HAL open science

Analyse des contenus audiovisuels.

Peter Stockinger

To cite this version:

Peter Stockinger. Analyse des contenus audiovisuels. : Métalangage et modèle de description. Editions Lavoisier - Hermès Science, 2012, Traitement de l'information, Fabrice Papy, 9782746232921. hal-01946336

HAL Id: hal-01946336

https://inalco.hal.science/hal-01946336

Submitted on 5 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Analyse des contenus audiovisuels

Métalangage et modèle de description

Peter Stockinger

(Preprint de l'ouvrage du même titre paru en 2012 aux éditions Lavoisier – Hermès Science, Paris)

Lambach - Paris
2009 - 2011

in memoriam
Monique et *Charles Morazé*

Table des matières

AVANT-PROPOS **8**

PREMIERE PARTIE LE CONTEXTE PRATIQUE, TECHNIQUE ET THEORIQUE **17**

CHAPITRE 1 L'ANALYSE D'UNE RESSOURCE AUDIOVISUELLE **18**

1.1. INTRODUCTION 18
1.2. DES CORPUS FONCTIONNELLEMENT DIFFERENTS 20
1.3. DES MODELES DE DESCRIPTION 26
1.4. DE L'ACTIVITE DE L'ANALYSE DE CORPUS AUDIOVISUELS 27
1.5. DE L'ACTIVITE DE L'INDEXATION 30
1.6. QUELQUES REFLEXIONS AU SUJET DU CADRE THEORIQUE DE REFERENCE 31

CHAPITRE 2 LE STUDIO *ATELIER DE SEMIOTIQUE AUDIOVISUELLE* – UNE BREVE PRESENTATION **38**

2.1. UN ENVIRONNEMENT DE TRAVAIL POUR L'ANALYSE DE CORPUS DE TEXTES AUDIOVISUELS 38
2.2 PRESENTATION SYNTHETIQUE DE L'ATELIER DE DESCRIPTION ASA 42
2.3 QUATRE APPROCHES D'ANALYSE DU TEXTE AUDIOVISUEL 48
2.4 MODELES DE DESCRIPTION ET FORMULAIRES INTERACTIFS DE TRAVAIL 51

CHAPITRE 3 UN EXEMPLE CONCRET D'UN MODELE DE DESCRIPTION DU CONTENU AUDIOVISUEL **55**

3.1. INTRODUCTION 55
3.2. LA SELECTION DU MODELE APPROPRIE DANS LA BIBLIOTHEQUE DE MODELES DE DESCRIPTION DU CONTENU AUDIOVISUEL 56
3.3. LES SEQUENCES D'UN MODELE DE DESCRIPTION DU CONTENU 59
3.4. CHAMP DE DESCRIPTION ET ORGANISATION SEQUENTIELLE D'UN FORMULAIRE D'ANALYSE 61
3.5. LE PLAN DES SCHEMAS DE DEFINITION ET LES PROCEDURES DE DESCRIPTION 64

CHAPITRE 4 MODELE DE DESCRIPTION ET TRAVAIL D'ANALYSE .. **66**

4.1. INTRODUCTION 66
4.2. L'ORGANISATION STRUCTURALE D'UN MODELE DE DESCRIPTION DU CONTENU AUDIOVISUEL 67
4.3. L'ORDRE SYNTAGMATIQUE CANONIQUE D'UN FORMULAIRE DE DESCRIPTION 69
4.4. TYPES D'ANALYSE, TACHES D'ANALYSE, PROCEDURES DE DESCRIPTION ET ACTIVITES DE DESCRIPTION 73
4.5. TACHES PARTICULIERES DE L'ANALYSE DU CONTENU D'UN CORPUS AUDIOVISUEL 76
4.6. REMARQUES CONCLUSIVES 79

DEUXIEME PARTIE TACHES D'ANALYSE D'UN CORPUS AUDIOVISUEL 80

CHAPITRE 5 LA TACHE D'ANALYSE « DESCRIPTION DE L'OBJET DE CONNAISSANCE » 81

5.1. INTRODUCTION 81
5.2. UN EXEMPLE SIMPLE DE DESCRIPTION REFERENTIELLE 82
5.3. STRUCTURE THEMATIQUE, STRUCTURE TOPIQUE ET OBJETS REFERENTIELS ... 84
5.4. UNE BIBLIOTHEQUE DE SEQUENCES POUR LA DESCRIPTION REFERENTIELLE .. 88
5.5. DES ARCHITECTURES FONCTIONNELLES ALTERNATIVES POUR DEFINIR DES SEQUENCES DE DESCRIPTION REFERENTIELLE 90

CHAPITRE 6 LA TACHE D'ANALYSE « CONTEXTUALISATION DU DOMAINE DE CONNAISSANCES » 96

6.1. INTRODUCTION 96
6.2. LA CONTEXTUALISATION PAR LOCALISATION SPATIALE 98
6.3. LOCALISATION ET CONTEXTUALISATION PAR PAYS 99
6.4. LOCALISATION ET CONTEXTUALISATION GEOGRAPHIQUE PHYSIQUE 103
6.5. LA CONTEXTUALISATION PAR LOCALISATION TEMPORELLE 108
6.6. LA CONTEXTUALISATION PAR EPOQUE HISTORIQUE 111
6.7. CONTEXTUALISATION HISTORIQUE ET PERIODISATION 115
6.2. LA CONTEXTUALISATION THEMATIQUE 117

CHAPITRE 7 LA TACHE D'ANALYSE DE LA MISE EN DISCOURS D'UN SUJET 122

7.1 INTRODUCTION 122
7.2. PROCEDURES DE MISE EN DISCOURS 123
7.3 ANATOMIE DE LA DESCRIPTION DE LA MISE EN DISCOURS D'UN SUJET 129
7.4 EXEMPLES ILLUSTRANT L'ANALYSE DE LA MISE EN DISCOURS 131
7.5 DE L'EXPERTISE TEXTUELLE ET DISCURSIVE 135

TROISIEME PARTIE PROCEDURES DE DESCRIPTION 136

CHAPITRE 8 DEFINITION DU DOMAINE DE CONNAISSANCES ET CONFIGURATION DE LA STRUCTURE TOPIQUE 137

8.1. INTRODUCTION 137
8.2. QUELQUES RAPPELS ET PRECISIONS 138
8.3. (RE-)CONFIGURER ET ADAPTER UNE STRUCTURE TOPIQUE EXISTANTE 142
8.4. (RE-)CONFIGURER DES STRUCTURES TOPIQUES PLUS COMPLEXES 145

CHAPITRE 9 LA PROCEDURE DE LA DESCRIPTION LIBRE D'UN CORPUS AUDIOVISUEL 151

9.1. INTRODUCTION 151
9.2. ANATOMIE DE LA PROCEDURE APPELEE « DESCRIPTION LIBRE » 152
9.3. L'ACTIVITE DE DESCRIPTION [DESIGNATION MINIMALE] 155
9.4. L'ACTIVITE DE DESCRIPTION [DESIGNATION CONTEXTUALISEE] 158
9.5. LES ACTIVITES [REALISATION D'UNE PRESENTATION SYNTHETIQUE] ET [DESIGNATION DU REFERENT EN LANGUE D'ORIGINE] 161
9.6. L'ACTIVITE DE DESCRIPTION [DESIGNATION DU REFERENT PAR MOT CLE] .. 162
9.7. VARIANTES PRAGMATIQUES ET ONOMASIOLOGIQUES DE L'ACTIVITE [DESIGNATION MINIMALE] 165

CHAPITRE 10 LA PROCEDURE DE LA DESCRIPTION CONTROLEE D'UN CORPUS AUDIOVISUEL 167

10.1 INTRODUCTION 167
10.2 ANATOMIE DE LA PROCEDURE DE LA DESCRIPTION CONTROLEE 168
10.3 TRAVAILLER AVEC PLUSIEURS MICRO-THESAURUS 171
10.4 SELECTIONNER, CLASSER ET HIERARCHISER A L'AIDE D'UN MICRO-THESAURUS 173
10.5 UNE APPROCHE COMBINEE DE DESCRIPTION CONTROLEE ET LIBRE 175

QUATRIEME PARTIE LE SYSTEME DES RESSOURCES METALINGUISTIQUES ASA 179

CHAPITRE 11 UNE VUE D'ENSEMBLE DES RESSOURCES METALINGUISTIQUES ASA 180

11.1. INTRODUCTION 180
11.2. PRESENTATION GLOBALE DU SYSTEME DES RESSOURCES METALINGUISTIQUES ASA 181
11.3. LE META-LEXIQUE DES TERMES CONCEPTUELS ASA 186
11.4. LE THESAURUS ASA 188
11.5. LES SCHEMAS DE DEFINITION 192
11.6. LES SEQUENCES DE DESCRIPTION 195
11.7. LES RESSOURCES EXTERNES AU SYSTEME ASA 199
11.8 L'ATELIER DE MODELISATION ASA 202

CHAPITRE 12 LE META-LEXIQUE REPRESENTANT L'UNIVERS DU DISCOURS ASA 208

12.1. INTRODUCTION 208
12.2. « TERME CONCEPTUEL » ET « THEME » – QUELQUES EXPLICATIONS 209

12.3. La structure definitionnelle d'une topique 211
12.4. L'univers du discours ASA..... 213
12.5. L'organisation generale du vocabulaire relatif aux objets d'analyse dans l'univers du discours ASA..... 217
12.5. Questions relatives a l'organisation du vocabulaire des termes conceptuels des objets d'analyse ASA..... 221
12.6. Le processus de l'elaboration du vocabulaire des termes conceptuels des objets d'analyse ASA..... 225

CHAPITRE 13 PRESENTATION PLUS DETAILLEE DU VOCABULAIRE CONCEPTUEL [OBJET D'ANALYSE] 228

13.1. Introduction..... 228
13.2. Les deux branches [Objet naturel] et [Objet de valeur]..... 229
13.3. Questions d'organisation du meta-lexique ASA 232
13.4. Comment tenir compte de classifications differentes ?..... 237
13.5. Le domaine conceptuel represente par le terme [Objet materiel fonctionnel]..... 240
13.6. Le domaine conceptuel represente par le terme [Objet social] ... 244
13.7. Le domaine conceptuel represente par le terme [Objet culturel]247
13.8. Domaines taxinomiques faisant partie de la branche [Objet symbolique primaire]..... 249
13.9. Domaines taxinomiques faisant partie de la branche [Objet symbolique secondaire]..... 253
13.10. Les domaines taxinomiques de la branche [Objet perdurant] 257
13.11. Les domaines taxinomiques de la branche [Objet region] 260

CHAPITRE 14 LE META-LEXIQUE DES ACTIVITES D'ANALYSE DE L'OBJET TEXTUEL..... 263

14.1 Introduction..... 263
14.2. Quatre classes d'activites d'analyse textuelle 264
14.3. La classe des activites [Procedure d'analyse structurale de l'objet textuel] 267
14.4. La classe des activites [Procedure d'analyse de l'objet textuel a l'aide du thesaurus ASA] 271
14.5. La classe des activites [Procedure d'analyse de l'objet textuel a l'aide d'une reference ASA externe]..... 272
14.6. La classe des activites [Procedure d'analyse pragmatique de l'objet textuel] 275
14.7. Activite d'analyse et schemas d'indexation..... 277
14.8. La bibliotheque des schemas d'indexation..... 279

CHAPITRE 15 LE THESAURUS ASA 285

15.1. Introduction..... 285
15.2. Presentation generale du thesaurus ASA 286

15.3. FACETTES ET LISTES D'EXPRESSIONS NORMALISEES 289

CHAPITRE 16 LES BRIQUES CONFIGURATIONNELLES DES MODELES DE DESCRIPTION 293

16.1. INTRODUCTION..... 293
16.2. ANALYSE D'UN TEXTE AUDIOVISUEL ET MODELES DE DESCRIPTION..... 294
16.3. LA BIBLIOTHEQUE DE SEQUENCES COMPOSANT LE MODELE DE DESCRIPTION THEMATIQUE 296
16.4. DEFINITION ET INSERTION D'UNE SEQUENCE DANS UN MODELE DE DESCRIPTION..... 301
16.5. PRESENTATION SYNTHETIQUE D'UNE BIBLIOTHEQUE DE SCHEMAS DE DEFINITION 305

CONCLUSION ET PERSPECTIVES 309

BIBLIOGRAPHIE..... 313

INDEX 320

GLOSSAIRE DES ACRONYMES 324

GLOSSAIRE DES TERMES SPECIALISES 334

Avant-propos

Un des grands enjeux pour les bibliothèques et archives numériques d'aujourd'hui est celui de permettre aux usagers une *appropriation active* de leurs ressources textuelles et, plus particulièrement, audiovisuelles. *Appropriation active* veut bien dire, adaptation des données audiovisuelles aux besoins et intérêts spécifiques d'un utilisateur ou d'un groupe d'utilisateurs, un groupe d'utilisateurs pouvant être tout aussi bien une petite « communauté » fermée comme, par exemple, les participants à un projet de recherche ou les membres d'une équipe enseignante, que des réseaux « communautaires » virtuels et agissant tout autour du globe. Prendre en considération ce processus de l'appropriation active passe par :

1) une prise en compte effective et systématique de l'*organisation structurale interne* du texte audiovisuel, c'est-à-dire plus particulièrement par le recours à la sémiotique du texte ou du discours [STO 03] ;

2) la définition et le développement de modèles et d'outils de traitement physique et/ou intellectuel (sous forme, par exemple, d'analyses et d'interprétations) du texte audiovisuel par toute personne ;

3) la possibilité offerte à toute personne de devenir *auteur* au sens de quelqu'un qui se propose de produire et de publier une ou plusieurs nouvelles versions d'un texte ou d'un corpus de textes audiovisuels *déjà existants* – nouvelles versions plus adaptées à un contexte d'usage spécifique ou encore, d'une manière plus large, au profil (culturel) et aux attentes d'un public donné.

Une question particulièrement centrale dans ce contexte est celle du *métalangage de description**[1] qui est nécessaire pour permettre à toute personne de procéder à des *analyses** de toute sorte d'un texte ou d'un corpus de textes audiovisuels

[1] Les expressions suivies par un astérisque « * » sont définies et décrites dans le glossaire des termes spécialisés se trouvant à la fin du livre.

documentant l'*univers du discours** d'une archive audiovisuelle. Ce livre lui est consacré.

La conception, l'élaboration et l'expérimentation concrète d'un métalangage de description pour l'analyse de corpus audiovisuels ainsi que le développement d'un environnement informatique pour tester ce métalangage a été rendu possible grâce à un financement de l'Agence Nationale de la Recherche (ANR)[2] sous forme d'un projet de recherche et de développement s'intitulant « Atelier de Sémiotique Audiovisuelle pour l'analyse de corpus audiovisuels en Sciences Humaines et Sociales » (acronyme du projet ASA-SHS)[3].

L'environnement qui porte le nom Studio ASA est présenté en détail dans l'ouvrage collectif *Les archives audiovisuelles* [STO 11a]. Dans un autre ouvrage collectif intitulé *Nouveaux usages des archives audiovisuelles* [STO 11b] sont présentés et discutés des exemples variés montrant l'intérêt potentiel de cet environnement pour des projets dans le cadre de la constitution et l'exploitation d'archives audiovisuelles ainsi que, notamment, pour l'*analyse de corpus audiovisuels* spécifiques.

Le contenu de ce livre est divisé en *quatre grandes parties*. La *première partie*, composée des quatre premiers chapitres, trace le contexte général de notre travail consacré à la définition et à l'élaboration d'un métalangage de description en vue de l'analyse d'un texte ou d'un corpus de textes audiovisuels.

Pour commencer, nous problématiserons ce que nous entendons par *analyse* d'un texte ou d'un corpus de textes audiovisuels (chapitre 1). Nous insisterons ici plus particulièrement sur ce qu'on appelle dans la littérature « anglo-saxonne » et en référence aux travaux de Michel Foucault le *linguistic turn* dans la recherche consacrée aux archives et/ou bibliothèques numériques (voir par exemple [CRA 08]). Au centre de cette approche on trouve la question de l'explicitation de l'univers du discours d'une archive ou d'une bibliothèque, c'est-à-dire du *sens* ou encore du *contenu* des ressources (écrites, audiovisuelles, etc.) qui composent une archive ou une bibliothèque. Et cette façon de problématiser à la fois l'*intérêt* (potentiel) et l'*appropriation active* d'un texte et/ou d'une archive de textes (audiovisuels) par ses usagers correspond parfaitement au point de vue développé dans ce livre.

---

[2] Voir http://www.agence-nationale-recherche.fr/.

[3] Le projet ASA-SHS (« Atelier de Sémiotique Audiovisuelle pour l'analyse de corpus audiovisuels en Sciences Humaines et Sociales » (acronyme du projet ASA-SHS ») a eu une durée de vie de trois ans. Il a démarré début 2009, dans le cadre du Programme Blanc de l'ANR (Agence Nationale de la Recherche), et s'est terminé fin décembre 2011.

Dans le chapitre suivant (chapitre 2), nous présenterons d'une manière très synthétique l'environnement de travail appelé Studio ASA que notre équipe de recherche en Sémiotique Cognitive et nouveaux Médias (ESCoM) à la Fondation Maison des Sciences de l'Homme (FMSH) à Paris a pu mettre en place entre 2000 et 2016 pour tester et valider l'approche présentée dans ce livre. Comme nous le montrerons, le métalangage de description propre à l'univers du discours d'une archive, s'articule sous forme d'une *bibliothèque* de *modèles de description** à laquelle l'analyste (la personne ou le groupe de personnes procédant à l'analyse d'un texte audiovisuel) a accès via des interfaces spécialisées (des *formulaires interactifs de saisie* d'informations ou de données) qui font partie du Studio ASA. Ceci dit, le métalangage de description en tant que tel reste bien entendu indépendant par rapport à un environnement concret de travail. Autrement dit, il peut être rendu opérationnel dans des environnements les plus variés.

Après la présentation synthétique du Studio ASA, nous introduirons dans le chapitre suivant (chapitre 3) un exemple concret de l'analyse d'un texte audiovisuel à l'aide d'une approche qui se base sur une *bibliothèque de modèles de description** représentant l'univers du discours* d'une archive. L'exemple concret développé dans ce troisième chapitre nous servira en quelque sorte de fil rouge pour développer et expliciter notre approche pendant tout le livre.

Ainsi une fois l'exemple présenté d'une manière aussi concrète que possible, nous en tirerons, dans le chapitre 4, un ensemble de conséquences plus générales qui seront discutées bien plus en détail dans les trois parties restantes du livre : typologie de pratiques d'analyse ; modèles de description appropriés à tel ou tel type d'analyse ; organisation hiérarchique et syntagmatique interne d'un modèle de description ; principales « briques » composant un modèle de description.

La deuxième partie est composée de trois chapitres (c'est-à-dire des chapitres 5, 6 et 7) et est consacrée à une discussion détaillée de trois tâches d'analyse qui nous semblent être très récurrentes dans une perspective de description – indexation du *contenu* d'une archive audiovisuelle. Il s'agit des tâches suivantes :

- *analyse du domaine d'expertise* (c'est-à-dire de l'objet référentiel, du *référent* d'un texte ou un corpus de textes audiovisuels) ;
- *analyse de la contextualisation* (spatiale, temporelle, sociale, etc.) *du domaine d'expertise* dont est question dans un texte ;
- et enfin, *analyse des stratégies de la mise en discours du référent* dans le texte.

Avec deux autres tâches – celle de l'*analyse de l'expression verbale et audiovisuelle* du contenu d'un texte et celle du *commentaire*, du *métadiscours de l'analyste* portant sur son analyse – que nous ne développons pas davantage dans ce livre, les trois tâches citées constituent un ensemble canonique de tâches définissant

un *type spécifique d'analyse* consacré à la description, indexation et interprétation du *contenu* d'un texte ou d'un corpus de textes audiovisuels.

Insistons sur le fait que l'analyse du contenu d'un texte ou d'un corpus de textes audiovisuels *ne forme qu'un type spécifique* et bien particulier d'analyses (voir chapitre 4). D'autres types d'analyse, qui sont développés et rendus opérationnels dans le Studio ASA mais dont nous ne parlons guère dans ce livre (consacré, répétons-le, surtout à la question de l'analyse du contenu d'une ressource audiovisuelle) sont, par exemple :

– l'*analyse paratextuelle** (type d'analyse qui s'intéresse – à la manière du standard Dublin Core[4] – à la description de l'identité formelle d'un texte : auteur, titre, droits, genre, etc.) ;

– l'*analyse audiovisuelle** *stricto sensu* (analyse qui s'intéresse avant tout à la description des plans visuels et sonores d'un texte audiovisuel) ;

– l'*analyse pragmatique** (analyse qui s'intéresse à l'ancrage contextuel – social, éducatif, linguistique, etc. – du texte analysé) ;

– la méta-description* forme particulière d'analyse qui sert à l'explicitation des objectifs et enjeux de l'analyse d'un objet audiovisuel.

Le chapitre 5 est réservé à une discussion approfondie des différents aspects théoriques, méthodologiques et techniques de l'analyse d'un domaine d'expertise thématisé dans un texte audiovisuel. Nous rencontrerons ici la notion théorique centrale – empruntée à la théorie sémiotique de Greimas – de la *configuration thématique** qui nous permet de décrire le sujet* d'un texte ou d'un corpus de textes audiovisuels. Le sujet du texte est constitué à la fois *de ce dont parle le texte* et du *comment il en parle* et *comment il exprime ce dont il parle*. Plus précisément parlant, la structure thématique se compose, entre autres :

– de la *structure topique** *stricto sensu* (c'est-à-dire de la structure définissant l'objet référentiel thématisé dans un texte) ;

– de la structure de la mise en discours* (c'est-à-dire de la structure définissant l' « approche » de l'objet référentiel thématisé dans le texte par son auteur) ;

– ainsi que de la structure de l'expression verbale et audiovisuelle* (c'est-à-dire de la structure définissant le langage verbal, visuel ou audiovisuel à l'aide duquel un contenu est exprimé).

Dans le chapitre 6, nous discuterons l'analyse du contexte de l'objet ou du domaine thématisé dans un texte audiovisuel. Il s'agit notamment de la *contextualisation spatiale* (géographique, géopolitique) et *temporelle* (chronologique, historique), complétée par une troisième forme de contextualisation que nous appelons *contextualisation thématique*. Cette troisième forme identifie le

[4] Voir http://dublincore.org/.

contexte institutionnel, social ou culturel, épistémique ou mental qui est pertinent pour un objet thématisé dans un texte audiovisuel.

Les procédures de la mise en discours d'un objet thématisé et contextualisé seront discutées dans le chapitre 7. Tout en étant conscient qu'il s'agit ici d'un domaine de recherche complexe et diversifié, nous nous concentrerons sur quelques aspects qui nous paraissent importants pour mieux cerner, par exemple :

– l'intérêt que présente pour son auteur l'objet référentiel thématisé ;

– la valeur ajoutée de la thématisation de l'objet référentiel dans un texte donnée ;

– le caractère spécifique (original, « nouveau », etc.) du traitement de l'objet référentiel dans le texte ;

– les objectifs que poursuit l'auteur en thématisant tel ou tel objet référentiel, et ainsi de suite.

Comme nous le verrons dans la *deuxième partie* de ce livre, la réalisation d'une tâche d'analyse spécifique (analyse de l'objet ou du domaine, analyse du contexte de l'objet, analyse de la mise en discours de l'objet, etc.) s'appuie sur des *séquences de description** qui composent un modèle de description donné. Ainsi distinguons-nous, par exemple :

– des collections de séquences réservées à l'analyse du domaine ou de l'objet référentiel thématisé dans un texte ou un corpus de textes audiovisuels ;

– des collections de séquences réservées à l'analyse du contexte d'un objet référentiel thématisé ;

– des collections de séquences réservées à l'analyse de la mise en discours de l'objet référentiel thématisé.

Chaque séquence, à son tour, est constituée d'au moins *une paire canonique de schémas de définition* – l'une des deux définit l'objet d'analyse, l'autre la procédure employée pour analyser l'objet en question.

Nous distinguons, par ailleurs, entre *deux procédures de description** de base qui peuvent être employées par chaque tâche d'analyse. Leur présentation et discussion est l'objet principal de la *troisième partie* de ce livre composée de trois chapitres (c'est-à-dire des chapitres 8, 9 et 10).

Dans le chapitre 9, nous traiterons de la première des deux procédures. Il s'agit de la procédure de description dite libre*. L'analyste est – comme l'expression le suggère déjà – libre de produire telle ou telle information, telle ou telle donnée pour renseigner ou encore indexer un texte audiovisuel afin de rendre compte du domaine traité et de la façon dont le domaine est abordé et exprimé par l'auteur.

La deuxième procédure est présentée et discutée dans le chapitre 10. Il s'agit de la procédure que nous appelons *contrôlée**. Celle-ci s'effectue à l'aide d'un *micro-thesaurus** approprié à l'aide duquel l'analyste renseigne l'objet référentiel thématisé dans un texte audiovisuel, le contexte spatial, temporel ou thématique, telle ou telle stratégie de mise en discours ou encore telle ou telle façon d'exprimer le sujet développé dans le texte soumis à l'analyse.

