVASAN, S., AMIR, A. et al.** Key to effective video retrieval: Effective cataloging and browsing. *In: ACM Multimedia 98 Proc.*, Septembre 1998, pp. 99–107.
- [Poole, 1993] **POOLE, J.** Similarity in legal case based reasoning as degree of matching in conceptual graphs. *In: Proceedings: First European Workshop on Case-Based Reasoning (EWCBR'93)*, 1993, Kaiserslauten, Germany, pp. 54–58.
- [Poulet *et al.*, 1997] **POULLET, L., PINON, J. et CALABRETTO, S.** Semantic structuring of documents. *In: Proceedings of IEEE Conference on Data Management Systems, BIWIT'97*, Juillet 1997, Biarritz, France, pp. 118–124.
- [Prié, 1995] **PRIÉ, Y.** *Contribution à une clarification des rapports entre sémantique interprétative et informatique.* Brest, Mémoire de DEA Informatique, option Intelligence Artificielle et Sciences Cognitives, Université de Rennes 1, Sep 1995. 99 p.
- [Prié *et al.*, 1998a] **PRIÉ, Y., JOLION, J.-M. et MILLE, A.** Sesame: modèle conceptuel de description de documents audiovisuels et assistants d'annotation de contenu. *In: 4èmes Journées d'Etudes et d'Echanges «COmpression et REprésentation des Signaux Audiovisuels»*, CNET — France Télécom, Juin 1998, Lannion, pp. 227–234.
- [Prié *et al.*, 1998b] **PRIÉ, Y., MILLE, A. et PINON, J.-M.** Ai-strata: A user-centered model for content-based description and retrieval of audiovisual sequences. *In: Int. Advanced Multimedia Content Processing Conf.*, Novembre 1998, Osaka, Lecture Notes in Computer Science, vol. 1554, pp. 328–343.
- [Prié *et al.*, 1999a] **PRIÉ, Y., MILLE, A. et PINON, J.-M.** A context-based audiovisual representation model for audiovisual information systems. *In: International and Interdisciplinary Conference on Modeling and using Context*, Septembre 1999, Trento, Lecture Notes in Artificial Intelligence, vol. 1688, pp. 296–309.

- [Prié *et al.*, 1999b] **PRIÉ, Y., MILLE, A. et 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. *In: Ingénierie des Connaissances 1999*, Afia, Juin 1999, Palaiseau, pp. 21–30.
- [Propp, 1970] **PROPP, V.** *Morphologie du conte*, Paris: Gallimard, 1970. 246 p.
- [Purchase, 1998] **PURCHASE, H.** Defining multimedia. *IEEE Multimedia*, 1998, vol. 5, n° 1, pp. 8–15.
- [Rao *et al.*, 1995] **RAO, R., PEDERSEN, J., HEARST, M. et al.** Rich interaction in the digital video library. *Comm. of the ACM*, 1995, vol. 38, pp. 29–39.
- [Rastier, 1987] **RASTIER, F.** *Sémantique interprétative*, Paris: P.U.F., 1987. 284 p.
- [Rastier, 1995a] **RASTIER, F.** Communication ou transmission? *Césure*, 1995, vol. 8, pp. 151–195.
- [Rastier, 1995b] **RASTIER, F.** Le terme: entre ontologie et linguistique. *La banque des mots*, 1995, n° 7, pp. 35–65.
- [Roussey *et al.*, 1998] **ROUSSEY, C., CALABRETTO, S. et PINON, J.-M.** Indexation sémantique de corpus multilingues: Application aux manuscrits anciens. *In: Proceedings of INFOR-SID'98*, May 1998, Montpellier, France, pp. 289–304.
- [Rouvray et Balaban, 1979] **ROUVRAY, D. et BALABAN, A. T.** *Applications of Graph Theory*, San Diego, CA: Academic Press, 1979, chap. Chemical applications of graph theory, pp. 177–221.
- [Rowe *et al.*, 1995] **ROWE, L., BORECZKY, J. et EADS, C.** Indexes for user access to large video databases. *In: Storage and Retrieval of Image and Video Databases II*, 1995, San Jose, pp. 150–161.
- [Sabry-Ismail *et al.*, 1997] **SABRY-ISMAIL, L., ROISIN, C. et LAYAÏDA, N.** Navigation in structured multimedia document using presentation context. *Hypertextes et hypermédias*, 1997, vol. 1, n° 2-3-4, pp. 27–38.
- [Sack et Davis, 1994] **SACK, W. et DAVIS, M.** Idic: Assembling video sequences from story plans and content annotations. *In: Int. Conf. on Multimedia Computing and Systems*, 1994, Boston, pp. 30–36.
- [Salton et Buckley, 1990] **SALTON, G. et BUCKLEY, C.** Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 1990, vol. 41, n° 4, pp. 288–297.
- [Salton et Mac Gill, 1983] **SALTON, G. et MAC GILL, M.** *Introduction to Modern Information Retrieval*, New-York: Mac Graw Hill Book Company, 1983. 448 p.
- [Santini et Jain, 1996] **SANTINI, S. et JAIN, R.** Similarity queries in image databases. *In: Proceedings of CVPR 96, IEEE International Conference on Computer Vision and Pattern Recognition*, Juin 1996, San Francisco, pp. 646–651.