Le chapitre 8 par lequel débute cette troisième partie, est réservé, quant à lui, à des questions plus pointues qui concernent la définition de la structure topique*, c'est-à-dire de la partie de la *configuration thématique** qui est réservée à l'analyse du domaine d'expertise – de l'*objet référentiel** – thématisé et traité dans un texte ou un corpus de textes audiovisuels.

Enfin, la *quatrième partie* de ce livre – la plus importante – réunit 6 chapitres (c'est-à-dire les chapitres 11, 12, 13, 14, 15 et 16) qui présentent d'une manière aussi systématique et détaillée que possible le système métalinguistique et ses différents composants dont nous avons besoin pour spécifier le métalangage de description* (sous forme d'une *bibliothèque* de modèles de description) pour décrire les ressources d'une archive audiovisuelle.

Dans le chapitre 11 sera proposée une vision synthétique des différents composants du dit système métalinguistique ainsi que de leurs fonctions dans la spécification d'un modèle de description ou d'une *bibliothèque* de modèles de description.

Les chapitres 12, 13 et 14 seront consacrés à une présentation détaillée du « cœur » du système des ressources métalinguistiques qui est composé par deux méta-lexiques (ontologies) de *termes conceptuels** (de *concepts*) :

– le premier méta-lexique identifie et désigne tous les *objets d'analyse* de l'univers du discours ASA : *objets référentiels*, *objets de localisation* (spatiale, temporelle, etc.), *objets discursifs* ou encore *objets d'expression* textuelles, linguistiques, visuelles, etc. ;

– le deuxième méta-lexique identifie toutes les *activités d'analyse* pour décrire, indexer, annoter, illustrer, etc., les objets d'analyse de l'univers du discours ASA.

Dans le chapitre 15, nous présenterons une autre ressource centrale du système métalinguistique ASA celle du *thesaurus** dont nous avons besoin pour mettre en œuvre la procédure de la description contrôlée. Le thesaurus, lui, est constitué :

– par une *bibliothèque* de *facettes* (c'est-à-dire de dimensions sémantiques) et,

– par une *bibliothèque* d'*expressions normalisées* (de « descripteurs ») qui sont classées selon les facettes dont dispose le thesaurus.

Le dernier chapitre de ce livre – le chapitre 16 – sera, enfin, consacré à une discussion plus systématique des *briques** génériques qui forment un modèle de description ou une bibliothèque de modèles de description. Il s'agit de deux types de briques génériques :

– les *séquences de description** (référentielle, contextuelle, discursive, de l'expression verbale, de l'expression audiovisuelle, etc.) et,

– les *schémas** définissant chaque séquences (schémas définissant les *objets à analyser* et schémas définissant les *activités* à réaliser *pour analyser un objet*).

Séquences et schémas forment des configurations qui sélectionnent et positionnent un ou plusieurs *termes conceptuels** provenant des deux méta-lexiques cités ci-dessus. Ils forment, à côté du thesaurus et des deux méta-lexiques des termes conceptuels, une *troisième classe* importante des ressources métalinguistiques ASA.

Ce livre constitue l'aboutissement (provisoire) d'une partie importante des recherches qui nous ont préoccupées depuis longtemps déjà. En effet, notre intérêt portant sur le métalangage de description de textes (audiovisuels) remontent aux alentours des années 1983 et 1984 lorsque A.J. Greimas, avec qui nous avons eu la chance de travailler pendant plus que 15 ans, nous a amené à nous intéresser aux questions théoriques et techniques soulevées par les *réseaux sémantiques* ou *conceptuels*, alors très en vogue dans les recherches en intelligence artificielle et consacrées aux systèmes de représentations des connaissances.

Dans le cadre de notre travail de chercheur au CNRS entre 1985 et 1991 et, d'une très stimulante coopération de plusieurs années avec IBM France à la fin des années 80, nous avons pu travailler d'une manière concrète sur les questions relatives à l'extraction et la représentation du contenu dans des bases de données lexicales et textuelles à l'aide de modèles d'inspiration sémiotique et structurale (« structurale » au sens de la linguistique structurale). Dès 1995/96, et grâce à toute une série de projets de recherche européens, nous avons pu élargir la problématique de la définition d'un métalangage de description sous forme de modèles de description[5] aux bibliothèques numériques et enfin, dès 2000/2001, aux corpus et archives audiovisuelles numériques.

Si les prévisions de CISCO[6] s'avèrent exactes, à savoir celles qui nous pronostiquent qu'à l'horizon de 2014, 91 % des données numériques circulant sur Internet seront de nature audiovisuelle, alors on peut mesurer l'énorme enjeu que

---

[5] Nous parlions alors de *scripts sémiotiques* ou encore de *scripts thématiques* appliqués d'une part à l'analyse du contenu de bases de données textuelles et d'autre part à la conception et réalisation de produits et services d'information en ligne (voir [STO 94]).

[6] Voir http://newsroom.cisco.com/dlls/2010/ts_111510.html.

présente, pour les années à venir, la problématique des *modèles conceptuels* (et donc d'un *système de ressources métalinguistiques** permettant de les élaborer) pour traiter (c'est-à-dire analyser, adapter, publier, faire circuler, conserver) des corpus audiovisuels des plus variés. Disposer d'un tel système de ressources métalinguistiques constitue, sans aucun doute, une condition *sine qua non* pour une gestion et une exploitation intelligente de ces données !

Nous sentons une certaine satisfaction personnelle pour le travail accompli depuis maintenant presque 25 ans tout en étant tout aussi conscients de ses limites, de ses zones d'ombre et, enfin, de son caractère très provisoire. Face à la complexité empirique des textes (audiovisuels) et de la diversité des attentes (très rarement explicites) de leurs publics, nos certitudes d'antan ont fait place à une posture bien plus modeste. Tout métalangage est une construction intrinsèquement limitée et faillible, élaboré sur des bases parfois peu stables et profondément enracinée dans les visions culturelles et les croyances (scientifiques ou autres) de son concepteur. Mais il s'agit ici d'un « outil » qui est indispensable pour tout « travail » autour du texte audiovisuel (voir aussi [STO 11a, STO 11b]).

En guise de conclusion, nous voudrions encore exprimer nos remerciements à toutes celles et tous ceux qui ont joint pendant un certain temps notre équipe de recherche en sémiotique cognitive et nouveaux médias à la Fondation Maison des Sciences de l'Homme à Paris pour contribuer à la mise en place du programme des Archives Audiovisuelles de la Recherche (voir [STO 03c, STO 10a]) et à l'avancement de nos recherches consacrées à l'analyse et à la publication de corpus audiovisuels. Une pensée toute particulière à Elisabeth de Pablo qui nous a accompagnés dans tous nos projets de recherche et qui nous a tant aidés dans notre travail scientifique.

Plus nous avançons dans l'âge, plus nous avons l'impression d'acquérir – pour citer Claude Lévi-Strauss – la *bonne distance* qui nous permet de mieux saisir la valeur des personnes qui ont pu croiser notre chemin de chercheur. Parmi les personnes qui nous restent les plus chères et dont notre souvenir ne s'effacera jamais, nous comptons en tout premier lieu, Monique et Charles Morazé envers qui notre dette n'est pas seulement intellectuelle – bien loin de là ! Ce livre leurs est dédié en mémoire fidèle et « filial ».

Peter STOCKINGER
(Lambach – Paris, 2011)

Première partie

Le contexte pratique, technique et théorique

Chapitre 1
L'analyse d'une ressource audiovisuelle

1.1. Introduction

L'objectif de ce livre est de présenter une approche opérationnelle basée sur la *sémiotique** du *texte** audiovisuel [STO 03] pour l'analyse, c'est-à-dire la description, l'interprétation et l'indexation de corpus audiovisuels numériques.

La notion centrale, utilisée pour cette approche, est celle du *modèle de description** d'un objet audiovisuel, telle qu'une vidéo, qui se base sur un ensemble de critères servant à la sémiotique pour traiter l'objet *texte** et que nous présenterons plus en détail dans le troisième chapitre de ce livre. Il s'agit principalement des critères suivants :

- le critère du texte en tant qu'entité *compositionnelle* (un texte peut être, en principe, décomposé en unités textuelles « plus petites » et il fait partie, lui-même d'un environnement textuel, d'un – métaphoriquement parlant – *text-* ou *mediascape*) ;

- le critère du texte en tant qu'entité *structurale* pourvue d'un ensemble de constituants caractéristiques (constituants tels que le constituant thématique, le constituant narratif, le constituant rhétorique et discursif, le constituant de

l'expression multimodale d'un contenu ou encore le constituant de l'organisation formelle et physique du contenu dans le texte) ;

– et enfin le critère du texte en tant qu'entité *historique* (le texte en tant que genre) et *évolutif* (le texte en tant que *produit* en principe toujours modifiable d'un savoir-faire).

L'hypothèse de ce livre est que tout travail d'analyse d'un corpus textuel en général et audiovisuel en particulier – peu importe son niveau de spécialisation – repose sur des représentations, des « visions », voire des théories : 1) de l'*objet texte* et 2) de l'*activité* de l'*analyse* du texte*.

Un modèle de description n'est ainsi, en fin de compte, rien d'autre que la partie explicitée, formalisée (au sens large du terme) d'une théorie ou d'une vision qui guide le travail de l'analyse d'un corpus textuel (audiovisuel, dans notre cas).

L'écart de « satisfaction » pouvant exister entre le modèle et la théorie, ou la vision sous-jacente au travail d'analyse, pourra être expliqué soit par un implicite plus ou moins important qui guide l'analyste dans son travail et que le modèle ne peut pas ou n'est pas capable de prendre en considération, soit encore par les simplifications obligatoires qu'il faut opérer par rapport à un référentiel théorique pour développer une approche explicite et opérationnelle de l'analyse de corpus textuels, c'est-à-dire audiovisuels.

Toujours est-il que le travail de définition, de développement, de validation et du suivi de modèles de description de corpus textuels en général et audiovisuels en particulier, représente tout un *métier*, c'est-à-dire tout un ensemble de compétences et de savoir-faire spécialisés faisant appel à une culture et des connaissances variées qui couvrent non seulement des domaines pratiques et technologiques tels que les technologies de l'information et des connaissances, les sciences appliquées de la documentation, de l'archivistique, de la bibliothéconomie ou encore de la gestion de patrimoines culturels lato sensu, mais aussi – et cela, à notre avis, d'une manière cruciale – tout un ensemble des disciplines en sciences humaines tel que les sciences du texte (dont, notamment la *sémiotique**), les sciences du langage ou encore cet ensemble hétéroclite émergeant d'approches et de problématiques que l'on classe sous le sigle général des « sciences de la culture » (*Kulturwissenschaften*, en allemand).

Le métier en question est celui du *modélisateur** parfois aussi appelé *concept-designer**, voire *cogniticien* ou *ingénieur des connaissances*, la terminologie est en effet encore très floue et peu stabilisée. C'est néanmoins un *rôle central* dans le *processus de travail** [SDL 11] définissant la constitution, l'analyse et la publication/diffusion de patrimoines de connaissances qui sont véhiculés par des

corpus audiovisuels. Le modélisateur prépare, développe et gère toutes les ressources métalinguistiques nécessaires aux autres acteurs intervenant dans ce processus de travail afin que ces derniers puissent réaliser leur travail.

1.2. Des corpus fonctionnellement différents

Dans le cadre du processus de travail de numérisation de patrimoines de connaissances, nous pouvons distinguer une série de classes de modèles (c'est-à-dire de ressources métalinguistiques) nécessaires pour accomplir les différentes activités composant le dit processus.

Comme décrit dans [SDL 11], le processus de constitution d'un patrimoine de connaissances sous forme, par exemple, d'une archive numérique, se déroule en plusieurs étapes canoniques qui sont notamment :

1) l'étape de la *préparation d'un terrain de collecte de données* documentant un patrimoine de connaissances ;
2) l'étape de la *conduite, de la réalisation* à proprement parler *du terrain*[7] *;*
3) l'étape du *traitement technique et auctorial* des données collectées (incluant, entre autres, le *derushage* de données audiovisuelles, le montage et la postproduction de données audiovisuelles collectées) ;
4) l'étape de *l'analyse* (de la description, de l'indexation mais aussi de l'adaptation pragmatique) des données collectées et documentant un terrain ;
5) l'étape de la *publication* et de la *diffusion* des données collectées et/ou analysées et, enfin, l'étape de la *conservation* des données collectées et/ou analysées et/ou publiées.

Or, chaque étape dans ce processus de numérisation d'un patrimoine de connaissances a obligatoirement à faire avec un certain type *fonctionnellement spécialisé* de *corpus** (dans notre cas : de corpus audiovisuels) :

---

[7] Dans le cadre du Programme de recherche AAR (Archives Audiovisuelles de la Recherche) que nous avons mis en place en 2001/2002 à la Fondation Maison des Sciences de l'Homme (FMSH) à Paris, nous utilisons d'une manière délibérée le terme *terrain* dans un sens très large. Ainsi, incluons-nous sous ce terme la collecte de données documentant d'une part un *« événement » ou une « manifestation » ponctuelle* et *circonscrite* (manifestation ponctuelle et circonscrite telle que séminaire, cours, colloque, réunion de travail, etc. ponctuant la vie de la recherche et de l'enseignement supérieur) et d'autre part la *recherche de terrain* (*field research*, en anglais) au sens habituel du terme en sciences humaines et sociales (comme, par exemple, la mission ethnologique visant la collecte – réalisée normalement en plusieurs temps et à durée très variable – de données de toute sorte relatives à l'organisation sociale d'une communauté, de sa vie sociale et culturelle).

1) L'étape de la préparation d'un terrain de collecte de données audiovisuelles ne peut se concevoir qu'en référence à un *corpus* déjà existant ou encore à constituer *de connaissances et de sources d'information* nécessaires à la bonne conduite du terrain (connaissances et sources d'information pouvant recouvrir aussi bien des références bibliographiques, des ressources en ligne, des informations personnelles, des « bonnes pratiques », des exemples de projets similaires en cours ou déjà réalisés, des annuaires, etc.)[8].

2) L'étape de collecte des données débouche sur la constitution ou la mise à jour/l'enrichissement d'un *corpus de terrain** déjà existant. Le corpus de terrain n'est pas seulement constitué par les données produites pendant le terrain. Prenons le cas de l'enregistrement d'un *terrain* aussi bien circonscrit que celui d'un séminaire de recherche dont les séances à filmer s'échelonnent sur une année universitaire. Le corpus des données documentant le terrain *séminaire de recherche* n'est pas (obligatoirement) réduit aux enregistrements audiovisuels des différentes séances. Il recouvre toutes les données supposées pertinentes soit pour rendre compte de ce terrain (c'est-à-dire pour en constituer une archive de connaissances à proprement parler), soit pour rendre possible une analyse de qualité de tel ou tel aspect du séminaire filmé, soit pour disposer d'un fonds documentaire en vue d'une ou de plusieurs publications (en ligne) du séminaire, soit encore pour le transformer (tel quel ou après un processus de sélection de documents à préserver « absolument ») en un *corpus patrimonial* (voir *infra*) documentant, par exemple, l'histoire d'une discipline ou d'une institution de recherche[9].

---

[8] Ainsi, par exemple, avons-nous mis en place dans le cadre du Programme AAR toute une procédure de collecte d'informations, de connaissances et autres données utiles pour mener à bien un projet d'entretien avec un chercheur – projet qui peut se concrétiser sous forme d'un seul entretien ou d'une série d'entretiens, de séries d'entretiens *in situ* (c'est-à-dire sur le lieu de travail du chercheur), d'entretiens composés de différents participants discutant avec le chercheur invité, etc. Cette procédure se base sur un modèle simple définissant les critères à prendre en considération pour préparer un entretien avec un chercheur – critères tels que la connaissance du chercheur et de son parcours personnel, la connaissance de ses principaux domaines de recherche, la connaissance du contexte académique et scientifique, etc. Mais ce modèle exige également, de la part de la personne en charge de la préparation d'un entretien, de collecter des données facilitant d'une part la conduite de l'entretien et d'autre part l'exploitation des données documentant l'entretien en vue de son traitement technique et auctorial, de son analyse et de sa publication. Ainsi, la personne en charge d'un entretien doit, par exemple, identifier avec le chercheur invité, en amont de l'entretien, les principaux sujets qui seront développés tout en respectant le scénario-type d'un entretien dans le cadre du programme AAR.

[9] Dans le cadre du Programme AAR, nous avons en effet établi toute une série de procédures simples de constitution de tels corpus de terrain lors de l'enregistrement de ce que nous avons appelé *événements scientifiques* (entretien avec un chercheur, colloque ou encore séminaire de recherche).

3) L'étape de traitement technique et auctorial s'appuie sur une sélection de données collectées faisant partie d'un corpus de terrain ou, le cas échéant de plusieurs corpora de terrain ou encore de données provenant de différentes périodes dans la vie d'un corpus de terrain (un corpus de terrain peut être mis à jour, enrichi, etc.). En tout cas, un *corpus de traitement** est composé de données sélectionnées en vue, par exemple, d'être montées ensemble pour former une *nouvelle réalisation audiovisuelle* correspondant à une intention auctoriale de publication (c'est-à-dire à un scénario définissant une telle réalisation). Ainsi, une intention de publier les enregistrements d'un séminaire de recherche peut viser la diffusion d'une certaine problématique traitée pendant le dit séminaire. Dans ce cas, pas « tout » le séminaire fait l'objet d'une intention de publication mais seulement les parties de celui-ci dans lesquelles la problématique choisie est traitée. Mais même au cas où il y a une décision de publier « tout » le séminaire, les enregistrements réalisés pendant la phase du terrain doivent subir un traitement technique (encodage, vérification de la qualité visuelle et sonore, élimination de passages inutilisables, etc.) avant d'être mise à la disposition d'une publication du séminaire et de ses différentes séances sous forme par exemple d'un site web. Donc, peu importe si l'étape du traitement se réduit à une « simple » activité technique ou si elle recouvre aussi une activité auctoriale à proprement parler, à chaque fois se pose la question de la définition et de la constitution du corpus de traitement. Ajoutons aussi que dans le contexte des archives numériques de connaissances, un *corpus de traitement* peut se nourrir non seulement de données provenant d'un ou de plusieurs corpus de terrain mais également de données déjà publiées et « réinjectées », réutilisées dans le cadre d'un nouveau traitement technique et surtout auctorial. Concrètement parlant, une problématique scientifique traitée dans un séminaire et faisant l'objet d'un scénario de montage pour une publication en ligne peut parfaitement bien intégrer, à côté des données originales (provenant, par exemple, d'un nouveau corpus de terrain) des parties de contributions déjà publiées.

4) L'étape d'analyse de données collectées est celle qui nous intéresse le plus ici et à laquelle est consacré ce livre. Retenons pour le moment que l'analyse d'une donnée textuelle (audiovisuelle, dans notre cas) ne se réduit ni à une « simple » indexation libre, ni à une indexation contrôlée selon tel ou tel standard, tel ou tel langage documentaire. L'analyse recouvre *toutes* les activités intellectuelles – de l'indexation documentaire jusqu'à l'interprétation la plus personnelle en passant par les différentes formes de l'expertise professionnelle de l'information – qui « utilisent », « exploitent » l'objet *texte** pour satisfaire un besoin (un désir, voire une simple curiosité) de connaissances. Or, un tel besoin ou désir peut reposer sur des motifs très variables et se réaliser dans des contextes sociaux et culturels les plus différents. Toujours est-il que l'analyse en tant qu'*activité de satisfaction d'un besoin ou désir de connaissances* ne peut se réaliser avec succès que si elle dispose de l'objet approprié, de sa *matière première* qui est le *texte** ou, plutôt, le corpus de

textes. Dans le cadre du processus de la constitution et de la diffusion d'un patrimoine de connaissances, le *corpus d'analyse**, c'est-à-dire dans notre cas le corpus des données audiovisuelles soumis à l'analyse, n'est pas obligatoirement – bien loin de là – coextensif avec un corpus de terrain. En effet, tout dépend de l'objectif de l'*analyse** et, d'une manière plus générale, de la *politique d'analyse** (par exemple, dans le cadre de l'exploitation des contenus d'une *archive** de connaissances). Si l'analyse est conçue comme une activité de description et de classification de données préalablement collectées et documentant un terrain particulier en vue, par exemple, de leur publication en ligne, corpus de terrain et corpus d'analyse se rapprochent, deviennent similaires sans se confondre pour autant. Si l'analyse est conçue d'une manière indépendante de l'activité de collecte, le corpus des données audiovisuelles dont l'analyse a besoin pour satisfaire ses objectifs n'a évidemment plus rien à voir avec tel ou tel corpus de terrain. Le corpus d'analyse se construit et s'enrichit exclusivement en fonction des objectifs de l'analyse elle-même. Dans [STO 11b], on trouve deux exemples de la constitution d'un corpus d'analyse qui se nourrissent de données provenant de différents corpus de terrain : le premier exemple est celui de l'analyse de la fabrication traditionnelle du pain en France et au Portugal [DEP 11b], le deuxième celui de l'analyse comparative du regard sur les *Mille et une nuits* et des usages créatifs que certains artistes en font [CHE 11b]. Dans l'un et l'autre cas, le corpus d'analyse se compose[10] non seulement de données provenant de différents corpora de terrains (terrains réalisés dans le cadre du programme Archives Audiovisuelles de la Recherche, AAR) mais en plus possède sa propre dynamique d'enrichissement ou encore de mise à jour.

5) L'étape de publication/diffusion de données s'appuie à son tour sur un corpus de données – le *corpus de publication** qui n'est pas nécessairement coextensif avec le corpus de terrain ni avec celui d'analyse. Par exemple, dans le cadre du processus de publication d'un séminaire de recherche ou d'un entretien filmé et analysé, les quatre corpus fonctionnellement distincts – le corpus de terrain, le corpus de traitement, le corpus d'analyse et, enfin, le corpus de publication – peuvent se rapprocher, voire (partiellement) se confondre. Cependant, ils restent

[10] Par exemple, dans le cadre du programme « Archives Audiovisuelles de la Recherche » (AAR), nous avons développé en 2007 un portail à thème consacré uniquement à l'Amérique Latine. Or le corpus analysé et publié sur ce portail se compose de données provenant de toute une diversité de terrains. En plus, ces terrains ont été en partie réalisés à l'intérieur du programme AAR (et selon les procédures définissant la préparation et la conduite d'un terrain dans le cadre de ce programme) mais en partie aussi à l'extérieur de celui-ci - par des acteurs individuels (chercheurs, documentaristes, etc.) ou collectifs (institutions de recherche, etc.). Cet exemple montre bien l'indépendance fonctionnelle entre *corpus de terrain** et *corpus d'analyse**.

fonctionnellement distincts, obéissent à des motifs et objectifs qui leurs sont propres, évoluent et sont gérés selon des critères spécifiques aux exigences des activités pour lesquelles ils constituent l'objet principal de travail. Ainsi, le corpus de publication peut se composer à la fois de données provenant d'un ou de plusieurs corpus de données analysées, de données déjà publiées, de données collectées mais pas obligatoirement analysées, etc. Le critère crucial pour évaluer la qualité d'un corpus de publication est celui de sa capacité à satisfaire une *intention de publication* (voir [NAN 05] et, plus particulièrement, des règles formelles régissant un genre choisi de publication [DPS 11]. Insistons sur le fait qu'un corpus de publication peut être composé de données *déjà publiées*[11]. Nous touchons ici la problématique devenue, aujourd'hui, très discutée de la *réingénierie* (en anglais, le *repurposing*) documentaire[12], la *re-médiation* de données textuelles déjà médiatisées – problématique qui concerne directement la protection juridique du contenu et la traçabilité des usages faits d'un contenu[13].

---

[11] De nouveau dans le cadre du Programme AAR et grâce à plusieurs projets de R&D français et européens (dont notamment le projet français SAPHIR, financé par l'Agence Nationale de la Recherche, et le projet européen LOGOS, financé dans le cadre du 6ème PCRD), nous avons pu développer et tester des *modèles de republication* d'entretiens de recherche qui ont été publiés initialement sur le site AAR sous forme d'un vidéo-livre interactif. Ainsi, par exemple, un entretien avec l'ethnomusicologue Sabine Trebinjac, réalisé par Aygun Eyyubova pour les AAR en juin 2007 à la FMSH à Paris, a été re-publié (moyennant les adaptations et enrichissements appropriés) sous forme de *dossiers bilingues* (français/anglais ; français/chinois ; français/turc, etc.), d'un *dossier pédagogique* de niveau M1, d'un *dossier thématique* consacré à la musique ouïgoure ou encore, ensemble avec d'autres publications, sous forme d'un *vidéo-lexique* consacré aux musiques du monde.

[12] Jirasri Deslis montre, d'une manière exemplaire, comment utiliser les nouveaux médias sociaux et le web 2 pour une meilleure diffusion des connaissances scientifiques et comment contribuer à une meilleure « acculturation » scientifique des communautés virtuelles composant cette (selon l'expression de l'UNESCO) « nouvelle diversité culturelle » issue de la révolution technologique du numérique (voir aussi [DES 11a, DES 11b, DES 11c]).