-
- [Sawhney *et al.*, 1997] **SAWHNEY, N., BALCOM, D. et SMITH, I.** Authoring and navigating video in space and time. *IEEE Multimedia*, 1997, vol. 4, n° 4, pp. 30–39.
- [Schreiber *et al.*, 1993] **SCHREIBER, G., WIELINGA, B. et BREUKER, G.** (édité par). *KADS: a Principled Approach to Knowledge-Based System Development*, London: Academic Press, 1993, *Knowledge-Based Systems*, vol. 11. 457 p.
- [Seyrat *et al.*, 1998] **SEYRAT, C., DURAND, G. et FAUDEMAY, P.** Méthode d'indexation multimédia fondée sur les objets visuels. In: *4èmes journées d'études et d'échanges compression et représentation des signaux audiovisuels: Coresa 98*, 1998, Lannion, pp. 213–215.
- [Shapiro, 1991] **SHAPIRO, S.** *Principles of semantic networks*, San Mateo, CA: Morgan Kaufmann, 1991, chap. Cables, paths, and "Subconscious" reasoning in propositional semantic networks, pp. 137–156.
- [Simonnot, 1995] **SIMONNOT, B.** A cooperation model for video document retrieval. In: *Proceedings of SPIE, Storage and Retrieval for Image and Video Databases III*, Février 1995, San Jose, CA, pp. 307–317.
- [Simonnot, 1996] **SIMONNOT, B.** *Modélisation multi-agent d'un système de recherche d'information multimédia à forte composante vidéo*. Thèse de doctorat, Université Henri Poincaré — Nancy I, Janvier 1996. 259 p.
- [Smaïl, 1994] **SMAÏL, M.** *Raisonnement à base de cas pour une recherche évolutive d'information: Prototype Cabri-n. Vers la définition d'un cadre d'acquisition de connaissances*. Thèse de doctorat, Université Henri Poincaré — Nancy I, Octobre 1994. 268 p.
- [Smith et Kanade, 1997] **SMITH, M. et KANADE, T.** Video skimming and characterization through the combination of image and language understanding techniques. In: *IEEE Computer Vision and Pattern Recognition (CVPR)*, Juin 1997, San Jose, Puerto Rico, pp. 775–781.
- [Smith et Chang, 1997] **SMITH, R. et CHANG, S.-F.** Visually searching the web for content. *IEEE Multimedia Magazine*, 1997, vol. 4, n° 3, pp. 12–20.
- [Smoliar et Wilcox, 1997] **SMOLIAR, S. et WILCOX, L.** Indexing the content of multimedia documents. In: *2nd Int. Conf. on Visual Information Systems*, Décembre 1997, San Diego, pp. 53–60.
- [Sorlin, 1992] **SORLIN, P.** *Esthétique de l'audiovisuel*, Paris: Nathan Université, 1992. 222 p.
- [Sowa, 1984] **SOWA, J.-F.** *Conceptual Structures — Information Processing in Mind and Machine*, Reading, MA: Addison-Wesley, 1984. 481 p.
- [Sowa, 1991] **SOWA, J.-F.** (édité par). *Principles of Semantic Networks. Explorations in the Representation of Knowledge*, San Mateo, CA: Morgan Kaufmann, 1991. 582 p.