[13] Voir à ce propos une expérimentation très innovatrice décrite par Francis Lemaitre et Valérie Legrand-Galarza qui s'appuie sur une nouvelle technologie (le V.D.I. – *Versatile Digital Item* ; http://www.ict-convergence.eu/) permettant de tracer tout usage d'une vidéo de sorte que les propriétaires du contenu de la vidéo utilisée ou réutilisée peuvent ainsi suivre son destin sur le web [LEL 11]. L'expérimentation s'appuie ici sur le cas concret des archives PCIA (Patrimoine Culturel Immatériel Andin) réalisée par Valérie Legrand-Galarza en coopération avec les communautés quechuas concernées. Un des buts principaux de ces archives est de permettre à tout enseignant, tout chercheur l'exploitation, l'usage des vidéos composant le fonds PCIA tout en respectant les droits moraux et juridiques à proprement parler des communautés concernées – dont notamment le droit à la révocation d'un contenu diffusé et/ou réinterprété et re-publié (voir [LEL 11].

6) Enfin l'étape de conservation de données collectées, traitées, analysées ou publiées dans le cadre d'un projet ou programme de constitution et de diffusion d'un patrimoine de connaissances s'occupe plus particulièrement de la sélection de ces données pour en faire un *corpus patrimonial* à proprement parler avec, entre autre, comme objectif de sauvegarder à long terme et de transmettre d'une génération à une autre un ensemble de connaissances et de valeurs, c'est-à-dire, *in fine*, une *culture* et ses *traditions*. Formellement parlant, un corpus patrimonial peut se rapprocher à l'un ou l'autre type de corpus précédemment identifié et discuté mais fonctionnellement parlant, il diffère de tous les autres. Cela implique également que même si techniquement parlant la constitution d'un corpus patrimonial se réduit à une simple sélection de données provenant des différents corpus qui caractérisent le travail de la constitution, de l'analyse et de la diffusion d'un patrimoine de connaissances, cette activité technique de la sélection est obligatoirement soumise à une *politique patrimoniale* (à proprement parler ou lato sensu). Ainsi, la constitution d'un corpus patrimonial documentant, par exemple, une problématique scientifique spécifique peut avoir comme but de créer et de faire évoluer une véritable culture en la matière avec ses valeurs, ses traditions, sa doxa, ses hétérodoxies, mais aussi ses grandes découvertes et inventions, ses « héros », ses événements et dates marquants, ses savoir-faire, etc. La constitution – et l'enrichissement – d'un corpus patrimonial peut aussi avoir comme but de promouvoir l'identité et la notoriété d'une institution – son excellence, sa compétitivité, son attrait, etc. Comme on le sait, la promotion de l'identité d'une institution n'est plus du tout l'apanage des sociétés et entreprises commerciales mais est devenue, suite aux profonds changements affectant le cadre global de la recherche et de l'enseignement public, une préoccupation majeure des universités et autres établissements publics de la recherche et de l'enseignement supérieur[14].

[14] Comme preuve au moins indirecte pour cette affirmation, on peut citer la prolifération de vidéothèques et autres « chaînes » en ligne documentant les activités (de recherche, d'enseignement, etc.) de telle ou telle université, tel ou tel établissement de recherche. Cependant, l'impression d'une certaine « naïveté » prédomine lorsqu'on consulte bon nombre de ces initiatives. Cette impression de naïveté renvoie au fait que les concepteurs de ces « chaînes » ou vidéothèques institutionnelles semblent considérer que le fonds audiovisuel diffusé *constitue déjà en soi* un corpus *patrimonial* (ou autre : *de terrain*, *analysé*, *publié/republié*, etc.) à proprement parler. D'une manière plus générale, très rare sont les initiatives qui distinguent nettement entre *fonds* de données audiovisuelles, dont dispose une vidéothèque ou une chaîne institutionnelle, et *corpus* de ressources audiovisuelles – un corpus étant une sélection de données audiovisuelles qui proviennent d'un fonds, qui remplissent une *fonction*, et qui obéissent à un *objectif* (de documentation d'un terrain, d'interprétation d'un sujet, de publication d'un événement, de promotion d'une identité institutionnelle, etc.).

1.3. Des modèles de description

Le travail concret de la constitution et de l'exploitation de ces différents corpus fonctionnellement distincts est obligatoirement déterminé par des *modèles (conceptuels* ou *métalinguistiques)* formant le *cadre de référence*, à la fois intellectuel et pratique, pour les acteurs engagés dans ce processus complexe qui est celui de la numérisation d'un patrimoine de connaissances, c'est-à-dire la collecte, le traitement et l'analyse, la publication ou la republication et la préservation à plus long terme de documents de toute sorte témoignant du savoir et du savoir-faire, des croyances et des valeurs, des normes et des règles, des habitudes et des comportements, etc. d'une personne, d'un groupe social, d'une institution ou encore d'un pays, d'une région, d'une époque, etc.

On comprend dès lors, peut-être mieux la position stratégique occupé par le *modélisateur* dans le processus de travail de constitution et de diffusion de patrimoines de connaissances. Le modélisateur prépare, développe, gère, etc. tous les *modèles* nécessaires pour mener à bien les activités impliquées dans la constitution, l'exploitation, la diffusion et la conservation d'un patrimoine de connaissances sous forme de corpus de données textuelles (dans notre cas, plus particulièrement audiovisuelles). En comparaison aux différents types fonctionnellement distincts des corpus qui jalonnent le processus de constitution et de diffusion d'un patrimoine de connaissances, il s'agit notamment :

– des *modèles de description** dont l'*analyste** (autre rôle phare dans le processus de travail en question) a besoin pour procéder à la description, indexation et annotation de corpus audiovisuels ;

– des *modèles de collecte* dont a besoin le *producteur du terrain*, c'est-à-dire la ou les personnes ayant en charge la définition et la réalisation d'un terrain (lato sensu) consistant en la collecte *raisonnée* de données pour documenter un domaine de connaissances, une manifestation ou un événement ;

– les modèles qui servent aux spécialistes du *derushage* et aux monteurs pour procéder au traitement *technique* et *auctorial* d'un corpus de données collectées (dans le cas des données audiovisuelles il peut s'agir, par exemple, d'un formulaire guidant la sélection de rushes pertinents et d'un scénario type guidant le montage des rushes en une nouvelle production audiovisuelle) ;

– les *modèles de publication* qui servent au *rédacteur-auteur* pour organiser et réaliser une ou une série de publications/re-publications d'un corpus audiovisuel ;

– et, enfin, le ou les modèles permettant de sélectionner et de réaliser des *corpus patrimoniaux* à proprement parler.

Comme déjà dit, dans ce livre nous sommes concernés par un, et un seul type de modèles, à savoir par les modèles de description qui constituent la ressource métalinguistique indispensable à l'analyse et à l'indexation à proprement parler mais aussi à l'annotation et à l'interprétation plus ou moins libre de corpus audiovisuels destinés, par exemple, à être publiés sous forme d'un site portail, d'un dossier thématique, d'un vidéo-lexique, etc. Nous traiterons dans [SDL 12] plus en avant la question des modèles pour procéder à la constitution de corpus de terrain et de corpus patrimoniaux ; dans [DPS 11], on trouvera des éléments relatifs à la définition de modèles nécessaires pour procéder à la constitution de corpus de données audiovisuelles pour publication selon des genres (ou formats) particuliers.

Insistons sur le fait que la spécification, le développement et le suivi de ces différentes classes de modèles doivent reposer sur *l'un et le même métalangage de description*. Dans notre cas, il s'agit du *métalangage de description ASA**[15] (que nous appelons également *ontologie générique ASA*). Même si nous ne l'utilisons ici que pour développer les *modèles de description* dont a besoin l'analyste pour procéder à la description du contenu et/ou des plans visuels et acoustiques d'un corpus de textes audiovisuels, la prétention de ce métalangage est de pouvoir servir, *in fine*, à la définition et au développement de toutes les classes de modèles que nous venons de citer.

1.4. De l'activité de l'analyse de corpus audiovisuels

L'analyse d'un *texte** audiovisuel ou d'un corpus de textes audiovisuels constitue à la fois, comme nous venons de le voir, un ensemble d'activités concrètes ayant comme objectif de satisfaire à un manque de connaissances ou d'informations, et, une des étapes principales dans le processus de la constitution et diffusion d'un patrimoine de connaissances.

Comme déjà dit, les objectifs de l'analyse peuvent être très disparates, obéir à des intentions particulières fort diverses. En parlant de l'*analyse d'un corpus textuel ou audiovisuel*, on pense souvent soit à des activités de recherche académique sur ou

[15] L'acronyme « ASA » veut dire *Atelier de Sémiotique Audiovisuelle* et renvoie d'une part à notre référence théorique principale qui est la sémiotique du texte (voir [GRE 76, GRE 79, STO 83, STO 03]) et d'autre part au projet ASA-SHS, financé par l'ANR pendant la période comprise entre début 2009 et fin 2011, qui nous a permis de développer et de tester « à grandeur réelle » le dit métalangage de description.

à partir de corpus[16] soit encore à des activités d'expertise professionnelle (dans le cadre, par exemple de la veille d'information [STO 11e] et de l'analyse critique et comparative de données et de connaissances « stratégiques » pour un secteur d'activité, une entreprise, une institution, ...).

Mais l'analyse de corpus audiovisuels est tout aussi importante et incontournable dans d'autres activités phare de notre vie économique, sociale et culturelle. Ainsi est-elle un élément indispensable de l'*activité pédagogique* aussi bien en amont qu'en aval (« *in situ* », « en classe ») : en amont, elle fait partie de la préparation même de l'enseignement, en aval, elle est une des activités les plus importantes dans l'appropriation des connaissances – linguistiques ou autres (voir à ce sujet les très intéressantes études et explications dans [MCK 06, BRA 07, HQS 07, PBR 07] ou encore [KEM 02].

Les activités de l'analyse forment également une activité centrale dans la constitution et le suivi d'une bibliothèque (ou d'une vidéothèque) numérique. Elles servent aux objectifs « classiques » de description, de classification et d'indexation des données textuelles (et, plus particulièrement audiovisuelles) composant le fonds d'une bibliothèque ou d'une vidéothèque. Les objectifs sont variés incluant ceux qui font partie du métier du bibliothécaire, de l'archiviste ou du documentaliste. Il s'agit de classer les données (par collection, par auteur, par sujet, par genre, par langue, par année, etc.), d'identifier et de décrire des « paragraphes » ou segments qui possèdent un intérêt particulier et de les rendre accessibles à un public intéressé, d'en produire des *versions enrichies* afin de les rendre pertinents à des objectifs plus spécialisés (tels que l'enseignement ou l'apprentissage) ou encore d'en proposer, pour certaines parties choisies, des *adaptations linguistiques*, voire *pragmatiques* à destination d'un public ne pouvant apprécier la valeur des données originales faute de compétences linguistiques et culturelles adéquates (voir également [SAC 11]).

Mentionnons également le cas de la publication et/ou de la republication d'une donnée audiovisuelle ou d'un corpus entier de données audiovisuelles. L'analyse joue un rôle central et incontournable « *en amont* » de cette activité – aucune publication prétendant à une certaine valeur intellectuelle ne peut se dispenser d'une analyse préalable des données qu'elle propose à un public visé ou intéressé. De nouveau, l'analyse peut consister en une série d'activités d'identification, de description et de classification de données possédant une pertinence pour une publication visée. Mais elle peut consister également en une expertise approfondie des données sélectionnées pour une publication – expertise servant de base à l'auteur

[16] En linguistique, il existe toute une branche spécialisée consacré à cet objet ; en rhétorique et analyse du discours, on s'appuie sur les corpus textuels pour comprendre le développement syntagmatique de genres particuliers de discours (voir par exemple [BCU 07, MCK 06]).

(individuel ou collectif) pour étayer son point de vue, sa vision sur la question qui motive la publication.

Enfin un dernier exemple montrant la centralité des activités de l'analyse de corpus textuels et, plus particulièrement, audiovisuels, est celui de la constitution et du suivi des archives personnelles comme, par exemple, la gestion des données audiovisuelles documentant la vie (c'est-à-dire l'agenda temporel et le cheminement narratif) d'une personne ou d'une famille sous forme de fêtes, de réunions et réceptions, de voyages, d'activités rituelles, etc. On le sait bien que depuis la « banalisation » des outils d'enregistrement audiovisuel, d'innombrables photos et vidéos sont réalisés à tout instant par un nombre de plus en plus croisant de personnes individuelles, de familles, de groupes informels de personnes, etc., pour documenter l'agenda temporel de leur vie de tous les jours. Or, cette masse gigantesque de données produites doit obligatoirement subir un minimum d'activité de sélection, de classification et de description pour « devenir » une archive personnelle, familiale, d'amis, etc., au sens propre du terme. Autrement dit, le souvenir personnel, le passé familial documenté sous forme de corpus patrimoniaux qui composent l'archive d'une personne, d'une famille, d'un groupe d'amis, etc., sont les produits de l'analyse – sans analyse, pas d'archive et pas d'histoire non plus !

Toujours est-il que l'analyse – peu importe le contexte dans lequel elle se déploie et les objectifs (scientifiques, pédagogiques, professionnels ou « simplement » personnels) qu'elle poursuit – repose nécessairement sur un ou des systèmes métalinguistiques, un *métalangage de description** sans lequel elle ne pourrait pas être exécutée. L'*usage concret* d'un tel métalangage prend le plus souvent la forme de formulaires (dynamiques), c'est-à-dire d'interfaces de champs typés par une sélection de concepts pertinents à l'usage en question (voir par exemple, la figure 4.2 dans le chapitre 4). Nous allons, dans les chapitres suivants, montrer en quoi ressemblent tels formulaires dynamiques et comment on les élabore.

Pour terminer cette petite présentation générale consacrée à l'analyse de données audiovisuelles, remarquons qu'il – c'est-à-dire le travail d'analyse d'un texte audiovisuel – peut être comparé à un *travail de terrain* « tout court » en sciences humaines et sociales (il peut être parfaitement rapproché, par exemple, à un travail de terrain en ethnologie ou en archéologie). Afin pour qu'un travail de terrain aboutisse (bien ou mal, c'est une autre question), l'analyste a recours à des *modèles* (implicites ou plus ou moins explicites) pour « donner du sens » à son terrain ainsi qu'à une méthodologie de travail. Ainsi, l'ethnologue travaillant dans une communauté villageoise aura recours à des connaissances linguistiques pour pouvoir collecter et classer les données du parler villageois, il aura recours à des connaissances sociologiques pour pouvoir collecter, classer et interpréter des

données relatives à l'utilisation de l'espace naturel du village et à la construction d'un espace social commun, il aura recours à des connaissances musicologiques pour pouvoir collecter, classer et interpréter les données pertinentes à la culture musicale des villageois, et ainsi de suite. Ces connaissances peuvent posséder un statut plus ou moins implicite ou, au contraire, explicite, peuvent être plus ou moins appropriées ou inappropriées par rapport à l'objet, peuvent évoluer dans le temps et suite aux expériences accumulées par l'ethnologue et ses collègues travaillant avec lui ou sur des sujets comparables. N'empêche la collecte, le classement et l'interprétation des données d'un terrain réfèrent obligatoirement à un *schéma conceptuel* qui guide et oriente, qui *cadre* (au sens de l'anglais *framing*) le travail de l'analyste.

Les choses sont les mêmes lorsqu'on parle de la description ou de l'indexation d'un corpus textuel ou, comme c'est notre cas, d'un corpus de textes audiovisuels. Le corpus, c'est le terrain ; l'analyste, lui, peut être un chercheur, un enseignant, quelqu'un qui possède un intérêt professionnel ou simplement personnel pour ce terrain. Or, nous avons investi un très grand effort tant intellectuel que technique pour rendre ce travail de « terrain » le plus riche possible mais aussi le plus souple et adaptable aux intérêts et objectifs de l'analyste. Ce livre est consacré à cette question qui est, comme nous le verrons encore, d'une assez grande complexité.

1.5. De l'activité de l'indexation

Lorsqu'on parle de l'indexation, on a presque obligatoirement en tête une certaine vision stéréotypée de cette activité telle qu'elle est pratiquée dans un contexte de documentation, d'archivage ou encore de catalogage bibliothécaire.

Dans notre approche, *cataloguer* ou *classer* ne sont que deux activités particulières dans le processus de l'appropriation par l'utilisateur d'un objet textuel en général et audiovisuel en particulier, c'est-à-dire dans le processus de la transformation progressive de cet objet en une *ressource* sui generis* destinée à un public, un groupe d'utilisateurs visés. L'ensemble de cette transformation est supporté par une série d'activités typiques et récurrentes. Ces activités produisent certaines modifications dans l'objet *texte :*

- parfois, il s'agit de modifications dans l'organisation (formelle, physique, etc..) initiale même du texte (c'est le cas, par exemple, pour les activités de la segmentation et de l'extraction de *segments** (passages) dans un objet filmique, c'est aussi le cas de l'activité du montage) ;
- parfois, il s'agit d'adaptations du contenu « intellectuel » aux contraintes d'un contexte social spécifique (c'est le cas notamment de la description de l'objet dans son ensemble ou d'une de ses parties, de son enrichissement par des commentaires, d'aides, de référencements, etc.).

D'une manière générale, ces modifications ont comme résultat le repositionnement de l'objet textuel source dans le champ intertextuel propre au « paysage textuel » (au *textscape*) d'un acteur social donné. Ainsi, elles peuvent concerner l'organisation, la structure *textuelle* de l'objet textuel source comme, par exemple, son organisation syntagmatique et linéaire. Elles peuvent aussi concerner les rapports *méta-textuels* entre l'objet textuel source et le *regard*, le *point de vue* sur ce texte source exprimé soit par son auteur soit par une tierce instance sous forme de commentaires, d'évaluations, de comparaisons, de (ré-)interprétations et de (re-)classifications ou encore sous forme de conseils et de guides (d'utilisation). Elles peuvent porter plus spécifiquement sur une (re-)définition des rapports *hypertextuels* que l'objet textuel source entretient avec d'autres objets faisant partie du paysage textuel d'un acteur social sous forme de citations, de références, de « liens », etc. Elles peuvent également porter sur le *paratexte* de l'objet textuel source : elles peuvent concerner plus particulièrement l'« identité signalétique » de l'objet textuel – son *péritexte*, selon Genette [GEN 87] – qui peut être modifié, par exemple, à l'aide de nouveaux titres, de nouvelles préfaces ou postfaces. Elles peuvent enfin porter sur l'« impact » *épitextuel* de l'objet textuel source sur son environnement textuel sous forme, par exemple, d'activités de diffusion ciblée (nous pensons ici aux nouveaux médias sociaux, aux *buzz* et autres possibilités de faire circuler et connaître un objet textuel).

Dans tous les cas, pour toutes ces activités d'appropriation, et donc de modification de l'organisation de l'objet textuel et de son positionnement dans un champ intertextuel, l'indexation constitue en quelque sorte la *partie matérielle du faire* : une fois elle se manifeste sous forme d'une production de mots clés libres, une deuxième fois, sous forme d'une production de textes explicatifs, une troisième fois, sous forme d'ajouts de liens pointant vers d'autres ressources en ligne, une quatrième fois sous forme d'annotation visuelle ou acoustique, et ainsi de suite.

Ainsi, tout en reconnaissant l'existence d'une pratique « orthodoxe » de l'indexation, nous la considérons ici *lato sensu*, au sens large du terme en tant que, la *partie matérielle* du travail de la transformation de l'objet textuel en une *ressource*, un *bien* à proprement parler pour un groupe d'utilisateurs.

1.6. Quelques réflexions au sujet du cadre théorique de référence

Dans l'approche qui est la notre, la définition et l'élaboration des modèles de description des corpus audiovisuels réfèrent à la théorie sémiotique du *texte** (*audiovisuel ;* voir [STO 03]). Les principales facettes sémiotiques et cognitives du texte audiovisuel (ou du texte *tout court*) qui nous intéressent ici plus particulièrement, peuvent être illustrées d'une manière intuitive et simple à l'aide des sept questions typiques suivantes :

1. Dans un corpus/un fonds de textes audiovisuels, quel est le *texte** ou encore quel est le *segment** – (la partie) spécifique du texte – qui suscite un intérêt, une attention particulière (pour l'analyste ou un public donné) ?

2. De quoi *parle-t-on*, que *montre-t-on* dans le texte audiovisuel ou dans un de ses segments identifiés (c'est-à-dire quels sont les *domaines de connaissance thématisés*, quels sont les *sujets* abordés) ?

3. Selon quel point de vue, selon quel *cadre auctorial* (c'est-à-dire d' « auteur ») ces domaines et sujets sont-ils abordés (c'est-à-dire comment les sujets sont-ils pris en charge par leur « auteur » et comment sont-ils *mis en discours*) ?

4. Comment un sujet thématisé et interprété discursivement dans un texte ou une partie spécifique du texte audiovisuel est-il *exprimé*, *représenté visuellement* et *acoustiquement* (ou encore, quelle est la *mise en scène audiovisuelle* d'un sujet) ?

5. Comment un sujet sélectionné, interprété et mis en scène par son « auteur » se développe-t-il en un *tout cohérent* (aussi bien au niveau de sa *linéarité textuelle* qu'au *niveau de son intégration syntagmatique (et narratif)* pour former une ressource (de connaissances, d'informations, etc.) potentielle pour un public et un contexte d'usage donné ?

6. Dans quelle tradition s'inscrit ce « travail » textuel qui consiste en la *sélection*, le *traitement discursif*, l'*expression/la mise en scène audiovisuelle* du sujet et son *intégration* linéaire et syntagmatique en un tout cohérent et finalisé (autrement dit, quel est le *genre** auquel réfère ou duquel fait partie le texte, ou tel segment du texte, qui suscite un intérêt, une attention particulière (étant donné un certain contexte d'usage) ?

7. Etant donné d'une part le cadre de référence culturel d'un public donné ainsi que ses attentes/besoins et d'autre part le profil spécifique (l'*identité auctoriale*) du texte, quelles sont les opérations nécessaires à effectuer pour rapprocher le *texte* et le *public* (c'est-à-dire quelles sont les activités identifiées pour transformer le texte au sens d'une *ressource (intellectuelle) potentielle en une ressource (intellectuelle) réelle*) ?

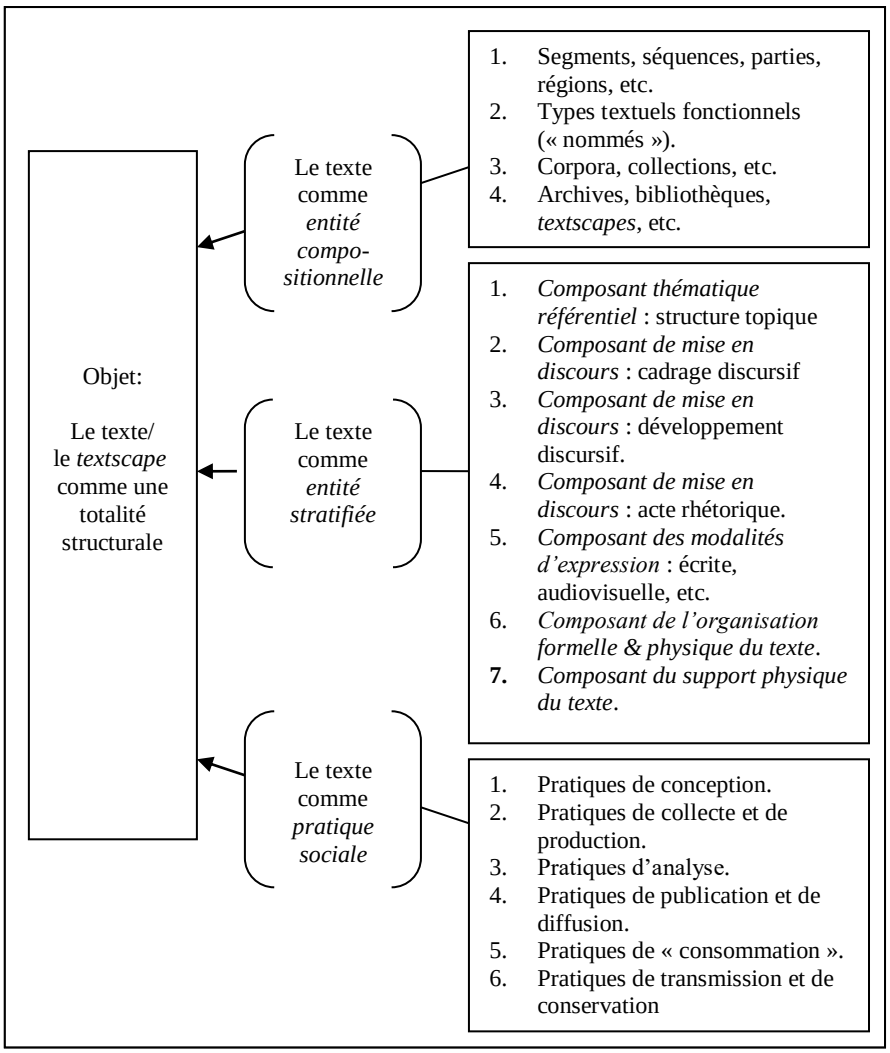

Figure 1.1. *Trois approches de l'objet texte (audiovisuel)*

Ces sept questions servent avant tout à « fixer » et à orienter les idées et les habitudes aussi bien *en amont* de la constitution (collecte, production) d'un corpus audiovisuel qu'*en aval* (c'est-à-dire pendant les phases de l'analyse à proprement parler d'un corpus ou d'un texte isolé pour en faire une ressource cognitive réelle : segmentation, description, classification, indexation, annotation ou adaptation).