- [Staff, 1997] **STAFF, C.** Hypercontext: Using context in adaptive hypertext. *In: Proc. of the International and Interdisciplinary Conference on Modeling and Using Context*, Février 1997, Rio de Janeiro, Brazil, pp. 243–255.
- [Stiegler, 1990] **STIEGLER, B.** Machines à lire. *Autrement*, 1990, n° 121, pp. 143–161.
- [Stiegler, 1994] **STIEGLER, B.** Machines à écrire et matières à penser. *Genesis*, ITEM, CNRS, 1994, n° 5, pp. 25–49.
- [Swanberg *et al.*, 1993] **SWANBERG, D., SHU, C.-F. et JAIN, R.** Knowledge guided parsing in video databases. *In: Image and Video Processing Conference; Symposium on Electronic Imaging: Science & Technology*, Février 1993, San Jose, California, SPIE, vol. 1908, pp. 13–24.
- [Takeshita *et al.*, 1995] **TAKESHITA, A., INOUE, T. et TANAKA, K.** Extracting text skim structures for multimedia browsing. *In: Working Notes of IJCAI Workshop on Intelligent Multimedia Information Retrieval*, éd. par Maybury M., Août 1995, Montreal, pp. 46–58.
- [Taniguchi *et al.*, 1997] **TANIGUCHI, Y., AKUTSU, A. et TONOMURA, Y.** Panorama excerpts: Extracting and packing panoramas for video browsing. *In: ACM Multimedia 97*, Novembre 1997, Seattle, pp. 427–436.
- [Taniguchi *et al.*, 1995] **TANIGUCHI, Y., AKUTSU, A., TONOMURA, Y. et al.** Efficient acces interface to real-time incoming video based on automatic indexing. *In: ACM Multimedia'95*, Novembre 1995, San Fransisco, pp. 25–33.
- [Terray, 1996] **TERRAY, P.** *Les outils de lecture des supports dynamiques, la reconstruction de l'objet textuel*. DEA, Université Technologique de Compiègne, 1996. 74 p.
- [Tonomura, 1997] **TONOMURA, Y.** *Handbook of Multimedia Information Management*, Upper Saddle River, New Jersey: Prentice-Hall, 1997, chap. Multimedia Interfaces — Multimedia Content Indication, pp. 189–209.
- [Tonomura *et al.*, 1993] **TONOMURA, Y., AKUTSU, A., OTSUJI, K. et al.** Videomap and videospaceicon: Tools for anatomizing video content. *In: Proceedings of INTERCHI'93 Conference on Human Factors in Computing Systems*, Avril 1993, Amsterdam, The Netherlands, pp. 131–136.
- [Ueda *et al.*, 1993] **UEDA, H., MIYATAKE, T., SUMINO, S. et al.** Automatic structure visualisation for video editing. *In: Proceedings of INTERCHI'93 Conference on Human Factors in Computing Systems*, Avril 1993, Amsterdam, The Netherlands, pp. 137–141.
- [Uehara *et al.*, 1996] **UEHARA, K., OE, M. et MAEHARA, K.** Knowledge representation, concept acquisition and retrieval of video data. *In: Proc. of International Symposium on Cooperative Database Systems for Advanced Applications, CODAS'96*, Novembre 1996, Kyoto, Japan, pp. 527–534.