En même temps, ces sept questions renvoient à l'échafaudage d'une s*émiotique du texte audiovisuel* qui s'articule autour de *trois grandes approches* (voir figure 1.1) complémentaires : l'approche traitant l'objet textuel comme une entité compositionnelle, l'approche traitant l'objet textuel comme une entité internement stratifiée et l'approche considérant le texte en relation à une pratique sociale. Ces trois approches caractérisent en effet toute analyse concrète d'un texte ou d'un corpus de textes audiovisuels.

Selon l'approche qui considère le texte (audiovisuel) comme une *entité compositionnnelle* (figure 1.1), le texte est un objet qui est, en principe, constitué par des *parties* « plus petites » (plus locales) dont la tâche est de contribuer à la réalisation progressive d'une intention, d'un objectif de communication. En même temps, tout objet textuel peut, lui-même, *faire partie* d'une entité plus grande pour laquelle il est censé accomplir des fonctions (d'information, de communication, etc.) plus ou moins précises (ainsi, par exemple, un texte audiovisuel peut faire partie d'une archive audiovisuelle où il documente un ensemble de sujets* qui relèvent de l'*univers du discours** de l'archive en question).

Dans le contexte de notre travail de recherche sur les archives audiovisuelles, l'approche considérant le texte comme une entité compositionnelle renvoie au fait qu'un texte audiovisuel considéré comme un *tout cohérent* peut être décomposé en unités (séquences, segments, parties, etc.) qui, à leur tour se « comportent » comme des textes audiovisuels « à part entière » formant ainsi des touts cohérents. Il renvoie également au fait qu'un texte audiovisuel n'est pas obligatoirement limité physiquement par le début et la fin d'une « vidéo » – un texte au sens d'un « signe porteur d'information » [STO 99] peut parfaitement se matérialiser en des « bouts » de vidéos physiquement séparés mais formant un tout sémantiquement. Un texte audiovisuel à support numérique doit donc être compris plutôt comme un réseau fonctionnel et hiérarchiquement intégré de parties textuelles qui, elles-mêmes, peuvent se présenter sous forme de documents se composant à leur tour de parties ou séquences.

Par exemple, un texte audiovisuel au sens d'une ressource de connaissances pour un public donné peut être un *film* en entier, c'est-à-dire, l'enregistrement d'un cours, ou d'un entretien. Il peut être aussi un *segment particulier du film* (c'est-à-dire un moment dans un séminaire ou cours consacré à une thématique, un sujet). Il peut être, enfin, une collection de segments appartenant à différents « films » (de cours, ou séminaires, etc.) mais qui, ensemble, font partie d'un nouveau tout cohérent du point de vue thématique ou du point de vue d'un usage spécifique (exemple : un ensemble de segments consacrés à une langue particulière « extraits » de différents films ou enregistrements composant un *nouveau film ré-linéarisé* ou un *dossier* hypermédias de segments filmiques destinés, par exemple, à un usage pédagogique).

Retenons donc qu'en tant que totalité structurale, le texte (audiovisuel) ne doit pas être réduit à une forme matérielle privilégiée – à un livre, par exemple, un journal, une vidéo, une photo, etc. En tant que totalité structurale, un texte (audiovisuel) peut être composé de toute une variété et d'un nombre ouvert de « formes matérielles privilégiées » telles que sont le livre, la photo ou le film. Autrement dit, un texte au sens d'une totalité structurale peut se présenter sous forme d'un corpus, d'une collection, d'une archive, d'une bibliothèque.

Enfin, le texte peut également être compris au sens d'un *textscape*. *Textscape* veut dire – d'une manière analogique à l'expression anglaise *landscape* – « environnement » (spatial, temporel, épistémique et culturel, etc.) de signes d'information dans lequel « baigne » un acteur social (un individu, un groupe, une institution, etc.) et qui lui sert de source et de ressources cognitives pour ses diverses activités. On peut en effet avancer le point de vue que le cadre de vie de tout acteur social est caractérisé, entre autre, par différents types ou genres de « paysages » dont fait partie le *textscape*. Pensons à l'espace urbain qui sert de scène aux agissements quotidiens de milliers de personnes. Celui-ci est pour ainsi dire « accoutré » de quantité de textes concrets sous forme d'affiches, d'enseignes, de publicités, de consignes, de tableaux d'information, de photos, de spots, de supports acoustiques et, enfin, de dispositifs mobiles permettant de compléter ce paysage d'information avec des informations provenant d'un ailleurs physique, voire temporel. On peut traiter indépendamment chaque représentant textuel d'un tel *textscape* ce que font la plupart des études consacrées à un genre textuel. Mais on peut également se poser la question de l'organisation globale d'un *textscape*, de sa cohérence interne, de son sens pour un public donné ou encore de sa production (de sa génération), de sa gestion et de son évolution. On voit bien que l'approche pour traiter l'objet « texte » reste identique, peu importe sa taille et sa complexité matérielle.

La deuxième approche considère le texte comme une *entité stratifiée* (comme une « pâte feuilletée », pour reprendre une image chère à A.J. Greimas). Selon cette approche, le texte est constitué d'un ensemble de plans spécifiques qui servent à traiter l'information (*lato sensu*) communiquée par le texte : choix d'un objet, prise en charge de l'objet choisi par l'énonciateur, expression écrite, orale ou encore audiovisuelle de la vision proposée par l'auteur de l'objet qu'il traite, etc.

Autrement dit, cette deuxième approche privilégie la compréhension et l'analyse de l'organisation interne de l'objet textuel, c'est-à-dire (dans la tradition de la linguistique structurale remontant à F. de Saussure) du *contenu* d'un texte et de son *expression* verbale ou, dans notre cas, audiovisuelle. Ainsi, un texte (audiovisuel) peut être décrit selon un ensemble de strates différentes parmi lesquelles on compte, notamment (voir aussi [STO 99]) :

– le plan *topique :* plan qui sert à la description de la sélection et de la thématisation d'un objet dont un texte ou un corpus de textes rendent compte ;

– le plan *énonciatif et discursif :* plan qui sert à la description de la façon dont l'énonciateur – l'« auteur » – du texte « façonne », « cadre » l'objet thématisé, comment il le traite en fonction aussi de son public, des objectifs poursuivis, etc. ;

– le plan *narratif :* plan qui sert à la description de la façon dont l'objet sélectionné et constituant le thème du discours de l'énonciateur ou de l'auteur du texte est introduit et développé « tout au long du texte », selon quel ordre, selon quelle logique syntagmatique ;

– le plan de l'*expression verbale et audiovisuelle :* plan qui sert à la description de la ou des modalités d'expression choisies pour communiquer l'objet thématisé et mis en discours ;

– le plan de l'*organisation scénographique* du thème de discours : plan qui sert à la description de l'inscription dans l'espace (par exemple, dans une surface bidimensionnelle) et dans le temps (par exemple, dans une durée ou à un moment précis) de l'objet thématisé, mis en discours et exprimé à l'aide d'un ou de plusieurs médias ;

– le plan de la *matérialisation* : plan qui sert à produire une instance matérielle du texte – sous forme, par exemple, d'un texte manuscrit ou imprimé, sous forme d'un texte numérique, etc.

C'est cette deuxième approche que nous privilégierons dans ce livre. Dès le troisième chapitre, nous verrons comment nous avons essayé d'intégrer la distinction entre les différents plans textuels dans le *métalangage de description** de l'*univers du discours** d'une archive audiovisuelle. Comme nous le verrons, un tel métalangage de description se « concrétise » sous forme d'une *bibliothèque de modèles de description** qui forme des configurations assez complexes dont les différents éléments proviennent d'un ensemble de *ressources métalinguistiques** que nous présenterons plus en détail dans la quatrième partie de ce livre.

Enfin, la troisième approche considère le texte comme une *pratique sociale* ou encore comme une entité autour de laquelle s'organise une pratique sociale. Selon cette approche, sont prises en considération davantage les activités autour du texte (en tant qu'entité compositionnelle et stratifiée) pendant toute sa « durée de vie », c'est-à-dire de sa conception et naissance jusqu'à son éventuelle disparition ou transformation en une autre entité textuelle. Autrement dit, tout texte est obligatoirement inséré dans un réseau de pratiques et d'activités. Ces pratiques et activités contribuent au *formatage*, au *profilage* du texte et sont elles-mêmes tributaires de son identité, de son profil spécifique. Ainsi, un texte audiovisuel se composant du montage d'un ensemble d'extraits d'entretiens avec des chercheurs sur une problématique spécifique possède obligatoirement son identité, son profil spécifique qui le rend (*a priori*) plus ou moins aptes ou, au contraire, récalcitrants à des activités d'adaptation et d'exploitation dans des contextes d'usages spécifiques.

Nous nous sommes intéressés au *texte au sens d'une pratique sociale* dans le cadre du programme Archives Audiovisuelles de la Recherche que nous avons mis en place en 2001. Le programme AAR étant un programme de recherche et de

développement consacré à la constitution, diffusion, conservation et exploitation de patrimoines scientifiques et culturels à l'aide de l'audiovisuel numérique, un des grands enjeux de ce dernier a été de définir et d'« instrumenter » les principales étapes et activités qui composent la *chaîne du processus de production et de diffusion de patrimoines audiovisuels*. Nous avons décrit d'une manière plus détaillée ce cadre dans [STO 11a].

Le point crucial que nous souhaitons souligner ici, est le fait que d'une étape à une autre, non seulement le rôle mais aussi l'identité du texte change et, cela en fonction, des tâches et activités en question. Or, cela a des conséquences directes, par exemple, sur la façon comment on considère ce qu'est un *corpus de textes* (dans notre cas, audiovisuels) et comment on le traite [STO 11a].

Ainsi, distinguons nous, comme déjà discuté ci-auparavant, entre différents types fonctionnellement spécialisés de corpus textuels : des corpus qui documentent un « terrain » (de recherche, par exemple), des corpus qui servent à documenter un patrimoine (en vue, par exemple, d'être diffusé sur le web sous forme d'une archive audiovisuelle) et qui peuvent très fortement varier par rapport à des corpus de terrain (aussi bien dans leur composition interne que par rapport à la « nature » des textes qui les composent éventuellement), des corpus fonctionnellement spécifiques servant à une publication particulière ou encore des corpus spécifiques de textes choisis pour être conservée pour une longue durée. Le profil du texte, son identité, pour parler ainsi, varie selon son appartenance à l'un ou l'autre des corpus fonctionnellement spécialisés. Il est donc en quelque sorte en vain de parler du texte « hors contexte » (et de le traiter dans ce sens).

Chapitre 2

Le studio *atelier de sémiotique audiovisuelle* – une brève présentation

2.1. Un environnement de travail pour l'analyse de corpus de textes audiovisuels

Avant de traiter les questions relatives à la conception et au développement de modèles de description de corpus audiovisuels à l'aide d'un métalangage de description, nous allons présenter d'une manière synthétique l'environnement de travail de l'analyste de corpus audiovisuels – environnement s'intitulant *Studio ASA*[17] (pour une présentation détaillée du Studio ASA, voir [STO 11a]).

Le Studio ASA repose, d'une part, sur une vision globale des activités donnant corps au processus de l'appropriation d'un texte audiovisuel par l'analyste pour un utilisateur visé et, d'autre part, sur l'approche *sémiotique* du *texte* (dans notre cas, du

[17] L'acronyme *ASA* est une référence au projet de recherche qui a permis le développement de cet environnement. Il s'agit, comme déjà précisé dans le chapitre introductif de ce livre, du projet *Atelier de Sémiotique Audiovisuelle pour la description de corpus audiovisuels en sciences humaines et sociales* (ASA-SHS), projet financé par l'Agence Nationale de la Recherche (ANR) pendant la période comprise entre début 2009 et fin 2011 (pour plus d'informations, voir le carnet de recherche consacré à ce projet sur le portail Hypothèses.org : http://asashs.hypotheses.org/).

texte audiovisuel) que nous venons d'esquisser rapidement dans le premier chapitre de ce livre (voir aussi [STO 03]).

La représentation et la simulation des activités de l'appropriation, c'est-à-dire de la transformation qualitative d'un texte audiovisuel en une ressource, un *bien* pour un public se concrétise sous forme d'un ou de plusieurs *scénarios* d'activités guidées par des *modèles* de description (voir aussi [STO 11a]). Le scénario d'activités de base étant celui définissant les principales étapes d'un projet de production, d'analyse et de diffusion d'un patrimoine de connaissances [SDL 11], le scénario définissant celle de la transformation qualitative d'un texte audiovisuel en une ressource à proprement parler (un *bien* intellectuel) pour un public donné est celui qui définit les principales activités faisant partie de l'étape de l'analyse du texte audiovisuel ou d'un corpus de textes audiovisuels (nous y reviendrons plus loin dans ce livre ; voir chapitre 5). Le modèle de description à son tour tient compte de la spécificité à la fois de la partie du texte audiovisuel qui doit être traitée par l'analyste et du type des résultats à fournir par l'analyste.

Pour donner un exemple simple, l'identification d'un segment audiovisuel (c'est-à-dire d'un moment choisi dans une vidéo que l'analyste juge pertinent pour son propos) est un travail spécifique qui recouvre un ensemble de *décisions* intellectuelles et *gestes techniques* tels que l'*isolement* du segment audiovisuel (c'est-à-dire son « découpage » virtuel) dans le flux audiovisuel, l'*attribution d'un intitulé* au segment, la *production d'une notice textuelle* résumant le contenu et/ou signalant la raison pourquoi ce segment a été identifié par l'analyste, et ainsi de suite (voir aussi [DEP 11a]).

« Isoler un segment audiovisuel », « lui attribuer un intitulé », « produire une notice explicative », etc. sont des activités concrètes qui se déroulent selon un certain ordre (pas obligatoirement linéaire ni complètement prédéterminé) et qui acquièrent leur sens d'une part en référence à l'objectif qu'elles ont à satisfaire (c'est-à-dire identifier le ou les segments les plus importants aux yeux de l'analyste) et d'autre part en référence aux modèles de description qui guident le travail de la segmentation et qui, en quelque sorte, « obligent » les différentes activités à apporter des réponses à des « lieux problématiques » tels que « début du segment », « fin du segment », « titre », « sous-titre », etc.

Une analyse particulière (composée donc essentiellement d'une série de *décisions intellectuelles* et de *gestes techniques* guidés par des *modèles*) d'un texte audiovisuel, s'appuie sur un certain nombre de *formulaires (dynamiques)* qui représentent matériellement les modèles de description sous-jacents et qui constituent l'*interface de travail* de l'analyste. Comme le montre la figure 2.1, un *formulaire* est une sorte de *texte interactif* qui guide l'analyste dans la saisie des informations dont a besoin le système informatique pour effectuer la transformation qualitative voulue par l'analyste afin de produire, à partir du texte audiovisuel existant, une ressource de connaissances pour lui-même ou pour un public visé.

La figure 2.1 montre une partie du formulaire qui invite l'analyste, souhaitant décrire le sujet *Civilisations en Asie* dans une vidéo, à procéder à un ensemble d'activités spécifiques. Nous discuterons de ces activités en détail tout au long de ce livre.

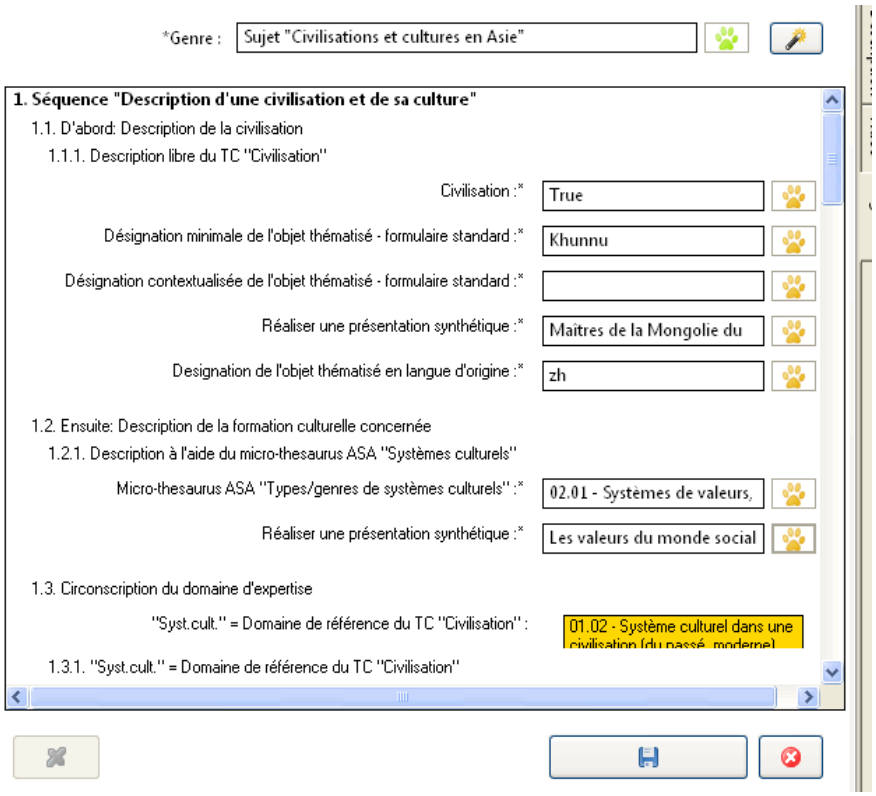

Figure 2.1. *Exemple d'un formulaire interactif de saisie d'informations relatives à un segment audiovisuel identifié comme pertinent par l'analyste*

Cependant, le sens global de ces activités d'analyse semble être intuitivement évident. En nous référant à la figure 2.1, il s'agit principalement :

- de désigner la civilisation en jeu par son nom ;
- de produire éventuellement une présentation textuelle ;
- de rendre compte – si pertinent – de l'expression en langue d'origine ;
- de développer les caractéristiques culturelles dont est question dans le segment audiovisuel analysé ;

- d'utiliser un micro-thesaurus pour identifier le ou les types de formations culturelles dont est question ;
- ainsi que, ce qui n'est plus visible dans la figure 2.1, localiser spatialement et temporellement la civilisation ou le fait civilisationnel en question ;
- et, enfin, produire éventuellement une analyse de la spécificité du discours consacré au fait civilisationnel (dans le cadre, par exemple, d'un enregistrement audiovisuel d'une conférence, d'un séminaire, d'un entretien, etc.).

Organisation et fonctionnement d'un formulaire tel que celui représenté dans la figure 2.1 seront expliqués plus loin (voir chapitre 5). Retenons ici que le Studio ASA se présente comme une interface composée d'une série de formulaires dont l'analyste peut se servir pour effectuer un travail plus ou moins simple et rapide ou, au contraire détaillé et systématique d'appropriation d'un texte audiovisuel.

Un formulaire, sorte de *texte interactif* qui guide l'analyste dans son travail, est un méta-texte interactif qui sert à l'analyste pour produire une description d'un objet-texte – vidéo ou une partie d'une vidéo.

Le formulaire interactif réfère, soulignons-le, obligatoirement à un *modèle de description** qu'il représente – pour parler ainsi – « matériellement ». Le modèle de description, lui, fait partie du *métalangage de description* propre à un domaine de connaissances ou d'expertise tel que celui du portail audiovisuel ARC (Archives *Rencontre des Cultures*)[18] consacré à la diversité culturelle et au dialogue interculturel.

La figure 2.2 propose un aperçu global de l'interface de travail de l'*Atelier de Description ASA*. L'Atelier de Description ASA est l'un des quatre composants du Studio ASA, les trois autres sont l'*Atelier de Segmentation* textuelle, l'*Atelier de Publication* de corpus de textes audiovisuels préalablement décrits et indexés ainsi que l'*Atelier de Modélisation* réservé à la ou aux personnes ayant en charge de préparer les ressources métalinguistiques (les modèles de description pour l'analyse) nécessaires au travail de l'analyste et du rédacteur-éditeur d'une publication. Nous le présenterons rapidement au chapitre 12 (section 12.3).

Le composant du Studio ASA appelé *Atelier de Publication* – à proprement parler – se base sur un outil informatique (sur une application web) s'intitulant *Semiosphère* et qui permet à une personne, ou à un groupe (une « communauté ») de personnes, de créer et de gérer ses propres archives audiovisuelles (également appelées « chaînes » au sens de You Tube ou de Daily Motion) et d'y publier ses

[18] ARC est un atelier d'expérimentation du projet ASA-SHS créé est coordonné par Elisabeth de Pablo et utilisé, entre autre, dans le cadre de la formation en communication interculturelle de niveau L2/L3 et M1/M2 à l'INALCO (Institut National des Langues et Civilisations Orientales) à Paris.

propres fonds sous forme, par exemple, de vidéothèques à thème, de lexiques et glossaires thématiquement organisés, de dossiers spécialisés, de dossiers bi ou multilingues, etc.

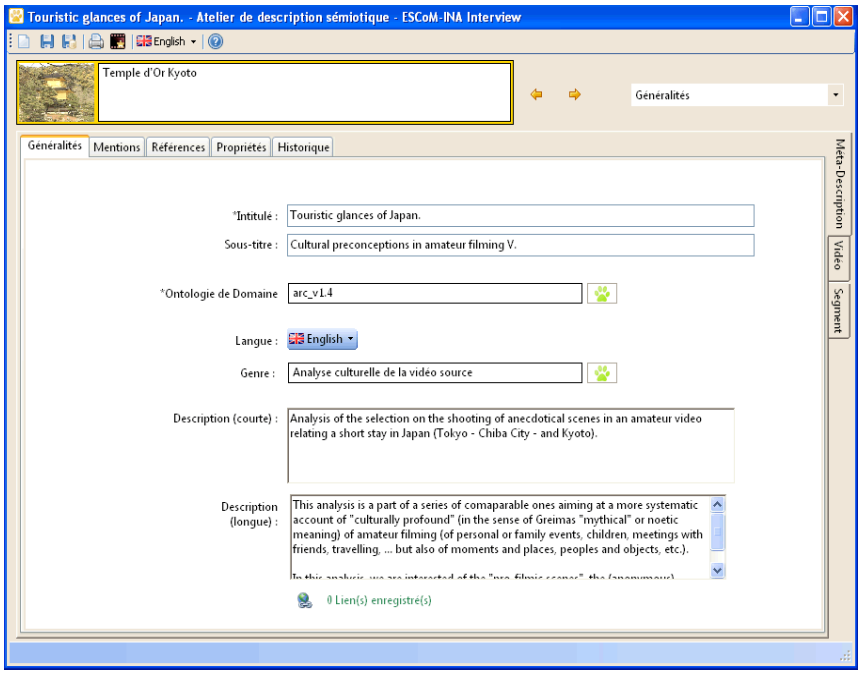

Figure 2.2. *Vue d'ensemble de l'interface de travail de l'Atelier de Description ASA*

Le composant du studio ASA appelé *Atelier de Segmentation* permet à l'analyste de visionner, une à une, les vidéos composant son corpus et d'y identifier le ou les segments (« passages ») qui l'intéressent et qu'il souhaite décrire, analyser davantage (bien sûr, l'analyste peut également se contenter d'analyser une vidéo « en entier » sans découpage en segments pertinents, etc.). Structure et usages possibles de l'atelier de segmentation sont décrits plus en détail dans [DEP 11a].

2.2 Présentation synthétique de l'Atelier de Description ASA

Voyons maintenant très rapidement le composant du Studio ASA appelé *Atelier de Description* – une présentation détaillée est fournie dans [STO 11a].

Comme la figure 2.2 le montre, cet atelier connaît *trois options d'analyse* ou de description et, pour chaque option, plusieurs *rubriques* qui identifient les facettes selon lesquelles chacun des trois niveaux d'analyse peut être abordé par l'analyste.

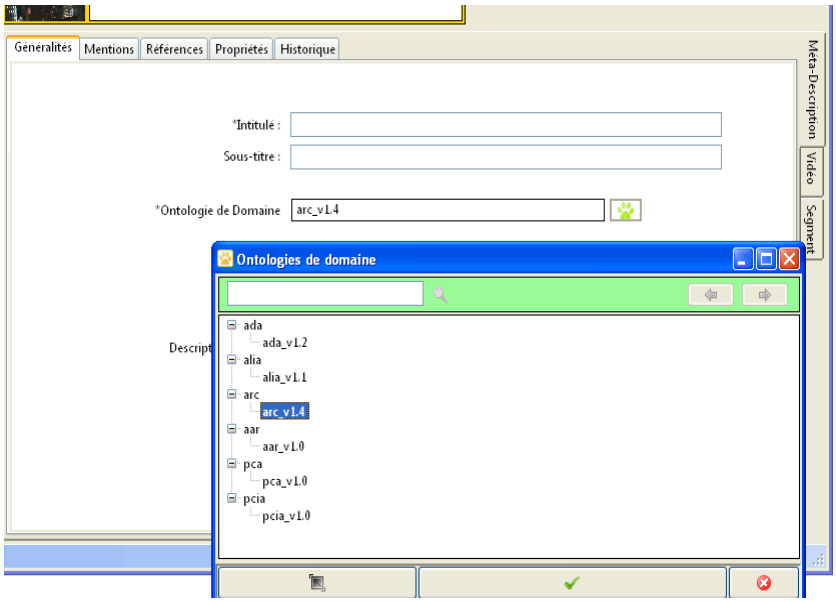

Figure 2.3. *L'interface de travail pour réaliser une méta-description*

La figure 2.2 présente plus particulièrement l'interface de travail accompagné de ses différentes rubriques permettant de réaliser une méta-description. Les rubriques composant cette option permettent à l'analyste de détailler davantage l'objet de son analyse et ses objectifs. Il ne s'agit donc pas d'une analyse du texte audiovisuel à proprement parler mais plutôt d'une explicitation du cadre de cette analyse – d'où le terme un peu docte de *méta-description*.