-
- [Uitdenborgerd et Zobel, 1998] **UITDENBORGERD, A. et ZOBEL, J.** Manipulation of music for melody matching. *In: ACM Multimedia 98*, Septembre 1998, Bristol, pp. 235–240.
- [Ullman, 1976] **ULLMAN, J.** An algorithm for subgraph isomorphism. *Journal of the ACM*, 1976, vol. 23, n° 1, pp. 31–42.
- [Vaillant, 1997] **VAILLANT, P.** *Interaction entre modalités sémiotiques : de l'icône à la langue*. Thèse de sciences cognitives, Université Paris-XI, Septembre 1997. 271 p.
- [van Rijsbergen, 1986] **VAN RIJSBERGEN, C.** A non-classical logic for information retrieval. *The computer journal*, 1986, vol. 29, n° 6, pp. 481–485.
- [Virbel, 1994] **VIRBEL, J.** Annotation dynamique et lecture expérimentale : vers une nouvelle glose? *Littérature*, « *Informatique et littérature* », 1994, vol. 9, pp. 91–105.
- [XML, 1998] **W3C.** Extensible markup language (XML) 1.0, [En ligne], Février 1998, [24.10.99]. Disponible sur Internet : <URL: <http://www.w3.org/TR/REC-xml>>.
- [RDF, 1999] **W3C.** Resource description framework (RDF) model and syntax specification, [En ligne], Février 1999, [24.10.99]. Disponible Internet : <URL: <http://www.w3.org/TR/REC-rdf-syntax/>>.
- [XML-Schema, 1999] **W3C.** Xml schema part 1: Structures / part 2: Datatypes, [En ligne], Sept 1999, [24.10.99]. Disponible sur Internet : <URL: <http://www.w3.org/TR/xmlschema-1/>> et <URL: <http://www.w3.org/TR/xmlschema-2/>>.
- [Wachman, 1996] **WACHMAN, J. S.** *A Video Browser that Learns by Example*. Masters' thesis, Vision and Modeling Group, MIT Media Laboratory, 1996. 86 p.
- [Wang et Adelson, 1994] **WANG, J. et ADELSON, E.** Spatio-temporal segmentation of video data. *In: Proc. of the SPIE: Image and Video Processing II*, Février 1994, San Jose, CA, SPIE, vol. 2182, pp. 120–131.
- [Weiss *et al.*, 1995] **WEISS, R., DUDA, A. et GIFFORD, D.** Composition and search with a video algebra. *IEEE Multimedia*, 1995, vol. 2, n° 1, pp. 12–25.
- [Winston *et al.*, 1987] **WINSTON, M., CHAFFIN, R. et HERMANN, D.** A taxonomy of part-whole relations. *Cognitive Science*, 1987, vol. 11, pp. 417–444.
- [Wold *et al.*, 1996] **WOLD, E., BLUM, T., KEISLAR, D. et al.** Content-based classification, search, and retrieval of audio. *IEEE Multimedia*, 1996, vol. 3, n° 3, pp. 27–36.
- [Wong *et al.*, 1989] **WONG, A., LU, S. et RIOUX, M.** Recognition and shape synthesis of 3-D objects based on attributed graphs. *IEEE Transactions on Pattern Analysis And Machine*, 1989, vol. 11, n° 3, pp. 279–290.

- [Wong, 1992] **WONG, E.** Model matching in robot vision by subgraph isomorphism. *Pattern Recognition*, 1992, vol. 25, n° 3, pp. 287–304.
- [Yeo et Yeung, 1997] **YEO, B. et YEUNG, M.** Retrieving and visualizing video. *Communications of the ACM*, 1997, vol. 40, n° 12, pp. 43–52.
- [Yeung et Yeo, 1996] **YEUNG, M. et YEO, B.** Time-constrained clustering for segmentation of video into story units. In: *Proceedings of 13th International Conference on Pattern Recognition, ICPR'96*, Août 1996, Vienna, Austria, pp. 375–380.
- [Yeung et Yeo, 1997] **YEUNG, M. et YEO, B.** Video visualization for compact presentation and fast browsing of pictorial content. *IEEE Transactions on Circuits and Systems for Video Technology, special issue on Multimedia Technology and Systems*, 1997, vol. 7, n° 5, pp. 771–785.
- [Yeung et al., 1996] **YEUNG, M., YEO, B. et LIU, B.** Extracting story unit from long programs for video browsing and navigation. In: *International Conference on Multimedia Computing and Systems*, Juin 1996, Vienna, Austria, pp. 296–305.
- [Yoshitaka et al., 1996] **YOSHITAKA, A., HOSODA, Y., YOSHIMITSU, M. et al.** Violone: Video retrieval by motion example. *Journal of Visual languages and Computing*, 1996, vol. 7, pp. 423–443.
- [Zettsu et al., 1997] **ZETTSU, K., UEHARA, K., TANAKA, K. et al.** A time-stamped authoring graph for video databases. In: *Databases and Expert Systems Applications*, 1997, Toulouse, Lecture Notes in Computer Science, n° 1308, pp. 192–201.
- [Zhang et al., 1995] **ZHANG, H., LOW, C. S. S. et WU, J.** Video parsing, retrieval and browsing: an integrated and content-based solution. In: *ACM Multimedia '95*, Novembre 1995, San Francisco, pp. 15–24.
- [Zhong et Chang, 1997] **ZHONG, D. et CHANG, S.** Video object model and segmentation for content-based video indexing. In: *IEEE Intern. Conf. on Circuits and Systems (ISCAS'97)*, Vol 2, Juin 1997, Hong Kong, pp. 1492–1495.