Comme les différentes rubriques de cette option le montre (figure 2.2), l'analyste mentionne, par exemple, l'intitulé de son travail, il peut fournir un résumé de celui-ci, préciser les droits d'auteur et les droits d'exploitation régissant son travail d'analyse, produire des références, etc. (pour connaître plus en détail cette interface, voir [CHE 11a]).

Est particulièrement important ici le choix du *domaine d'expertise* auquel se réfère l'analyste dans son travail. En effet, il s'agit du choix des ressources métalinguistiques à effectuer par l'analyste, plus particulièrement *la bibliothèque* des modèles de description* d'un texte ou d'un corpus de textes audiovisuels qui définit l'univers du discours d'une archive. Les domaines d'expertise de différentes

archives peuvent cependant se chevaucher. Ainsi, même si chaque archive dispose de « sa » bibliothèque de modèles de description, un modèle pris isolément peut figurer tel quel ou, moyennant quelques adaptations locales, dans différentes bibliothèques de modèles de description.

La figure 2.3 montre les huit domaines d'expertise actuellement disponibles parmi lesquels l'analyste peut choisir – à condition qu'il possède les droits appropriés pour le faire :

– *ada* est l'acronyme pour le domaine d'expertise *Atelier des Arkéonautes* dédié à la recherche en archéologie (il s'agit de la version 1.2 des ressources métalinguistiques consacrées à l'analyse des corpus audiovisuels documentant le domaine ADA) ;
– *alia* est l'acronyme pour le domaine d'expertise *Atelier Littéraire d'Ici et d'Ailleurs* consacré au patrimoine littéraire ;
– *arc* est l'acronyme pour le domaine d'expertise *Archives Rencontre des Cultures* dédié à la diversité culturelle et au dialogue interculturel ;
– *aar* est l'acronyme pour le domaine d'expertise couvert par le fonds audiovisuel de la vidéothèque de la Fondation Maison des Sciences de l'Homme (FMSH) à Paris ;
– *paca* est l'acronyme pour le domaine d'expertise de la vidéothèque *Patrimoine Audiovisuel Culturel de l'Azerbaïdjan* ;
– *pcia* est l'acronyme pour le domaine d'expertise de la vidéothèque *Patrimoine Immatériel Culturel Andin*.

Ces différents domaines d'expertise correspondent aux *ateliers d'expérimentation* du projet ASA-SHS. Il s'agit ici d'une liste parfaitement ouverte. Rien n'empêche d'y ajouter de nouveaux domaines d'expertise pour lesquels il faut, cependant, définir et concevoir les ressources métalinguistiques et les modèles de description appropriés. Ces ressources et modèles se constituent en partie de ressources et modèles déjà existants (c'est-à-dire définis et utilisés dans l'un des six ateliers d'expérimentation cités ci-dessus) et en partie de ressources et modèles nouveaux.

La figure 2.4 nous fournit un aperçu de l'interface de travail de la deuxième et la troisième option de l'Atelier de Description ASA. La deuxième option est intitulée *Vidéo*. Il invite l'analyste de décrire le texte audiovisuel dans sa *globalité* – contrairement à la troisième option réservée à l'analyse d'un ou de plusieurs segments spécifiques constituant un texte audiovisuel.

Cette troisième option permet effectivement à l'analyste d'analyser un texte audiovisuel plan par plan, sujet par sujet, d'une manière aussi détaillée qu'il le souhaite, selon différents points de vue, etc. Il permet également à l'analyste de ne prendre en considération que certains segments, voire qu'un seul.

Mais les deux options d'analyse sont abordées de la même façon – un passage, un segment d'un texte audiovisuel étant lui-même une *entité structurale*, un *texte** au même titre que le texte audiovisuel pris « dans sa globalité ».

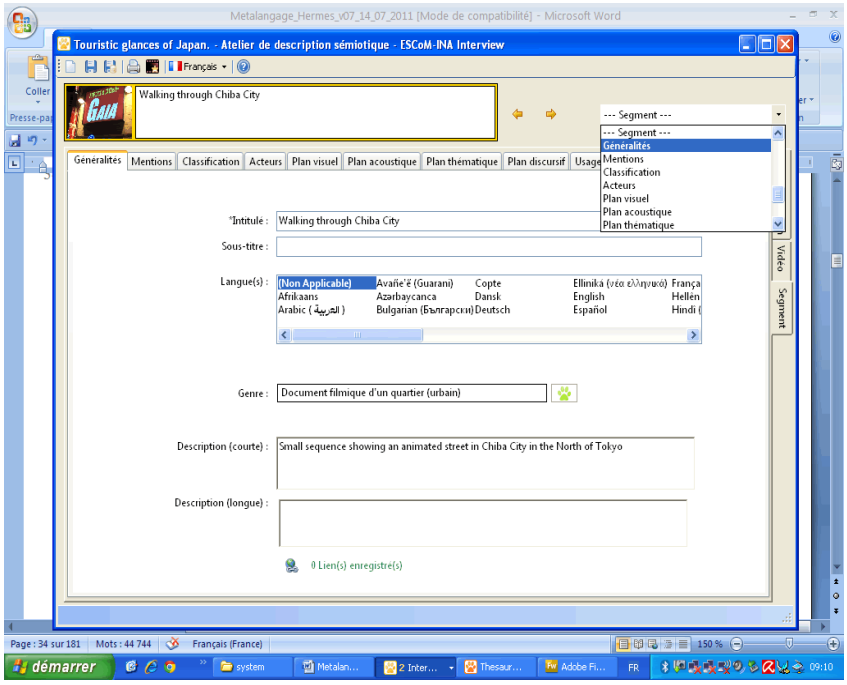

Figure 2.4. *L'interface de travail pour réaliser une description du texte audiovisuel « dans sa globalité » et/ou d'un segment audiovisuel spécifique*

La figure 2.5 montre les principales rubriques proposées à l'analyste pour procéder à une analyse d'un texte audiovisuel dans sa globalité ou de tel ou tel passage (« segment ») spécifique. Elles se regroupent pour former les *quatre types principaux* d'analyse que nous distinguons ci-dessous :

1) Les rubriques *Généralités*, *Mentions*, *Classifications* et *Acteurs* forment ce que nous appelons l'*analyse paratextuelle** d'un texte audiovisuel dans sa globalité ou d'un de ses segments. L'analyse paratextuelle est concernée par la description et l'explicitation de l'identité d'un texte audiovisuel et/ou de ses différents segments. Elle inclut également la description des aspects relatifs aux droits d'auteur, aux droits d'usage d'un texte audiovisuel (ou d'un de ses segments), etc.

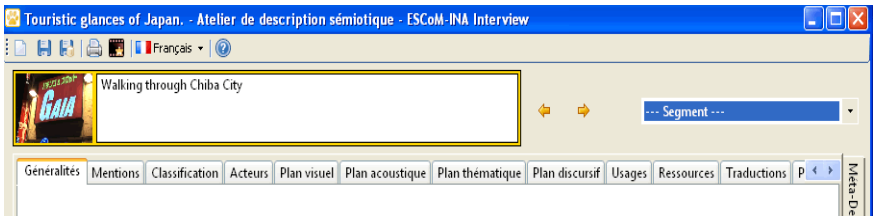

Figure 2.5. *Les principales rubriques pour réaliser une analyse d'un texte audiovisuel et/ou de segments spécifiques qui le constituent*

2) Les rubriques *Plan visuel* et *Plan acoustique* forment ce que nous appelons l'*analyse audiovisuelle** d'une vidéo ou d'un de ses segments. Il s'agit de l'analyse, plan par plan, des techniques de la mise en image ou de la mise en son d'un objet ou d'une situation représentée dans une vidéo. Ce type d'analyse est particulièrement intéressant dans le cadre de la constitution d'une bibliothèque de segments classés par types de plans, de cadrages, de mouvements visuels et/ou acoustiques, etc.

3) Les rubriques *Plan thématique* et *Plan discursif* forment ce que nous appelons l'*analyse du contenu** d'un texte audiovisuel, l'analyse des sujets traités et la façon dont ils sont traités dans un texte audiovisuel ou un corpus de textes audiovisuels. Cette forme d'analyse constitue le sujet principal de ce livre.

4) Enfin les rubriques *Usages*, *Ressources* et *Traductions* forment ce que nous appelons l'*analyse pragmatique** d'un texte audiovisuel ou d'un de ses segments. Celle-ci inclut notamment :

– l'explicitation des contextes pour lesquels un texte audiovisuel (ou l'un ou l'autre de ses segments) semble être approprié ;

– la préparation de la vidéo pour un usage particulier (sous forme de commentaires, de références, etc.) ;

– l'adaptation linguistique de la vidéo aux compétences de langue d'un public visé (cette adaptation peut se réaliser aussi bien au sens d'une traduction au sens classique qu'au sens d'un changement de registre de langue/de terminologie, etc., voir [STO 07, STO 10a, SAC 11].

Par ailleurs, l'analyste doit prendre, au début de son travail concret sur un texte ou un corpus de textes audiovisuels, une décision importante qui est celle de préciser le *genre d'analyse* qu'il projette de réaliser. La figure 2.6 montre les différents types

parmi lesquels l'analyste peut choisir. On y trouve les quatre principaux types d'analyse que nous venons d'identifier :

– l'analyse paratextuelle (ou « de base ») ;
– l'analyse thématique (du contenu) ;
– l'analyse audiovisuelle ;
– l'analyse pragmatique.

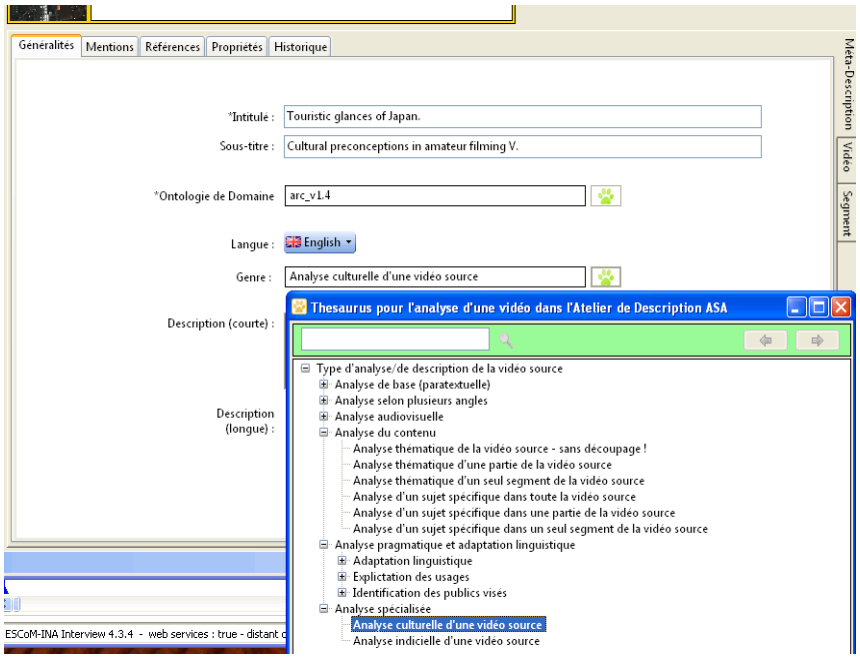

Figure 2.6. *L'interface permettant à l'analyste de sélectionner le type d'analyse approprié pour son travail*

Mais on y trouve encore d'autres formes d'analyse essayant de mieux rendre compte des différents intérêts et objectifs des parties prenantes dans l'analyse d'une archive. Nous avons ainsi identifié, par exemple :

– le type d'analyse selon *différents angles* (c'est-à-dire mêlant des intérêts audiovisuels, thématiques, pragmatiques, etc., en une seule analyse) ;

– le type d'analyse spécialisée telle que l'*analyse culturelle* (c'est-à-dire des préconstruits culturels motivant tel ou tel plan, telle ou telle partie d'une vidéo source) ou encore l'*analyse indicielle* au sens de l'historien italien Carlo Ginzburg (c'est-à-dire tel ou tel objet visuel comme indice pour tel ou tel sens figuratif [GIN 86]).

Cette liste est ouverte – elle peut être enrichie par de nouveaux types d'analyse, chaque type d'analyse faisant appel à une ou plusieurs rubriques, un ou plusieurs formulaires interactifs de l'Atelier de Description ASA.

Précisons cependant que même si nous l'avons prévu dès le départ, la version actuellement implémentée de l'Atelier de Description ASA ne permet pas une sélection dynamique des rubriques et de formulaires interactifs correspondant à un type spécifique d'analyse[19]. Espérons que l'on nous donnera l'occasion de pallier à cette faiblesse.

2.3 Quatre approches d'analyse du texte audiovisuel

Ces quatre types d'analyse forment autant d'approches pour expliciter, renseigner et, enfin, transformer qualitativement un objet audiovisuel quelconque en une ressource (intellectuelle, etc.) dans le cadre d'une bibliothèque ou d'une archive (en ligne), pour un certain public et pour un contexte d'usage.

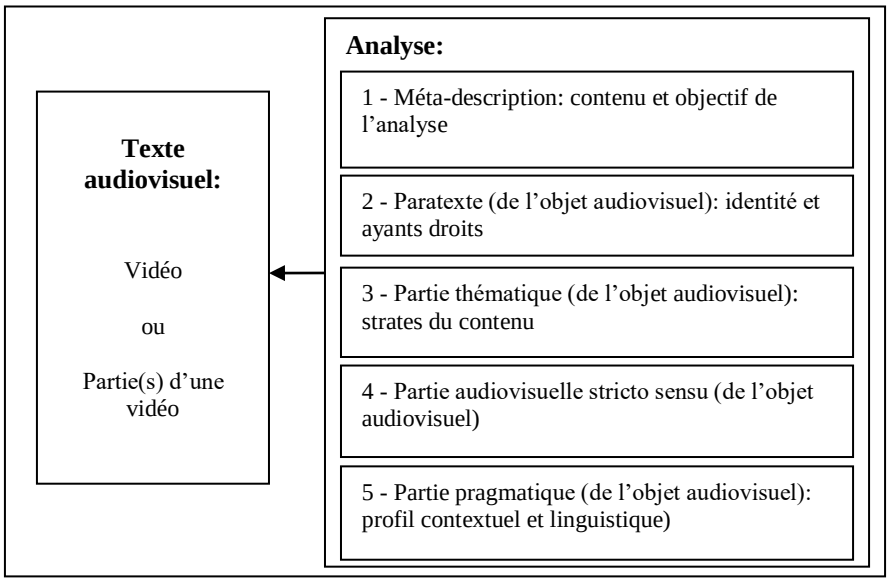

Figure 2.7. *L'analyse du texte audiovisuel – une vue d'ensemble*

[19] Comme nous l'avons déjà dit dans la présentation de ce livre, l'environnement de travail ASA a bénéficié d'un soutien remarquable de l'ANR, nous ayant permis avec une petite équipe de collaborateurs, d'en produire un prototype opérationnel et parfaitement utilisable, voire utilisé dans le cadre de notre programme de recherche AAR. Cependant, en l'état actuel, l'environnement ASA reste un prototype.

La figure 2.7 résume la vision générale qui sous-tend les rubriques et les formulaires organisant l'Atelier de Description ASA. Selon cette vision, l'analyse peut porter soit sur tout un corpus audiovisuel, soit sur une seule vidéo, soit encore sur une de ses parties qui intéresse plus spécifiquement l'analyste.

L'analyse peut se contenter d'une *simple description paratextuelle**. Techniquement parlant, celle-ci correspond *grosso modo* à produire des informations de base exigées par des standards aussi répandus et incontournables comme le Dublin Core[20] avec ses 15 éléments de description, l'*Open Archive Initiative* (OAI)[21] ou encore le schéma sémantique de la bibliothèque *Europeana*[22].

Cependant, l'analyse peut aller bien au-delà d'une « simple » identification paratextuelle. Dans ce cas – elle peut s'atteler, par exemple, à une *description du contenu** d'un corpus ou d'une vidéo. Celle-ci couvre, comme on le sait bien, toute une diversité d'approches, d'intérêts et d'enjeux dont le Studio ASA doit tenir compte. Un ensemble particulier mais très important de telles approches, intérêts et enjeux correspond à l'analyse du contenu d'un objet audiovisuel à l'aide d'un langage documentaire tel que le Web Dewey[23] ou le RAMEAU[24] de la Bibliothèque nationale de France (BnF), à l'aide d'un thesaurus[25], de glossaires et vocabulaires contrôlés[26] ou encore à l'aide d'une ontologie[27] afin de pouvoir participer aux

---

[20] Voir http://dublincore.org/
[21] Voir http://www.openarchives.org/
[22] Voir http://www.europeana.eu/schemas/ese/
[23] Voir http://www.oclc.org/ca/fr/dewey/versions/webdewey/default.htm
[24] Voir http://rameau.bnf.fr/
[25] Etant donné nos domaines de connaissances, nous pensons au thesaurus de l'UNESCO consacré aux sciences humaines et sociales et à la culture (http://databases.unesco.org/thesfr/), au Arts & Architecture Thesaurus (AAT) de la Getty Foundation (http://www.getty.edu/research/tools/vocabularies/index.html), du thesaurus MOTBIS pour l'éducation (http://www.cndp.fr/motbis/index.php/bibliographied/34-littterature-grise.html), etc.
[26] Pour le travail sur nos corpus audiovisuels [STO 11a, STO 11b]) ont été particulièrement importants, par exemple, les glossaires du site *Ethnologue – Languages of the World* (http://www.ethnologue.com/home.asp) pour désigner les langues du monde
[27] Pour notre travail, toute une série d'ontologies ont été importantes. Citons l'ontologie DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) du Laboratory for Applied Ontology de l'ISTE-CNR à Trento en Italie (http://www.loa.istc.cnr.it/DOLCE.html) qui nous a aidés à organiser nos propres ressources métalinguistiques ainsi que l'ontologie GOLD (General Ontology for Linguistic Description) qui nous a fournis les termes conceptuels (concepts) de base pour décrire le contenu de vidéos consacré à des recherches linguistiques (http://linguistics-ontology.org/); ou encore le CIDOC Conceptuel Reference Schema (http://www.cidoc-crm.org/scope.html) qui a fourni des références métalinguistiques nécessaires à la description de contenus audiovisuels relatifs à un patrimoine culturel.

évolutions du web sémantique vers une bibliothèque mondiale des ressources de connaissances.

Mais il ne s'agit ici que d'une vision très particulière de la description du contenu, fortement ancrée dans une certaine tradition de l'indexation telle qu'elle est pratiquée en documentation (scientifique, technique et professionnelle), bibliothéconomie ou encore archivistique.

N'oublions pas que la description du contenu fait également référence à l'analyse au sens d'une *expertise professionnelle* à proprement parler d'un contenu, voire au sens d'une *interprétation* et d'une *évaluation* critique d'un contenu donné. Nous-mêmes, par exemple, nous sommes particulièrement intéressés, dans notre enseignement et notre recherche en *sémiotique** audiovisuelle, à l'interprétation (herméneutique) des représentations culturelles, des connotations, des idées « préconçues » qui apportent un « plus de sens » au plan anecdotique des images le plus souvent totalement conventionnelles du monde social et naturel qui forment le matériau filmique des vidéos amateurs produites par tout un chacun pour documenter ses expériences, ses souvenirs, ses émotions [STO 10b].

Dans [LEG 11] est présenté un autre enjeu pour l'analyse systématique du contenu d'un corpus audiovisuel ouvert. Il s'agit de son emploi dans le cadre d'un programme de mise en valeur, de diffusion et de préservation d'un patrimoine culturel. L'exemple développé dans [LEG 11] est celui de l'expertise de réalisations et de pratiques communautaires (ici : des communautés quechuaphones au Pérou et en Bolivie) selon une approche proposée par l'UNESCO[28] pour son programme de la préservation du patrimoine culturel immatériel. Deux autres exemples très intéressants nous sont fournis dans [DEP 11b, CHE 11b] qui nous montrent l'expertise thématique dans un contexte de republication pédagogique d'une sélection de plans audiovisuels autour d'un sujet choisi tout en intégrant un ensemble de référence de base au standard LOM[29].

Cependant, ce type d'analyse (c'est-à-dire du contenu audiovisuel) ne constitue pas réellement le centre d'intérêt, par exemple, des grandes banques de données visuelles ou sonores. Ces banques de données sont plutôt intéressées de classer et de rendre accessibles leurs fonds sur la base de critères « formels » tels que *types d'objets visuels* ou *acoustiques*, *cadrage*, *plan* et *mouvement visuels* pour représenter une situation/un objet filmé, *durée* d'un plan visuel ou acoustique, etc. Une telle classification ne peut être réalisée que par une *description audiovisuelle** (au détriment d'une description du contenu).

---

[28] Il s'agit du cadre proposé par l'UNESCO pour analyser et évaluer un patrimoine culturel immatériel : http://www.unesco.org/culture/ich/index.php?lg=fr&pg=00001

[29] Dans notre cas, il s'agit de la version française LOMFR du standard *Learning Objets Models* (http://www.lom-fr.fr/).

Aucune des approches citées ne tient réellement compte, cependant, de la circulation des contenus et des ressources sur le web, du fait que l'Internet est un espace numérique intrinsèquement multilingue, couvrant des contextes et des usages des plus différents et constituant la base pour l'expression, l'échange mais aussi la confrontation d'une multitude d'identités et de références culturelles. Or, tout objet audiovisuel possède son identité, sa spécificité. C'est l'objectif de la *description pragmatique** de l'objet audiovisuel d'expliciter, voire d'adapter (c'est-à-dire de modifier plus ou moins profondément) cette spécificité en termes d'intérêt potentiel pour tel ou tel public, tel ou tel usage afin de soutenir et d'accompagner sa médiatisation (et/ou re-médiatisation) via les différents réseaux numériques.

Nous verrons dans les parties suivantes de ce livre, comment nous avons essayé de tenir compte de cette incroyable multitude d'exigences, d'intérêts, d'attentes, d'enjeux, de besoins, d'objectifs, etc. d'analyse tout en les traitant en référence à un *cadre théorique et méthodologique commun*.

2.4 Modèles de description et formulaires interactifs de travail

Comme nous venons le voir, l'interface de travail de l'Atelier de Description est composé de rubriques et chaque rubrique est composée d'un ou d'une bibliothèque de formulaires interactifs. Considérons maintenant plus en détail le formulaire interactif lui-même

Tout formulaire interactif composant l'interface de travail de l'Atelier de Description ASA représente, comme déjà dit, un *modèle de description*. Un *modèle de description** fait partie d'une *bibliothèque de modèles de description**. Chaque domaine de connaissances ou d'expertise possède *sa* bibliothèque de modèles de description. Certains modèles sont transversaux aux différents domaines, d'autres sont propres à un domaine donné, d'autres encore peuvent se retrouver dans plusieurs domaines (mais pas dans tous les domaines).

Un modèle de description est, à son tour, entièrement défini par les *ressources métalinguistiques ASA*. Comme nous le verrons encore dans la quatrième partie de ce livre, les ressources métalinguistiques ASA se constituent :

1. D'un *méta-lexique (ou vocabulaire) conceptuel** représentant les deux dimensions centrales qui forment tout domaine de connaissances/d'expertise : a) les *objets d'analyse** et b) les *activités* ou actions* d'analyse, c'est-à-dire les activités à l'aide desquelles on identifie, localise, dénomme, désigne, interprète, commente, etc., tel ou tel objet (ou type d'objets). Bien entendu, affirmer que les *objets d'analyse* et les *activités d'analyse* sont deux dimensions centrales dans la définition et le développement de modèles de description, ne signifie pas qu'elles sont les

seules dimensions à prendre en considération. D'autres dimensions, toutes aussi importantes, sont les *agents de l'analyse* (c'est-à-dire les personnes, groupes et autres acteurs performant une analyse), le *contexte de l'analyse* (par exemple, le contexte temporel) ou encore les *modalités* déterminant une analyse (comme les intentions, les consignes et autres obligations déterminant une analyse, le niveau des connaissances et du savoir-faire, etc.). Dans notre travail, nous avons privilégié de systématiser autant que se peut l'approche des deux premières dimensions (objets d'analyse et procédures d'analyse) tout en étant conscient de l'importance des autres dimensions citées ci-dessus à la fois pour une vision appropriée de toute activité de l'analyse et pour l'enrichissement du Studio ASA vers un véritable « atelier d'ingénierie et de gestion de connaissances ».