Résumé

La numérisation et la création numérique de flux audiovisuels (AV) permettent leur exploitation dans des systèmes d'information audiovisuels. Cela nécessite une modélisation et une instrumentation des contenus des documents AV qui en autorise l'accès direct et soit adaptée aux diverses utilisations possibles : recherche, indexation, navigation, *etc.* Après une revue des propositions actuelles de modélisation de documents AV, nous justifions quelques nécessités en vue de leur représentation. Nous proposons de représenter les documents AV en Strates Interconnectées par les Annotations (Strates-IA), c'est à dire d'« écrire » sur le flux à l'aide de termes (éléments d'annotation). Ceux-ci annotent des morceaux de flux (unités audiovisuelles), sont en relation entre eux, et sont instances d'éléments d'annotations abstraits décrits dans un graphe de relations conceptuelles. L'ensemble du système représente alors un graphe de connaissances global dans lequel nous définissons la notion de contexte comme extrémité de chemin contextuel et pour l'exploitation duquel nous décrivons divers outils contextuels basés sur la notion de graphe potentiel. Les graphes potentiels représentent des visées de description de l'utilisateur liées à ses tâches particulières. Ils s'instancient dans le graphe global (recherche d'isomorphismes de sous-graphes partiels) et nous en proposons un algorithme efficace de multi-propagation. Nous proposons également une modélisation des systèmes documentaires permettant le stockage intelligent de l'expérience sous la forme de cas d'utilisation en vue d'une réutilisation pour l'assistance. Enfin, nous discutons plus généralement les rapports entre documents et connaissances.

Mots-clés : Modélisation de documents et contenus audiovisuels, représentation de connaissances, exploitation contextuelle, isomorphisme de sous-graphes, Strates Interconnectées par les Annotations, expérience d'utilisation.

Abstract

Digitization and digital creation of audiovisual (AV) streams now allow their exploitation in audiovisual information systems. It is therefore needed to model AV documents so that random access and all kinds of usages be allowed: search, indexation, navigation... After a review of current proposals for modelling AV documents, we set up some necessities for AV representation. We propose to represent AV documents with Annotation Interconnected Strata (AI-Strata), which means « writing » on the stream with terms (annotation elements). These terms annotate parts of the stream (audiovisual units), are in relation with each other, and are instances of abstract annotation elements that are described in a conceptual relation graph. In fact, the whole system is a global knowledge graph in which we define a notion of context as contextual path extremity. We then describe some contextual manipulation tools based on the notion of potential graphs. A potential graph represents user's description aim and is directly linked with his task. It instanciates into the global graph (partial subgraph isomorphism search) and we propose an efficient algorithm for it, based on multi-propagation. We also propose a model of documentary information systems allowing intelligent stocking of use experience as use cases that can be reused for user assistance. Finally, we discuss the relations between documents and knowledge.

Keywords : Audiovisual documents and contents modelling, Knowledge representation, Contextual usage, subgraph isomorphism, Annotations-Interconnected Strata, use experience.