2. D'un *thesaurus* à facette* qui prédétermine pour certains termes conceptuels (ou concepts) du méta-lexique des listes de valeurs sous forme de *descripteurs* que peuvent utiliser les analystes de corpus audiovisuels. L'utilisation d'un thesaurus est particulièrement utile pour obtenir une certaine homogénéité des résultats d'analyse, soit pour des données *a priori* déjà standardisées ou fortement structurées (comme c'est le cas, par exemple, pour l'indexation de pays, de départements, régions ou provinces administratives d'un pays, de périodes chronologiques, de langues, etc.), soit pour des données pour lesquelles on souhaite obtenir des résultats « contrôlés », c'est-à-dire des résultats d'analyse obtenus grâce à un choix de termes ou de « descripteurs » imposés à l'analyste.

3. De *bibliothèques* de *briques** de modèles (appelées *séquences** ou *schémas**) déjà existants pouvant être réutilisés tels quels ou moyennant certaines modifications pour un projet d'analyse concret. Ainsi, existe-t-il des *séquences de localisation géographique* et de *localisation temporelle* d'un objet d'analyse, des séquences de l'*analyse du discours*, des séquences de l'*analyse du plan visuel* ou *acoustique*, des schémas définissant l'*objet d'analyse**, des schémas définissant les *procédures de description contrôlée* ou *libre* d'un objet d'analyse, etc. Toutes ces briques peuvent être reprises telles quelles ou moyennant certaines modifications dans le cycle de développement de modèles de description pour un domaine de connaissances/d'expertise donné.

Méta-lexiques conceptuels, thesaurus et bibliothèques de briques de modèles de description permettent de définir le *métalangage de description** approprié pour analyser les corpus audiovisuels documentant un domaine de connaissances/d'expertise donné, c'est-à-dire l'univers du discours d'une archive audiovisuelle donnée.

Un *métalangage de description** se manifeste sous forme d'une bibliothèque dynamique de modèles de description qui diffèrent les unes des autres, comme nous l'avons déjà vu, du point de vue de l'*objet d'analyse**. Ainsi :

1. Une classe de modèles de description est réservée à la *production de la méta-description* elle-même, c'est-à-dire à l'explicitation du contenu, des objectifs, des auteurs, du public visé, etc. d'une analyse concrète (voir pour cela [CHE 11a]).

2. Une classe de modèles de description est réservée à l'explicitation des *données paratextuelles* des objets audiovisuels soumis à l'analyse : intitulé de l'objet, auteur(s), genre, langue, propriété intellectuelle, etc. (voir [CHE 11a]).

3. Une classe importante de modèles est consacrée à l'analyse du *contenu* lui-même véhiculé par un corpus audiovisuel (celle-ci fait l'objet principal de ce livre).

4. Une classe tout aussi importante de modèle est consacrée plus particulièrement à la *mise en scène audiovisuelle du contenu* véhiculé par un corpus audiovisuel (il s'agit des modèles permettant l'analyse systématique des plans visuels et acoustiques et qui sont présentés plus en détail dans [DED 11]).

5. Une classe de modèles est réservée à l'*adaptation contextuelle* d'un corpus audiovisuel, à sa préparation de son usage dans des contextes visés par un projet d'analyse (il s'agit ici des modèles permettant à l'analyste de suggérer des usages spécifiques, de produire des références explicatives, de produire des versions bi ou multilingues, et ainsi de suite, voir [SAC 11]).

Parmi ces différentes classes de modèles de description, certaines peuvent être considérées comme relativement communes à tout travail d'analyse – peu importe la nature de l'analyse et peu importe aussi le type du texte audiovisuel soumis à l'analyse.

Il s'agit notamment des modèles relatifs à la description du plan sonore et visuel ainsi que des modèles servant à l'annotation d'un texte audiovisuel afin, par exemple, de l'enrichir à l'aide de références et de commentaires, de montrer son intérêt pour tel ou tel contexte d'utilisation. « Relativement commun » veut bien dire que ces modèles peuvent être adaptés aux exigences propres à tel ou tel type de travail d'analyse mais, en somme, ces modifications ne concernent que le niveau de sophistication voulue du travail d'analyse.

Une catégorie de modèles cependant dépend totalement du domaine de connaissances qui est documenté par le corpus des textes audiovisuels soumis à

l'analyse. C'est la classe des modèles thématiques qui permettent à l'analyste de décrire plus particulièrement le contenu (le ou les sujets, le propos, le message, etc.) d'un texte audiovisuel, d'un corpus de textes audiovisuels ou encore d'une partie spécifique d'un texte audiovisuel. Ces modèles sont essentiels et incontournables pour la description et l'interprétation de l'univers du discours d'une archive. Il n'est donc pas surprenant si nous affirmons qu'ils constituent le « cœur » de nos recherches et expérimentations – la partie restante de ce livre en témoignera.

Chapitre 3

Un exemple concret d'un modèle de description du contenu audiovisuel

3.1. Introduction

Dans le premier chapitre (section 1.6), nous avons esquissé rapidement le cadre théorique de référence qui nous guide dans l'élaboration des ressources métalinguistiques nécessaires pour mettre en place les *modèles de description** dont un analyste a besoin pour procéder à la description du contenu d'un corpus audiovisuel. La question qui se pose maintenant est de savoir comment ce modèle théorique de référence se reflète concrètement dans les dits modèles de description composant le *métalangage de description** de l'*univers de discours** d'une archive. Plus concrètement encore, comment ce cadre de référence s'exprime-t-il, dans tel ou tel formulaire interactif de travail faisant partie de l'interface du Studio ASA succinctement présenté dans le chapitre 2 (voir aussi [STO 11a]) ?

Dans ce chapitre, nous allons présenter et discuter brièvement un *modèle de description** du contenu d'un texte ou d'un corpus de textes audiovisuels. Dans le chapitre suivant (chapitre 4), nous en tirerons quelques conséquences plus générales,

pour l'approche développée ici de l'analyse du contenu de corpus ou d'archives entières de textes audiovisuels.

La section 3.2 présente très brièvement une *bibliothèque** de modèles de description pour l'analyse de corpus audiovisuels faisant partie d'une archive particulière. Dans la section 3.3 seront présentées et discutées succinctement les *séquences** qui composent un modèle de description. Une séquence est une *configuration thématiquement* et *fonctionnellement circonscrite* qui définit une *étape spécifique* de description dans le processus d'analyse d'un texte audiovisuel à l'aide d'un modèle de description.

Ensuite, dans la section 3.4, nous verrons que chaque séquence composant un modèle de description fait partie d'un type particulier de séquences répondant à un champ d'analyse spécifique. Enfin, la section 3.5 est réservée à une très rapide présentation de l'organisation interne canonique d'une séquence de description. Comme nous le verrons, chaque séquence est obligatoirement composée de deux types de *schémas de définition** – l'un définissant l'*objet d'analyse**, l'autre la *procédure d'analyse**.

3.2. La sélection du modèle approprié dans la bibliothèque de modèles de description du contenu audiovisuel

Pour commencer, nous allons discuter l'exemple concret d'un modèle de *description thématique** dont nous nous servons pour procéder à l'analyse du contenu de textes audiovisuels.

Prenons le cas de l'identification et de la description de sources audiovisuelles qui traitent des civilisations amérindiennes telles que la civilisation *aztèque*, la civilisation *inca*, la civilisation *chavin* ou encore la civilisation *mapuche*. Un corpus de textes audiovisuels traitant ce type de sujets fait partie des archives *A la Rencontre des Cultures* (ARC).

Les archives ARC (comme toutes les autres archives créées à l'aide de l'environnement de travail ASA exposé rapidement dans le deuxième chapitre de ce livre et d'une manière bien plus détaillée dans [STO 11a]) sont pourvues d'une *bibliothèque** de modèles de description du contenu audiovisuel. Une bibliothèque de modèles de description du contenu audiovisuel définit le métalangage de description de l'*univers du discours** d'une archive. Autrement dit, elle définit les *domaines de connaissances pertinents* (selon la politique, la ou les missions de l'archive) des textes audiovisuels qui font partie de l'archive.

Dans notre cas, la bibliothèque des modèles de description définit l'univers du discours des archives ARC. Comme le montre la figure 3.1, la bibliothèque des modèles de description du contenu audiovisuel ARC se présente sous la forme d'une

liste structurée d'intitulés de modèles tels que : *Patrimoine culturel immatériel*, *Formations culturelles par pays*, etc. Certains intitulés renvoient à des *collections de modèles*, d'autres à un *modèle spécifique*.

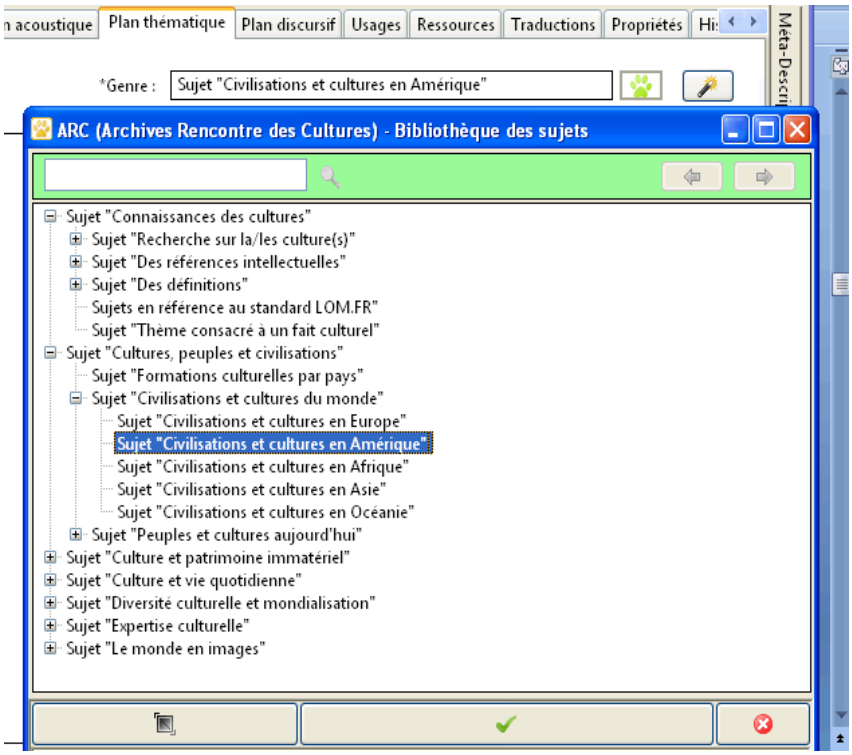

Figure 3.1. *Partie de la bibliothèque des modèles de description thématique de l'univers du discours des archives ARC – A la Rencontre des Cultures*

La figure 3.1 nous montre que la bibliothèque des modèles de description ARC contient un modèle spécifique qui s'intitule *Sujet « Civilisations et cultures en Amérique »*. Le modèle en question fait partie d'une *collection spécialisée* de modèles de description thématique de corpus audiovisuels traitant des civilisations dans le monde et de leurs formations culturelles.

Bien entendu, la collection de ces modèles peut être reprise – *telle quelle*, *partiellement* ou moyennant certaines *modifications/adaptations* – dans des bibliothèques de modèles de description définissant l'*univers de discours* d'autres archives audiovisuelles (numériques).

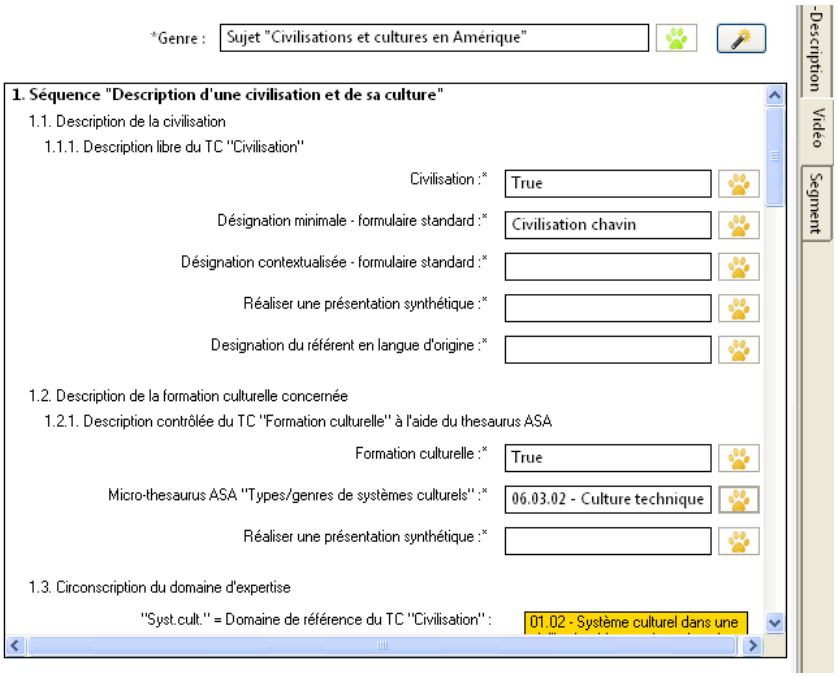

Figure 3.2. *La séquence de description réservée au domaine référentiel thématisé dans un texte ou un corpus de textes audiovisuels*

Ainsi, dans le cadre du projet ASA-SHS, la collection *Civilisations et cultures du monde* est reprise telle quelle mais accompagnée de simplifications fonctionnelles (notamment : suppression de la partie fonctionnelle « analyse du discours », voir infra) dans la bibliothèque de modèles des archives audiovisuelles FMSH-AAR dédiées à la diffusion des manifestations scientifiques de la Fondation Maison des Sciences de l'Homme (FMSH) à Paris. Elle est reprise partiellement dans la bibliothèque de modèles de description thématique définissant les archives audiovisuelles PCIA consacrées au patrimoine culturel immatériel des communautés andines (voir [LEG 11a]. « Partiellement » veut dire ici que seul le modèle relatif aux civilisations américaines (avec une restriction référentielle sur les civilisations andines) se trouve dans la bibliothèque des archives PCIA.

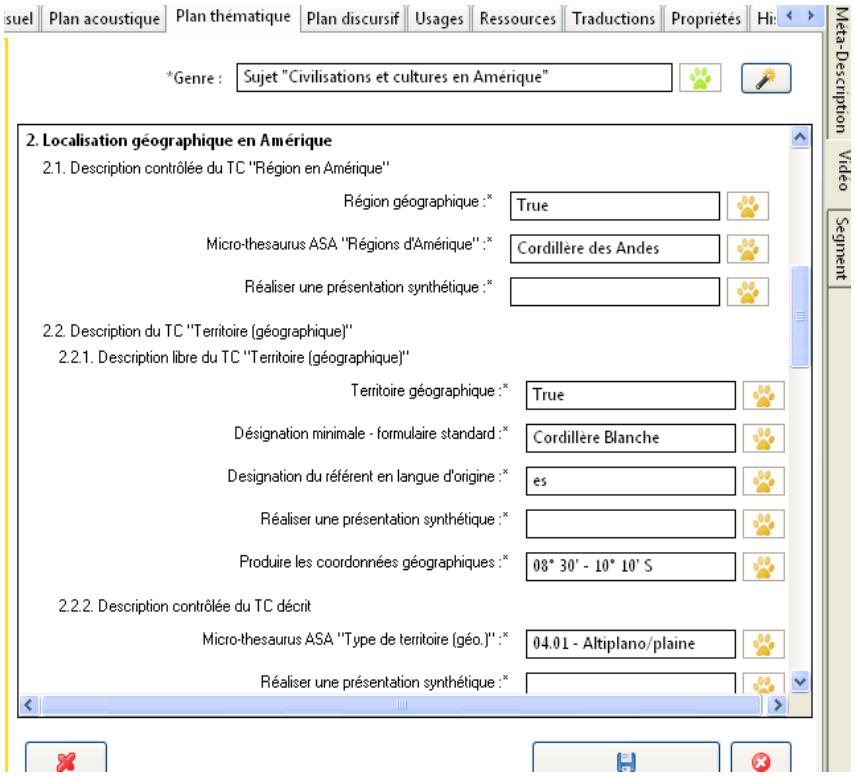

Figure 3.3. *La séquence de description réservée à la contextualisation géographique de l'objet référentiel d'analyse*

En tout cas, si l'analyste sélectionne l'intitulé *Sujet « Civilisations et cultures en Amérique »*, s'ouvre une interface sous forme d'un formulaire interactif de travail qui lui offre la possibilité de procéder à la description d'un texte audiovisuel selon la logique du modèle sélectionné.

3.3. Les séquences d'un modèle de description du contenu

En sélectionnant l'intitulé *Sujet « Civilisations et cultures en Amérique »*, l'analyste accède donc à un formulaire interactif qui forme son interface de travail afin de procéder à une analyse des corpus audiovisuels dont le contenu réfère à une civilisation – soit contemporaine soit historique – sur le sol américain.

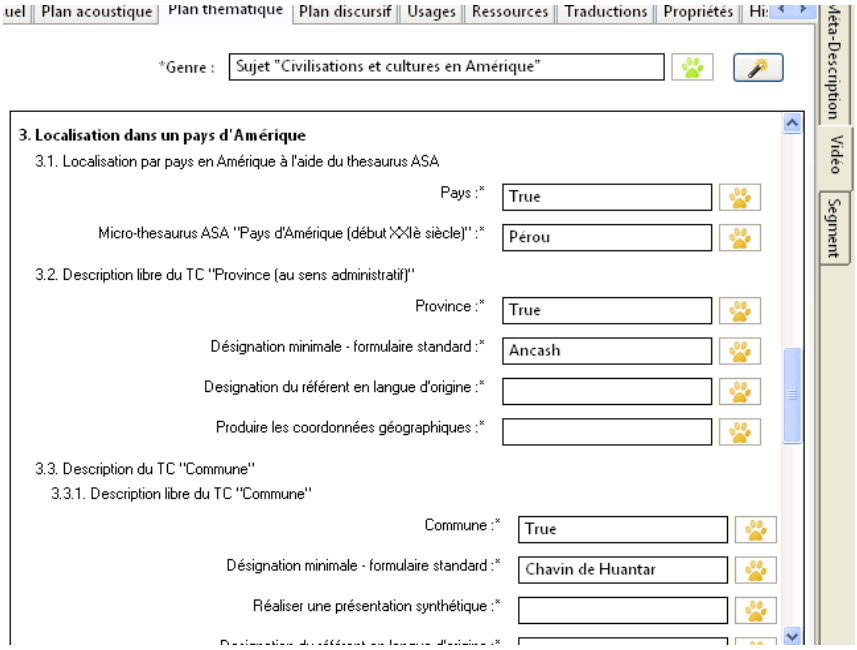

Figure 3.4. *La séquence de description réservée à la contextualisation par pays et territoire administratif de l'objet référentiel d'analyse*

Ainsi, la figure 3.2 montre une première *séquence** du modèle qui est fonctionnellement spécialisée dans l'analyse référentielle, c'est-à-dire dans l'analyse du domaine de connaissances qui fait l'objet d'une thématisation dans un texte audiovisuel ou un corpus de textes audiovisuels. Dans notre cas, la séquence porte sur deux *types d'objets** spécifiques du domaine de référence en question. Il s'agit du type d'objets désigné par le *terme conceptuel** [Civilisation] et du type d'objets exprimé par le *terme conceptuel** [Formation culturelle][30]. D'autres séquences du même type peuvent être plus élaborées, peuvent intégrer bien d'autres types d'objets. Notre séquence possède une structure interne simple et ne sert qu'à des analyses rudimentaires mais suffisantes dans le cadre, par exemple, d'un programme d'indexation thématique de base de grands corpus audiovisuels.

[30] Nous reviendrons encore plus loin sur la problématique des *termes conceptuels** qui constituent en effet le *vocabulaire* du métalangage ASA (voir chapitre 4 et la troisième partie de ce livre). Notons seulement ici que graphiquement, un terme conceptuel est mis entre les deux crochets comme suit : [Terme conceptuel]. La valeur du terme conceptuel (son instance, son occurrence ou encore son expression prédéfinie dans un thesaurus) s'exprime graphiquement comme suit : <Valeur du terme conceptuel> ou encore [Terme conceptuel : <Valeur du terme conceptuel>]

Les figures 3.3, 3.4 et 3.5 montrent trois séquences de description qui possèdent, toutes les trois, une même spécialisation fonctionnelle à savoir celle de localiser spatialement ou temporellement le domaine de référence thématisé dans un texte audiovisuel. Enfin, la figure 3.6 montre encore deux séquences fonctionnellement et thématiquement différentes. Il s'agit des deux séquences qui nous servent pour procéder, d'une part, à l'analyse de la mise en discours d'un domaine de connaissances thématisé dans un texte audiovisuel et, d'autre part, à l'explicitation du point de vue de l'analyste à propos du domaine de connaissances et/ou de sa mise en discours dans le texte analysé.

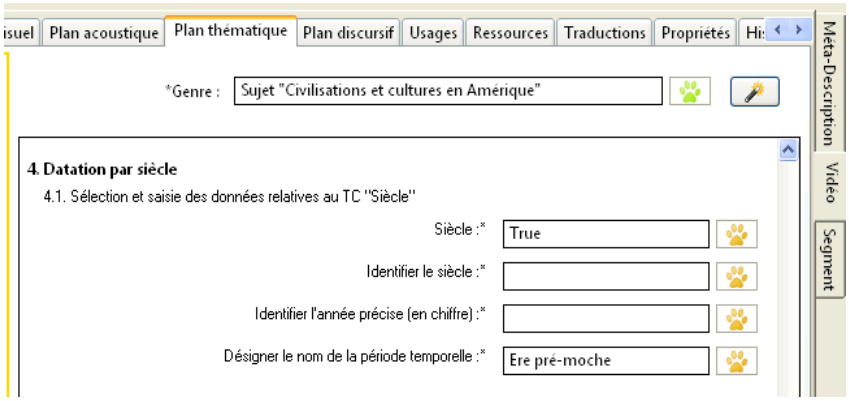

Figure 3.5. *La séquence de description réservée à la contextualisation temporelle de l'objet référentiel d'analyse*

Dans le chapitre 16, nous discuterons plus en détail de la constitution et de l'organisation d'une *bibliothèque de séquences** fonctionnellement spécialisées. Retenons ici les deux points suivants :

– tout *modèle de description** (du contenu audiovisuel) est composé d'un ou de plusieurs *séquences* ;*
– chaque séquence appartient à un type de séquences fonctionnellement spécialisées dans la réalisation d'une tâche spécifique d'analyse du contenu audiovisuel.

3.4. Champ de description et organisation séquentielle d'un formulaire d'analyse

En considérant les cinq figures (figure 3.2 à figure 3.6), on peut se rendre compte que le formulaire en question est structuré en quatre grands champs d'analyse :

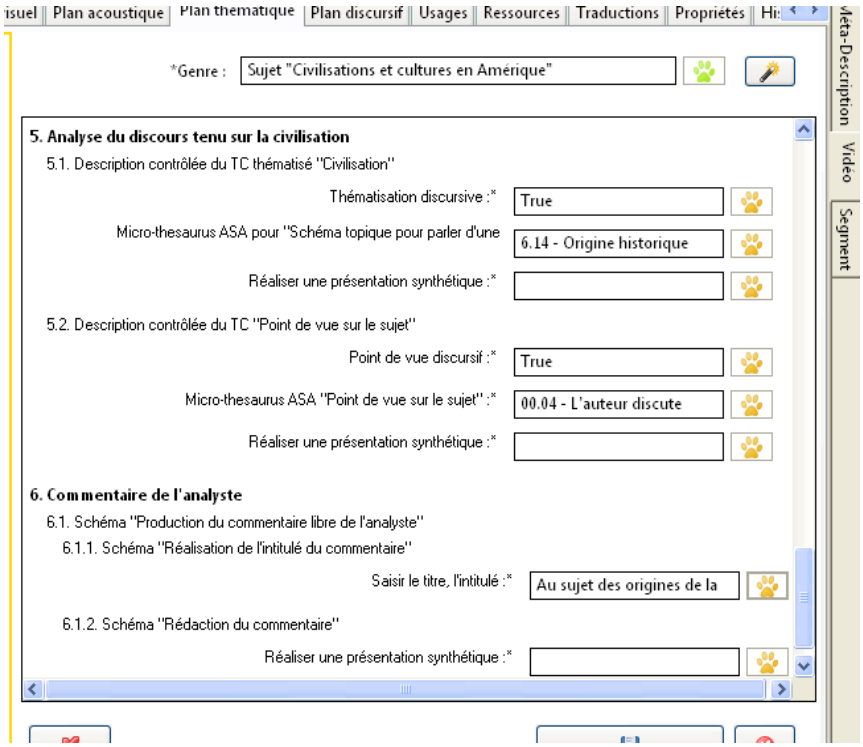

Figure 3.6. *La séquence de description réservée à l'analyse du discours tenu*

1) La séquence *Description d'une civilisation et de sa culture* (figure 3.2) correspond au *champ principal et nécessaire* de l'identification et de l'explicitation du domaine ou de l'objet de référence. D'une manière générale, cette séquence fait partie d'un *type spécifique de séquences* réservées à la tâche d'analyse qui s'intitule *Description de l'objet de connaissance thématisé*. Ainsi, la figure 3.2 montre la séquence qui est fonctionnellement spécialisée dans l'analyse référentielle et qui, thématiquement, porte sur deux types d'objets spécifiques du domaine de référence en question. Il s'agit du type d'objets désigné par le terme conceptuel [Civilisation] et du type d'objets exprimé par le terme conceptuel [Formation culturelle].

2) Les séquences *Localisation géographique en Amérique* (figure 3.3), *Localisation précise dans un pays en Amérique* (figure 3.4) et *Datation par siècle* (figure 3.5) correspondent au champ de la *localisation spatiale* (géographique) et *temporelle* (historique) du domaine ou de l'objet de référence traité. Elles font partie d'un type spécifique de séquences réservées à la tâche *Contextualisation de l'objet*

de connaissance thématisé (il s'agit de la tâche 1.2 dans la figure 4.4 ; voir chapitre 4.5).

3) La séquence *Analyse du discours tenu sur la civilisation* (figure 3.6) correspond au champ réservé à la description de certains *aspects énonciatifs* et de la *mise en discours* du domaine traité (focus d'intérêt, point de vue selon lequel l'objet est traité, etc.). Celle-ci fait partie d'un type de séquences spécifiques réservées à la tâche 2 (dans la figure 4.4 ; voir chapitre 4.5) *Analyse de la mise en discours de la topique*.

4) Et, enfin, la séquence *Commentaire de l'analyste* (figure 3.6) correspond au champ méta-textuel permettant à l'analyste de commenter soit son propre travail soit le traitement de l'objet dans la vidéo ou le segment soumis à l'analyse. Elle fait de nouveau partie d'un type spécifique de séquences réservées à la tâche de l'explicitation du cadre de l'analyse de la topique et de ses objectifs (= tâche 4 dans la figure 4.4 ; voir chapitre 4.5).

Soulignons le fait que seul le travail sur le premier champ, c'est-à-dire sur celui de l'identification et de l'explicitation de l'objet référentiel d'analyse, est *obligatoire* et *nécessairement présupposé* par le travail dans les trois autres champs, c'est-à-dire dans le champ de la localisation de l'objet référentiel, dans le champ de l'analyse de la mise en discours de l'objet référentiel thématisé et, enfin, dans le champ méta-textuel.

En somme et d'une manière simple, ces quatre champs d'analyse permettent de répondre aux questions de base nécessaires à toute analyse de contenu, questions que nous avons déjà discuté au premier chapitre (section 6) :

1. *Quoi ?* : De quoi parle-t-on ? *Où ?* et *Quand ? :* L'objet dont on parle, est localisé où et à quel moment/époque ?
2. *Comment ? :* Comment est traité et développé l'objet dans le texte audiovisuel ?
3. *Sous quelle forme ? :* quelles sont les expressions verbales et/ou audiovisuelles utilisées pour « parler » de l'objet ?
4. *Pourquoi* ? : Pourquoi l'ai-je traité ? pour quels motifs ? en ayant quelles intentions en tête ?

Un modèle de description thématique peut être complété, bien sûr, par d'autres champs fonctionnels d'analyse qui soit enrichissent les quatre champs que nous venons de citer, soit ouvrent l'analyse vers de nouvelles problématiques non prises en considération par nos quatre champs. Quoi qu'il en soit, tout formulaire interactif de travail composant l'interface entre l'analyste et un modèle de description doit tenir compte des quatre champs qui caractérisent le domaine de l'*expertise* du contenu d'un texte ou d'un corpus de textes audiovisuels :

1. les *objets du domaine de connaissances* auquel réfère le texte et qui constituent sa *structure topique* ;
2. la *mise en discours* et le *développement discursif* des objets de connaissance ;
3. *l'expression verbale, audiovisuelle ou* autre (gestuelle, spatiale, etc.) des objets de connaissance thématisés et développés sous forme de discours dans un texte ;
4. le *point de vue de l'analyste* (son intérêt, sa compréhension, etc.) soit relatif aux objets de connaissance thématisés dans le texte, soit relatif à la mise en discours des objets de connaissance, soit encore relatif aux deux.

En intégrant aussi systématiquement que possible ces quatre champs d'analyse dans un modèle de description thématique d'un corpus audiovisuel, nous pouvons ainsi tenir compte d'une manière plus adéquate du fait que toute *thématisation d'un objet de connaissance* dans un texte, s'effectue selon un *plan de médiatisation discursive*, c'est-à-dire selon le *point de vue* de l'auteur (lato sensu) et selon la capacité, la compétence de l'auteur à *hiérarchiser*, à *formuler et à développer linéairement* son point de vue sur l'objet thématisé dans un discours ou un texte.

3.5. Le plan des schémas de définition et les procédures de description

La séquence d'un modèle de description est obligatoirement composée d'un ou de plusieurs *schémas de définition**. Un schéma de définition est une *configuration* de termes conceptuels qui se différencie :

- soit en des schémas définissant les *objets d'analyse** (comme [Civilisation] ou [Formation culturelle]) ;
- soit en des schémas définissant les *activités d'analyse** (comme [Désignation minimale (d'un objet], [Illustration (d'un objet], etc.).

Ainsi, l'acte de description ou d'analyse est réglé au niveau des schémas de définition sous forme d' « instructions » ou de « consignes » implicites de type :

a) *d'abord*, sélectionner le type ou les types d'objets à analyse,
b) *ensuite* produire l'analyse en optant pour telle ou telle activité.

Précisons qu'une *tâche* particulière de description (description référentielle, description du cadre spatiotemporel, description de la mise en discours, etc.) s'effectue toujours selon des *procédures de description* précises qui intègrent une ou plusieurs activités concrètes de description portant sur un ou une sélection d'objets d'analyse. Nous distinguons plus particulièrement entre deux procédures centrales :

1. la *description libre** (c'est l'analyste qui détermine les valeurs d'un terme conceptuel) ;
2. et la *description contrôlée** (l'analyste s'appuie sur un thesaurus pour déterminer la ou les valeurs d'un terme conceptuel).

Ainsi, comme la figure 3.2 nous le montre, la *description référentielle** de l'objet de connaissance [Civilisation] se fait à l'aide d'une procédure de description libre. C'est l'analyste qui saisit l'expression nominale <Civilisation chavin>[31] pour désigner la valeur spécifique du terme conceptuel [Civilisation] dans le cadre de son analyse d'un texte audiovisuel thématisant « quelque chose » à propos de cette civilisation précolombienne d'il y a quelque 2500 ans.

Par contre, la description référentielle de l'objet de connaissance [Formation culturelle (de la civilisation identifiée)] est réalisée à l'aide d'une procédure de description contrôlée, c'est-à-dire à l'aide d'un thesaurus spécialisée dans le domaine de l'expertise culturelle. Pour assigner une valeur spécifique au terme conceptuel [Formation culturelle (concernée dans la civilisation chavin)], l'analyste utilise (dans notre cas) le micro-thesaurus ASA « Types/genres de systèmes culturels » et y sélectionne le descripteur <Culture technique> signalant par là qu'il est question dans le texte audiovisuel qu'il est en train d'analyser, de la culture technique de la civilisation *chavin*. Nous discuterons plus en détail des procédures de la description libres et contrôlées dans les chapitres 9 et 10.

[31] Rappelons que la *valeur* du terme conceptuel dans le texte audiovisuel soumis à l'analyse (c'est-à-dire l'*instance*, l'*occurrence* ou encore l'*expression prédéfinie* du terme conceptuel dans un thesaurus) s'exprime graphiquement comme suit : <Valeur du terme conceptuel> ou encore [Terme conceptuel : <Valeur du terme conceptuel>]

Chapitre 4

Modèle de description et travail d'analyse

4.1. Introduction

Suite à l'exemple concret d'un *modèle de description** du contenu d'un corpus de *textes audiovisuels**, présenté dans le chapitre 3, nous tirerons, dans ce chapitre, des conséquences plus générales relatives, d'une part, à la structure hiérarchique et syntagmatique de tout modèle de description et, d'autre part, relatives au travail d'analyse consacré à l'objet textuel.

Dans la section 4.2, nous discuterons très rapidement les 6 principaux constituants d'un modèle de description.

Dans la section 4.3, nous traiterons d'une manière générale l'ordre syntagmatique canonique d'un modèle de description qui détermine en même temps tout scénario concret d'analyse. Par « ordre syntagmatique » nous entendons, la structure de succession des principales séquences selon lesquelles se déroule le processus de l'analyse d'un texte audiovisuel.

Dans la section 4.4, nous discuterons plus en détail la distinction, qui nous semble capitale, entre *quatre niveaux hiérarchiques* distincts pour pouvoir traiter et modéliser adéquatement le travail d'analyse. Il s'agit de la distinction entre :

i) le *type* d'analyse ;

ii) la *tâche* d'analyse composant un type d'analyse ;

iii) la *procédure* de description à l'aide de laquelle une tâche d'analyse est réalisée ; et enfin ;
iv) l'*activité* ou les *activités* de descriptions concrètes qui composent une procédure de description.

Enfin, dans la section 4.5, nous rendrons compte plus particulièrement du type d'analyse appelé analyse du contenu (d'un texte ou d'un corpus de textes audiovisuels) – type d'analyse qui nous sert de référence dans ce livre.

4.2. L'organisation structurale d'un modèle de description du contenu audiovisuel

Un modèle de description du contenu audiovisuel possède une structure, une *organisation structurale canonique* qui conditionne et facilite la construction de nouveaux modèles ainsi que les modifications locales de modèles déjà existants. Cette organisation canonique est représentée dans la figure 4.1 et se base sur 6 constituants centraux :

1) Tout modèle de description thématique se compose d'un ou d'un petit ensemble de *champs d'analyse* thématiquement et fonctionnellement spécialisés dont font partie, par exemple, le champ de l'identification et de l'explicitation d'un objet de connaissance thématisé dans le texte audiovisuel, le champ de la contextualisation géographique et historique de l'objet thématisé, le champ de la mise en discours de l'objet thématisé, etc. (pour plus d'informations, voir chapitre 3).

2) Un modèle de description est composé de *termes conceptuels* (de concepts) faisant partie du vocabulaire d'un *méta-lexique de termes conceptuels* qui forme un composant essentiel du système métalinguistique sur lequel repose *in fine* tout modèle de description (dans notre cas, il s'agit du système métalinguistique baptisé « ASA », acronyme pour *Atelier de Sémiotique Audiovisuelle*). Nous distinguons entre deux méta-lexiques (deux vocabulaires conceptuels) : a) le méta-lexique réunissant les termes conceptuels qui représentent les objets d'analyse de l'*univers du discours** ASA et b) le méta-lexique réunissant les termes conceptuels qui représentent les procédures d'analyse de l'objet textuel[32].

3) Un modèle de description est une *structure hiérarchique* à l'intérieur de laquelle on peut distinguer différents niveaux de regroupement et d'intégration de termes conceptuels (dans notre approche, nous distinguons deux niveaux – le niveau des *schémas* de définition* des termes conceptuels et le niveau des *séquences* de description** des termes conceptuels : séquence et schéma sont triplement définis

[32] Pour une discussion plus approfondie du méta-lexique des termes conceptuels ASA, voir les chapitres 12, 13 et 14 de ce livre.

comme des unités *thématiques* (elles « s'occupent » d'un aspect spécifique de l'objet d'analyse), *fonctionnelles* (elles font partie de l'un ou de l'autre des champs d'analyse identifiés dans le chapitre 3) et, enfin, *hiérarchiques* (les deux unités font partie de deux niveaux d'intégration distincts)[33].

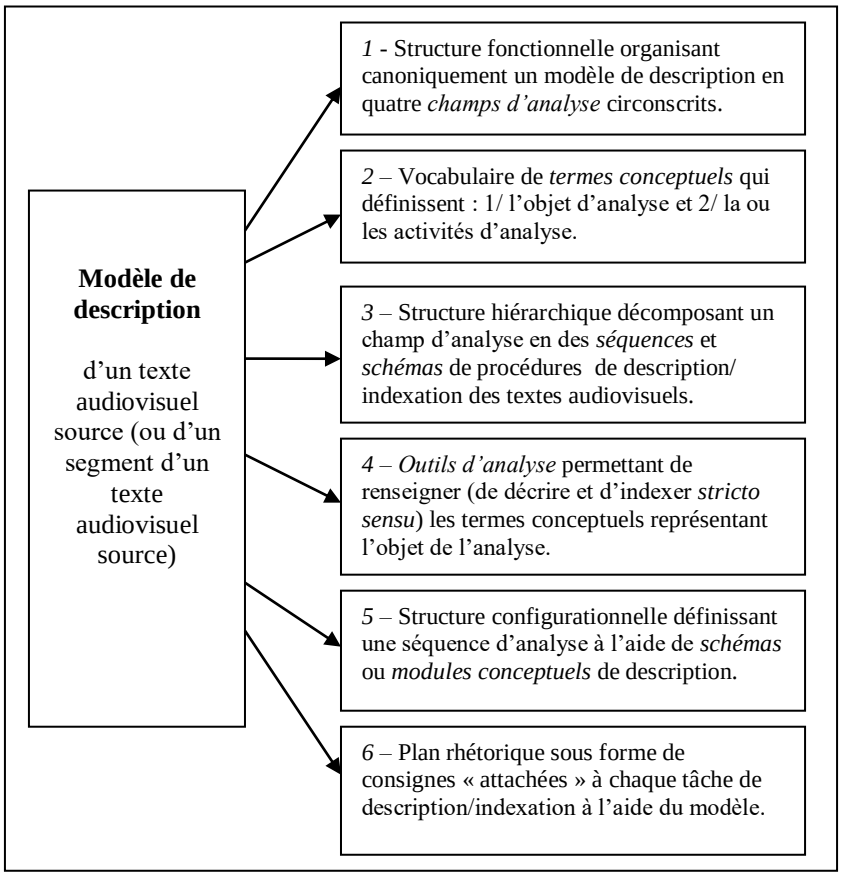

Figure 4.1. *Les constituants principaux d'un modèle de description d'un corpus de textes audiovisuels*

4) Un modèle de description est pourvu d'*outils d'analyse* de chaque terme conceptuel et/ou de chaque regroupement de termes parmi lesquels nous comptons, par exemple :

[33] Pour une discussion plus approfondie des deux unités séquence et schéma, voir les chapitres 15 et 16 de ce livre.

– les *outils d'indexation à proprement parler* d'un objet de connaissance thématisé dans un texte audiovisuel sous forme, par exemple, de désignation, de sa dénomination, de son identification sous forme de mots-clés, et ainsi de suite ;
– les *outils d'annotation textuelle* d'un objet analysé ;
– les *outils* – très importants – de *référencement* d'un objet à analyser à l'aide de thesaurus, terminologies ou encore standards ;
– les *outils* de *référencement géographique* et *chronologique* d'un objet de connaissance thématisé dans un texte audiovisuel.

Ces outils sont réunis sous forme d'une *bibliothèque de schémas d'indexation** que nous présenterons plus en détail dans le chapitre 14.

5) Les termes conceptuels définissant d'une part l'objet d'analyse et d'autre part les procédures d'analyse forment une *configuration*, c'est-à-dire une structure interne de *positionnement réciproque* et d'*intégration hiérarchique* en un « tout » qui est le *modèle de description**.

6) En fin de compte, un modèle de description est pourvu d'un ensemble de consignes qui forment son plan rhétorique et qui aident et guident l'analyste dans sa tâche d'analyse du contenu d'un texte audiovisuel.

4.3. L'ordre syntagmatique canonique d'un formulaire de description

Un formulaire de description d'un texte audiovisuel (ou d'une partie spécifique de celui-ci) débute toujours avec la séquence (parfois : les séquences) de description réservée à la *description référentielle** de l'objet de connaissance thématisé dans le texte audiovisuel (voir figure 4.2). Dans notre exemple développé dans le chapitre 3, l'analyste est invité tout d'abord à décrire la civilisation dont il est question dans le texte audiovisuel soumis à l'analyse, et ensuite à préciser de quel type de formation culturelle il s'agit. Comme le montre la figure 3.2 (chapitre 3), il est question – dans notre cas concret – de la civilisation *chavin* (une civilisation amérindienne ayant vécu son apothéose dans les derniers siècles avant J.C.) et de sa très remarquable culture matérielle. Il s'agit ici d'un modèle de description relativement simple mais qui représente l'avantage énorme de reposer sur un métalangage (une ontologie descriptive) explicite qui permet toutes sortes d'exploitations pratiques qui ont été discutées dans [STO 11a, STO 11b]. Plus loin, nous verrons que ce modèle de base de description référentielle peut faire place, si nécessaire ou souhaité, à des modèles de description beaucoup plus complexes mais exigeant également des compétences d'analyse très spécialisées et un temps d'analyse plus long.

La séquence de description référentielle est toujours suivie de la ou des séquences réservées à la description du cadre spatiotemporel du domaine de connaissances thématisé dans un texte audiovisuel (voir figure 4.2). En règle

générale, la séquence (les séquences) de la localisation spatiale (géographique, géopolitique, etc.) précède(nt) la séquences (les séquences) de la localisation temporelle (historique, etc.).

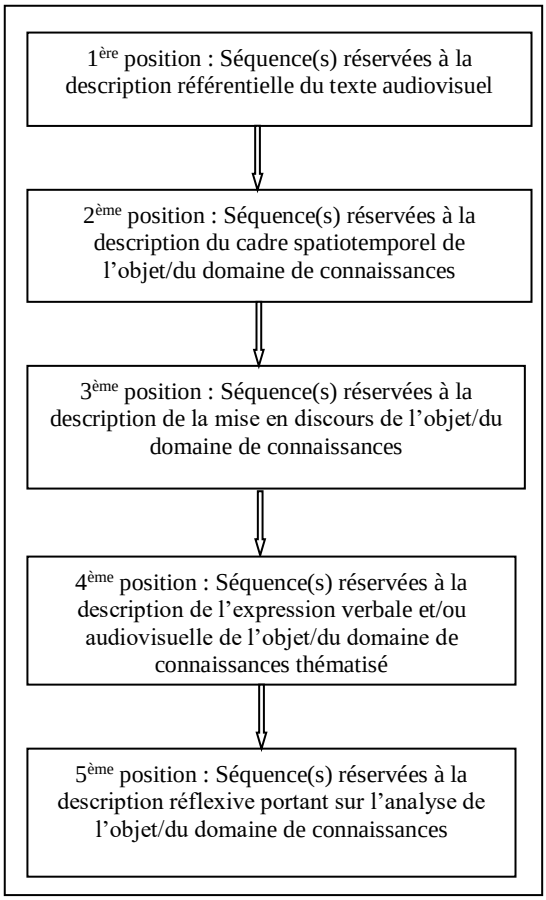

Figure 4.2. *Structure syntagmatique canonique d'un modèle d'analyse de corpus audiovisuels*

Pour reprendre notre exemple, développé dans le chapitre 3, celui-ci comporte en effet deux séquences de localisation géographique. La première séquence (voir chapitre 3, figure 3.3) invite l'analyste à produire des informations relatives à la localisation globale (de la civilisation *chavin*), c'est-à-dire relative à la région concernée sur le continent américain. Il s'agit, dans notre cas concret, de la <Cordillère des Andes> en <Amérique du Sud>. Si cela est pertinent, l'analyste peut préciser un territoire géographique plus circonscrit de la région identifiée. Dans

notre cas, l'analyste précise qu'il s'agit de la <Cordillère Blanche>. Il peut, en plus, procéder à une classification du territoire géographique en question à l'aide d'un thesaurus spécialisé. Dans notre cas, il précise que le territoire de la Cordillère Blanche est un territoire de type <Altiplano>.

Une fois les informations relatives à la localisation géographique produites, l'analyste peut fournir également des informations de nature géopolitique et administrative : sur le pays concerné, sur la province (le département, le canton, etc. selon la terminologie utilisée), sur la commune, etc. (voir chapitre 3 figure 3.4). Dans notre cas (voir chapitre 3 figure 3.4), l'analyste ajoute que le cadre administrativo-territorial pertinent pour localiser le sujet en question est le village <Chavin de Huantar> situé dans la province d'<Ancash> au <Pérou>.

Soulignons que, l'analyste est libre d'ajouter les informations nécessaires pour une localisation spatiale – géographique et/ou administrativo-territoriale. Pour enregistrer un modèle de description dans la *base de descriptions* du système ASA [LEM 11a], la seule séquence à renseigner d'une manière obligatoire est celle réservée à la description référentielle.

La localisation temporelle, de son côté peut être réduite à une séquence aussi simple que celle représentée dans la figure 3.5 (chapitre 3). Cette séquence invite l'analyste à procéder à une localisation temporelle (de la *civilisation chavin*) à l'échelle de siècles. Toujours dans notre cas, il s'agit d'une période se situant à peu près entre le 7ème et le 4ème siècle avant J.C. L'identification de ces siècles se fait, comme le montre la figure 3.5 (chapitre 3), via un micro-thesaurus spécialisé dans l'énumération des périodes de l'histoire (humaine) en siècles. Si les siècles identifiés pour localiser un objet de connaissance, forment une époque reconnue, alors l'analyste peut également saisir le nom de cette époque (voir chapitre 3, figure 3.5). Toujours est-il qu'il existe toute une bibliothèque de séquences de localisation temporelle et historique plus développée que l'analyste peut utiliser s'il en a besoin (nous y reviendrons plus loin, dans le chapitre 5).

Après les séquences réservées à la description référentielle et au cadre spatiotemporel, on trouve, comme le montre la figure 4.2, en troisième position la ou les séquences réservées à l'*analyse du discours* du domaine de connaissances thématisé. Pour reprendre notre exemple développé dans le chapitre, cette analyse est concentrée sur les deux questions de la thématisation discursive et du point de vue auctorial qui prévaut dans le traitement de l'objet de connaissance thématisé (voir figure chapitre 3, 3.6). Ici, l'analyste précise que ce sont (surtout) les origines historiques de la culture technique de la civilisation *chavin* dans la période comprise entre le 7ème et le 4ème siècle avant J.C. qui sont traitées par le texte audiovisuel soumis à l'analyse. Il ajoute également que cette question des origines historiques est traitée par l'auteur à la fois sous forme de références à ses propres travaux (de recherche) ainsi qu'en référence aux travaux (de recherche) d'autres auteurs. Les *micro-thesaurus** spécialisés mis à la disposition de l'analyste pour expliciter la

mise en discours de la topique spécifique <culture technique de la civilisation chavin dans la période comprise entre le 7ème et le 4ème siècle> constituent un essai, une tentative de typer, de classifier à l'aide d'un ensemble de traits récurrents et stéréotypés[34] les façons à aborder et à traiter (c'est-à-dire à médiatiser) par et dans le discours les objets du domaine de connaissances auquel réfère le texte audiovisuel soumis à l'analyse. Nous y reviendrons dans le chapitre 15 consacré à la présentation du *thesaurus** ASA.

La quatrième position est occupée, comme le montre la figure 4.2, par la ou les séquences réservées à la description de la *mise en scène (de l'expression) audiovisuelle* et/ou *verbale* du sujet traité dans un texte audiovisuel. Cette option n'est pas réalisée dans le formulaire qui nous sert d'exemple, c'est-à-dire dans le formulaire servant d'interface pour décrire des textes audiovisuels thématisant le domaine des civilisations et de leurs cultures sur le sol américain.

Enfin, comme le montre la figure 4.2, en cinquième et dernière position, arrive(nt) la ou les séquences réservées à l'explicitation du point de vue de l'analyste, de son interprétation, de son expertise. Dans notre exemple (chapitre 3, figure 3.6), il s'agit d'une séquence toute simple qui se réduit à la possibilité offerte à l'analyste d'ajouter, à son analyse du domaine de connaissances <culture technique de la civilisation chavin>, des commentaires, des compléments d'informations, etc. Cette position syntagmatique peut être enrichie par toute une bibliothèque de séquences plus spécialisées permettant d'intégrer assez facilement les aspects réflexifs les plus variés relatifs à une analyse réalisée.

Un point important que nous n'avons pas encore évoqué ici concerne le fait de rapprocher l'analyse d'un sujet précis (comme celui de la culture technique de la civilisation *chavin* traitée dans tel ou tel segment d'une vidéo choisie) aux différents standards et normes utilisés dans le cadre des bibliothèques et archives numériques, de l'enseignement et de l'apprentissage à distance, etc.

Or, une possibilité (il y en a d'autres !) de rapprocher une analyse concrète du style *ASA* est de renseigner une *séquence réservée* à la saisie des informations exigées par une norme ou un standard (tel que le *Dublin Core*, *OAI*, *LOMFR*, etc.) ou par un schéma de description de référence (tels que le schéma des éléments sémantiques de la bibliothèque européenne *Europeana*). Nous avons développé des exemples de ce rapprochement – notamment avec la norme LOMFR (une adaptation du standard LOM au système éducatif français) – mais n'avons pas eu encore l'opportunité de mettre en œuvre plus systématiquement cette approche.

Répétons encore une fois le fait que la structure syntagmatique représentée dans la figure 4.2 possède un *caractère canonique* pour tous les modèles d'analyse de corpus audiovisuels. Certains modèles sont pourvus de séquences d'analyse pour

[34] Voir [STO 03] pour plus de détails sur ce point.

chacune des cinq étapes, d'autres ne sont pourvus que de séquences pour telle ou telle étape.

Les modèles les plus simples, et qui se rapprochent le plus des formulaires standards de description/indexation, sont ceux qui ne sont pourvus que des séquences de l'analyse référentielle et/ou du cadre spatiotemporel. D'autres formulaires, en revanche, possèdent une structure très élaborée, exploitant au maximum la richesse de l'approche sémiotique de l'analyse de corpus audiovisuels.

On voit ainsi que la structure canonique représentée dans la figure 4.2 reste déterminante dans la définition d'un modèle de description ainsi que pour son développement sous forme d'un formulaire interactif de travail.

4.4. Types d'analyse, tâches d'analyse, procédures de description et activités de description

Comme nous l'avons évoqué, on voit que, (voir chapitre 1 ainsi que [STO 11a, STO 11b]) d'une manière générale l'analyse de corpus audiovisuels composant, par exemple, le fonds d'une archive, constitue l'une des activités importantes – dans la chaîne des activités – qui caractérise le *processus de travail** de la constitution, diffusion et préservation de patrimoines culturels ou scientifiques.

L'analyse elle-même peut prendre la forme d'une « simple » indexation d'un corpus audiovisuel ; elle peut également prendre l'allure d'une véritable expertise (dans le contexte de l'archivistique, la référence ici est l'étude diplomatique de documents), voire se manifester sous forme d'un faire de telle ou telle école ou approche interprétative. Elle peut, autrement dit, poursuivre des objectifs très différents et être performée par des acteurs (c'est-à-dire des analystes) possédant des compétences très variables.

En reprenant notre discussion dans le chapitre 2.3 au sujet des différents types d'analyse développés et pratiqués dans le cadre de nos travaux de recherche sur les archives audiovisuelles, la figure 4.3 propose maintenant une vision globale et structurée de cette activité si centrale. Cette figure nous servira de référence tout au long de ce livre.

Le travail de l'analyse peut être appréhendé, comme le montre la figure 4.3, selon 4 niveaux principaux. Le niveau le plus général est celui qui définit le *type d'analyse*. La définition du type d'analyse influence la sélection des *tâches d'analyse* appropriées (= 2ème niveau représenté dans la figure 4.3) ; une tâche d'analyse, à son tour, est composée d'une ou de plusieurs *procédures d'analyse* (= troisième niveau dans la figure 4.3) ; enfin une procédure d'analyse est elle-même composée par une ou plusieurs activités d'analyse (= quatrième niveau dans la figure 4.3).

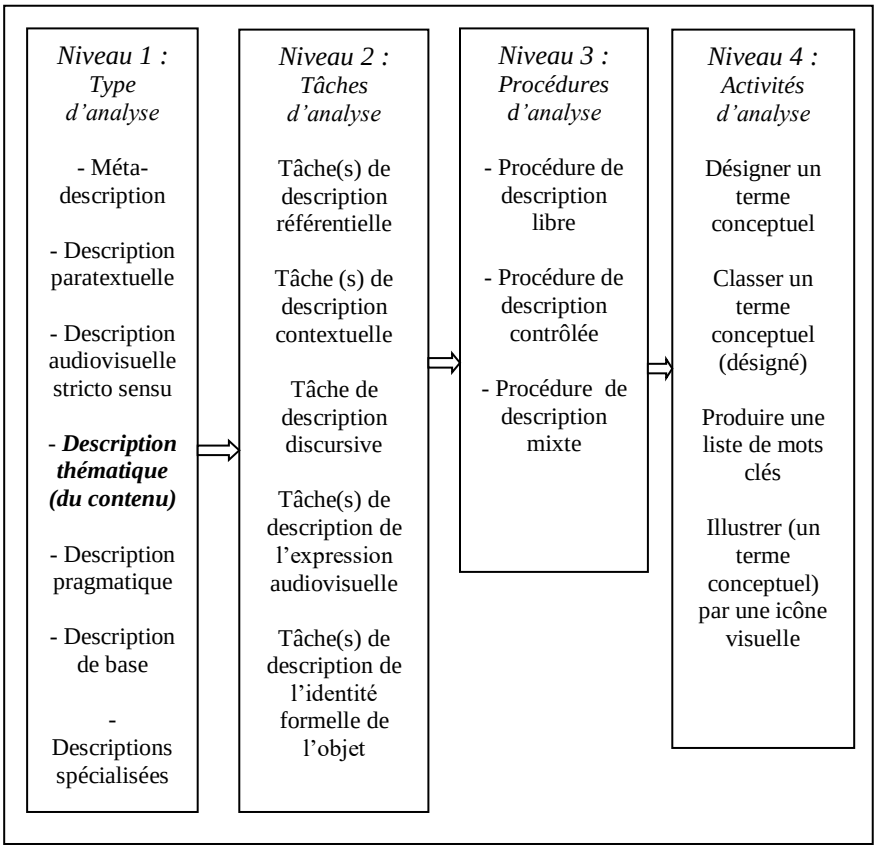

Figure 4.3. *Les quatre principaux niveaux hiérarchiques pour définir le travail d'analyse d'un corpus de textes (audiovisuels)*

Comme nous l'avons déjà explicité dans le chapitre 2, nous considérons qu'une analyse d'un texte ou d'un corpus de textes est toujours et nécessairement liée à un *certain point de vue* (de l'analyste) portant sur le texte à analyser, à un certain *intérêt* pour le texte ou encore à un *objectif*, un *but* que l'analyse se propose de satisfaire. Il est totalement vain de vouloir réduire tous ces points de vue, intérêts et objectifs en un seul cadre d'analyse. Mieux vaut laisser à l'analyste le choix de procéder à telle ou telle analyse, de réaliser selon ses souhaits une analyse d'un objet textuel. Sans vouloir prétendre que les ressources métalinguistiques ASA permettraient de couvrir tous les désirs ou besoins imaginables des analystes, nous pensons néanmoins qu'il est suffisamment souple pour répondre à une assez grande diversité de désirs ou de besoins.

Cela est une conséquence directe du fait que nous distinguons entre quatre niveaux hiérarchiques pour traiter le travail de l'analyse. Au premier niveau, nous distinguons, comme le montre la figure 4.3, et comme nous l'avons déjà brièvement vu au chapitre 2[35], plusieurs types d'analyse. Certains types sont concernés prioritairement par une strate particulière du texte audiovisuel. C'est le cas, par exemple, de la *description audiovisuelle* stricto sensu* qui s'intéresse, par exemple, prioritairement au repérage des plans visuels ou acoustiques, ou de la *description thématique** (du contenu) qui s'intéresse prioritairement à la thématisation d'un domaine de référence dans un texte ou un corpus de textes audiovisuels.

Certains autres types d'analyse se distinguent entre eux par rapport à la question : doit-on prendre en compte le texte audiovisuel dans son intégralité ou seulement tel ou tel passage, telle ou telle séquence. D'autres types d'analyse se distinguent encore par rapport au niveau de la spécialisation visée, c'est-à-dire par rapport au fait si une analyse doit ressembler davantage à une description de base (ne dépassant guère, par exemple, le niveau des 15 éléments de description composant le Dublin Core) ou si elle est entendue au sens d'une réelle expertise intellectuelle de l'objet textuel concerné.

Le choix pour tel ou tel type d'analyse (= niveau 1 dans la figure 4.3) entraîne obligatoirement une sélection des *tâches* plus spécifiques d'analyse (= niveau 2 dans la figure 4.3) à réaliser ainsi que, pour chaque tâche, une sélection des *procédures d'analyse* (= niveau 3 dans la figure 4.3) appropriées. Ainsi, le type d'analyse *description paratextuelle** vise uniquement – comme son intitulé l'exprime déjà – une analyse de l'objet audiovisuel du type *Dublin Core*, c'est-à-dire une analyse qui se réalise à l'aide d'une tâche qui s'intitule (figure 4.3) *description de l'identité formelle de l'objet*. Le modèle de description pour cette tâche est ainsi composé d'une « grosse » séquence pourvue d'objets d'analyse spécifiés selon le standard *Dublin Core* tels que [Auteur], [Titre], [Droit], [Description], etc. Pour renseigner chacun des objets d'analyse, une (parfois deux) procédure(s) de description leurs sont associés.

Comme nous le verrons encore dans les chapitres 9 et 10, nous distinguons notamment entre deux procédures de base : la *procédure de description libre** et la *procédure de description contrôlée* :*

– *description libre* signifie, que c'est l'analyste qui renseigne, selon ses habitudes et préférences (notamment linguistiques), un terme conceptuel (ou une configuration de termes conceptuels) représentant l'objet d'analyse ;
– par contre *description contrôlée* précise que l'analyste s'appuie sur un thesaurus pour réaliser son travail de description.

[35] Pour plus d'informations, merci de consulter [STO 11a].

Chaque procédure d'analyse est, à son tour, définie par une ou plusieurs *activités d'analyse**. Comme nous le verrons de nouveau au chapitre 14, nous disposons d'une riche ontologie descriptive d'activités d'analyse de l'objet textuel à l'aide de laquelle nous sommes capables de spécifier d'une manière très nuancée et fine les procédures et tâches d'analyse nécessaires pour réaliser un type d'analyse donné. Cela veut dire qu'une procédure d'analyse particulière telle que celle de la description libre peut varier énormément selon la *spécificité* de l'objet d'analyse et le niveau visé de précision d'une analyse. Ainsi, il est très différent de devoir désigner un objet conceptuel par une expression verbale, de l'illustrer par une icône visuelle ou encore, si pertinent, d'en indiquer les coordonnées géographiques, temporelles ou « conventionnelles » (sous forme d'une adresse, par exemple). Toutes ces éventualités font en sorte que le même type de procédure d'analyse peut varier énormément d'un type d'analyse à un autre, d'une tâche à une autre.

Retenons donc que le choix d'un type d'analyse entraîne une sélection de tâches pertinentes d'analyse ainsi que de procédures d'analyse appropriées à la fois par rapport à la spécificité de l'objet d'analyse et au cadre du type d'analyse choisi. Le système métalinguistique ASA prend en compte ces variations même si la réalisation informatique de l'actuel Studio ASA ne l'intègre que partiellement[36].

4.5. Tâches particulières de l'analyse du contenu d'un corpus audiovisuel

Voyons maintenant plus spécifiquement le type d'analyse qui nous intéresse le plus ici, à savoir la *description thématique* (du contenu)* d'un texte ou d'un corpus de textes audiovisuels. La figure 4.4 identifie les cinq *tâches principales d'analyse* du contenu d'un texte ou d'un corpus audiovisuel :

1. Analyse du domaine de référence (ou d'expertise) : analyse du contenu référentiel, de la structure topique *stricto sensu* (= tâche 1.1 dans la figure 4.4).
2. Analyse de la contextualisation référentielle (de la localisation spatiale et temporelle) de l'objet de connaissance thématisé dans le texte audiovisuel (= tâche 1.2 dans la figure 4.4).
3. Analyse du discours tenu sur l'objet de connaissance thématisé (= tâche 2 dans la figure 4.4).
4. Analyse de la mise en scène ou de l'expression audiovisuelle du discours tenu sur l'objet de connaissance thématisé (= tâche 3 dans la figure 4.4).

[36] L'actuel Atelier de Description ASA permet à l'analyste de choisir entre différents types d'analyse mais ce choix n'entraîne pas une sélection automatique de tâches et procédures pertinentes. L'analyste se doit de suivre un « simple » guide méthodologique pour réaliser d'une manière appropriée son analyse.

5. Commentaire métatextuel « libre » de l'analyste (= tâche 4 dans la figure 4.4).

Cela étant, l'analyste n'est pas obligé de réaliser chacune des tâches identifiées ci-dessus lors d'une analyse concrète du contenu d'un texte ou d'un corpus de textes audiovisuels. Il n'est pas obligé de décrire *in fine* la mise en discours d'une topique, d'un sujet spécifique ; il n'est pas non plus obligé de fournir des informations plus explicites sur son point de vue personnel à chaque fois qu'il procède à l'analyse d'une topique spécifique.

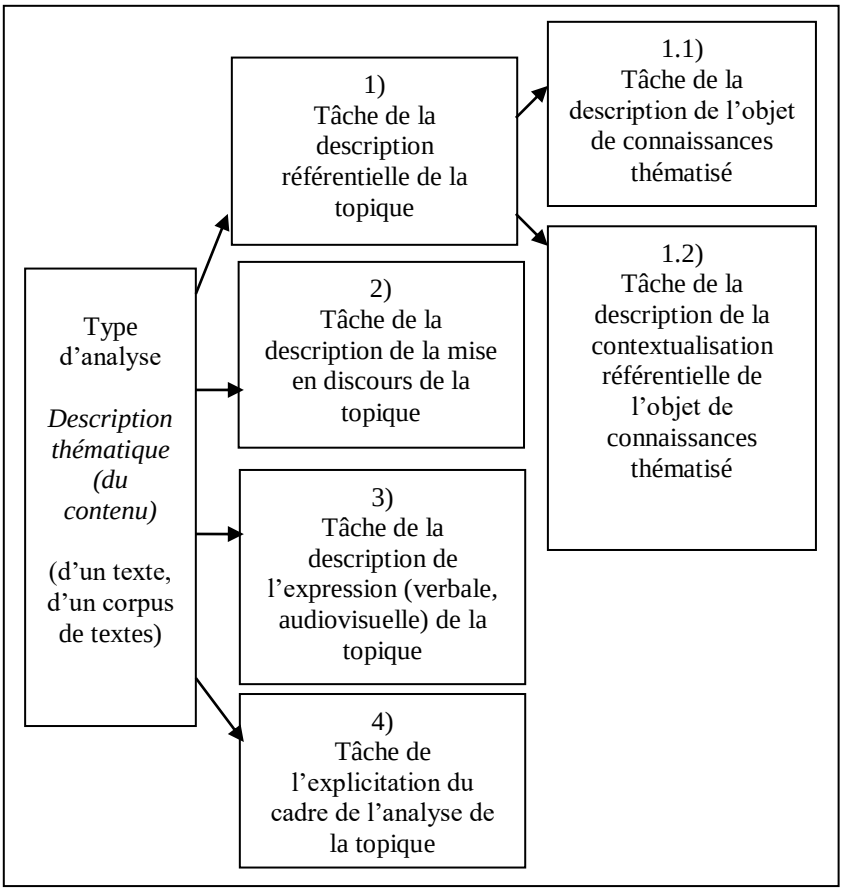

Figure 4.4. *Les principales tâches formant une analyse du contenu d'un texte ou d'un corpus de textes audiovisuels*

En effet, comme le suggère la figure 4.4, grâce aux différentes tâches identifiées ci-dessus, nous disposons d'une sorte de hiérarchie de spécialisation dans la description thématique couvrant aussi bien l'indexation thématique de base d'un domaine de connaissances que les diverses interprétations de ce dernier en passant par son traitement en tant qu'objet discursif et audiovisuel :

1. Dans le cadre de l'*indexation thématique de base*, l'analyste se contente de renseigner, d'identifier le ou les domaines de connaissances thématisé(s) dans un texte audiovisuel. L'identification ou le renseignement peut se faire, comme nous le verrons, soit d'une *manière libre* (c'est l'analyste qui désigne, dénomme, décrit, etc. verbalement, visuellement (à l'aide d'icones, par exemple) ou acoustiquement (à l'aide de *jingles*, par exemple) soit d'une *manière contrôlée* (à l'aide d'un thesaurus verbal ou visuel).

2. L'indexation thématique *de base* peut être complétée par une *analyse de localisation spatiale et/ou temporell*e du domaine de connaissances. Cette analyse se fait de nouveau à l'aide de procédures de description contrôlée et/ou libre.

3. Dans le cadre de l'analyse du discours à proprement parler, l'indexation thématique de base d'un domaine de connaissances est complétée par *une expertise du cadrage discursif* du domaine thématisé dans le texte audiovisuel et du *développement* du domaine thématisé dans le texte audiovisuel.

4. D'une manière encore plus poussée, l'analyse de la mise en discours peut être complétée par une analyse de l'*expression verbale* de la topique (par exemple, de la terminologie, des registres sociolinguistiques, etc. dans le cadre de vidéos contenant des témoignages, des entretiens, etc.) ; elle peut également être complétée par une analyse de l'*expression audiovisuelle* de la topique (par exemple, lorsque la thématisation d'un domaine de connaissances (d'un lieu, d'une personne, d'une activité, etc.) se fait essentiellement à l'aide de techniques visuelles (cadrage visuel, mouvement de caméras, plan visuel, etc., ou sonores).

5. Enfin, dans le cadre d'une expertise professionnelle ou d'une interprétation personnelle, l'indexation thématique peut encore être enrichie par des annotations (des commentaires, etc.) explicatives.

La possibilité offerte à l'analyste de choisir entre ces cinq approches (ou une combinaison de ces cinq approches) permet de tenir compte aussi bien des *intérêts spécifiques* que des *contraintes pragmatiques* qui influencent obligatoirement le travail de l'analyse du contenu d'un texte ou d'un corpus de textes audiovisuels (temps, ressources humaines, etc.).

4.6. Remarques conclusives

Dans la partie suivante de ce livre (*Partie II : Les tâches d'analyse d'un corpus audiovisuel*), nous allons considérer plus en avant certaines des tâches identifiées dans la figure 5.4 qui sont particulièrement importantes pour toute analyse du contenu. Ainsi nous discuterons plus en détail :

– la *description référentielle** d'un objet de connaissance (*Chapitre 6*) ;
– l'analyse de la *contextualisation référentielle* d'un objet de connaissance (*Chapitre 7*).

Nous ne discuterons, en revanche, ni de l'analyse de la mise en scène audiovisuelle ni de celle de l'expression verbale de l'objet de connaissance.

Dans la troisième partie de ce livre, nous discuterons plus spécifiquement des *procédures de description*. Enfin, la quatrième partie sera consacrée à la discussion d'une part des *objets d'analyse* (c'est-à-dire des objets dont « s'occupe » un travail d'analyse concret) et d'autre part des *activités et outils spécifiques de l'analyse*.

Deuxième partie

Tâches d’analyse d’un corpus audiovisuel

Chapitre 5
La tâche d'analyse « description de l'objet de connaissance »

5.1. Introduction

Dans ce chapitre nous considérerons plus particulièrement l'analyse des objets (situations, événements, etc.) composant le domaine de connaissances thématisé dans un texte audiovisuel ou un corpus de textes audiovisuels. Comme nous l'avons déjà dit, il s'agit ici de l'un des deux versants de la *description* dite *référentielle** de l'objet de connaissance thématisé dans un texte (audiovisuel, dans notre cas). L'autre versant de la description référentielle est celle de la contextualisation de l'objet de connaissance, c'est-à-dire de sa localisation (spatiotemporelle, sociale, historique, etc.) dont nous parlerons dans le sixième chapitre.

Rappelons également que la *description référentielle** constitue une des *tâches* centrales d'un *type* d'analyse spécifique du texte audiovisuel que nous appelons *description thématique** ou encore description du contenu audiovisuel. Les autres tâches qui composent le type d'analyse appelé description thématique sont les suivantes : la tâche de la *description de la mise en discours** d'un sujet développé dans le texte audiovisuel, la tâche de la *description de l'expression (ou mise en scène) audiovisuelle** et/ou verbale du contenu et, enfin, la tâche de la production

d'un commentaire par l'analyste lui-même au sujet de sa description (voir le chapitre 4, figure 4.4).

Nous commencerons ce chapitre en discutant un exemple simple de description référentielle (=la section 5.2). Pour cela, nous reviendrons à l'exemple développé dans le chapitre 3 relatif à la description d'un texte audiovisuel traitant de la *culture technique de la civilisation chavin*.

Ensuite (la section 5.3), nous préciserons plusieurs distinctions théoriques et méthodologiques qui nous paraissent importantes pour bien comprendre objet et enjeux de l'analyse référentielle. Il s'agit de la distinction entre *sujet** et *configuration* (ou structure) *topique** ainsi que de la distinction entre *configuration thématique** et *configuration topique* stricto sensu*.

La section 5.4 est consacré à la présentation rapide d'une bibliothèque de *séquences** pour la description référentielle de l'univers de discours d'une archive. Il s'agit de l'archive *ARC* (Archives *A la Rencontre des Cultures*)[37] qui forme un des chantiers d'expérimentation du projet ASA-SHS.

Enfin, la section 5.5 introduit la problématique de la modélisation conceptuelle comme une activité indispensable pour définir un modèle de description du contenu d'un corpus audiovisuel. Cette problématique sera développée davantage au chapitre 8 de ce livre.

5.2. Un exemple simple de description référentielle

L'extrait du formulaire d'analyse représenté dans la figure 3.2 (chapitre 3) nous montre le champ fonctionnel réservé à la tâche d'analyse *Description du/des objets de connaissance thématisés* dans le texte audiovisuel.

Dans notre exemple, ce champ fonctionnel est pourvu, comme déjà discuté dans le chapitre 3, d'une *séquence globale* composée de *deux sous-séquences*. Chaque sous-séquence est, à son tour, composée *de schémas de définition** au niveau desquels s'effectue le travail d'analyse ou de description à proprement parler d'un certain type d'objets. En effet, comme le montre la figure 4.2, chacune des deux sous-séquences :

– porte sur un *type spécifique d'objets du domaine de référence :* la première sous-séquence porte sur le type d'objets [Civilisation] et la deuxième sous-séquence sur le type d'objets [Formation culturelle] ;

[37] Voir http://semiolive.ext.msh-paris.fr/arc/

– intègre une *procédure de description** (chaque procédure étant composée de plusieurs *activités* spécifiques *de description*) à l'aide de laquelle l'analyste effectue son travail d'analyse consacré à un objet de connaissance : la première sous-séquence intègre la procédure de description dite *libre* (utilisée pour renseigner le terme conceptuel [Civilisation]) ; la deuxième sous-séquence intègre la procédure de description dite *contrôlée* (utilisée pour renseigner le terme conceptuel [Formation culturelle].

La figure 3.2 (chapitre 3) nous montre que le terme conceptuel [Civilisation] peut être renseigné par quatre activités de description qui forment donc ensemble la procédure de *description libre* – procédure qui sera présentée encore davantage au chapitre 9 :

1) l'activité de la *désignation du terme conceptuel sélectionné*, c'est-à-dire du terme [Civilisation]. Dans notre cas, il s'agit de produire l'expression <Civilisation chavin> via un formulaire d'indexation associé à cette activité.

2) l'activité de la *production éventuelle d'informations complémentaires* qui contextualisent davantage la simple mention <Civilisation chavin>. Une telle contextualisation peut être temporelle (historique), spatiale (géographique) ou encore thématique (institutionnelle, sociale.).

3) l'activité de la *production (éventuelle) d'une présentation textuelle* synthétique sous forme d'un petit texte explicatif.

4) enfin, si pertinent, l'activité de la *production d'une version en langue d'origine* de l'expression désignant le terme conceptuel.

Précisions que *seule la première activité*, celle de désigner le terme conceptuel [Civilisation] est obligatoire – toutes les autres sont facultatives. Bien sûr, plus une analyse s'appuie sur les différentes activités de description, plus elle s'enrichit mais plus elle exige aussi du temps de réalisation.

Comme le montre également la figure 3.2 (chapitre 3), une fois la description dite libre du terme conceptuel [Civilisation] terminée, l'analyste est invité à renseigner le ou les faits culturels dont est question dans le texte audiovisuel soumis à l'analyse. Ici, la description du fait culturel se fait via la *procédure de description contrôlée**. Cela veut dire que l'analyste utilise un *thesaurus** dans lequel il désigne le ou les termes prédéfinis (le ou les descripteurs) les plus appropriés, selon lui, pour caractériser le fait culturel en question. Dans notre cas concret, le terme prédéfini, le *descripteur*, que l'analyste a sélectionné afin de rendre compte du fait culturel dont est question dans son texte audiovisuel est le terme <Culture technique>.

En somme, le travail de description référentielle, accompli via le formulaire représenté par la figure 4.2, se résume à l'information produite par l'analyste précisant que le texte audiovisuel soumis à l'analyse « parle de » la *culture technique de la civilisation chavin.*

Notons encore que l'acte le plus simple de la description thématique se réduit à deux « gestes » bien précis :

– soit à la *sélection du concept* (du terme conceptuel) à renseigner (dans notre cas, le choix est limité à la sélection consécutive des deux termes conceptuels [Civilisation] et [Formation culturelle]) ;

– soit à la *saisie libre de l'expression nominale* représentant, selon l'analyste, au mieux le référent du concept ou du terme conceptuel (dans notre exemple, le référent choisi par l'analyste pour désigner le concept [Civilisation] est l'expression <Civilisation chavin>), soit à la sélection contrôlée d'un ou de plusieurs descripteur(s) dans un thesaurus (dans notre cas, il s'agit de la sélection du descripteur <Culture technique> pour désigner la valeur du concept [Formation culturelle]).

5.3. Structure thématique, structure topique et objets référentiels

L'analyse du contenu – du ou des sujets thématisés et développés – d'un texte ou d'un corpus de textes audiovisuels se fait, comme nous le savons déjà, à l'aide de *structures thématiques prédéfinies*. Une telle structure thématique prédéfinie est représentée par un terme conceptuel (TC) ou, plutôt, une *configuration de termes conceptuels*. Chaque terme conceptuel est obligatoirement défini dans le méta-lexique ASA. Rappelons que nous distinguons entre deux grands types de termes conceptuels :

– les termes conceptuels qui représentent les *objets d'analyse**,
– les termes conceptuels qui représentent les *activités d'analyse**.

Comme nous le verrons encore plus en détail, pour chacun des deux types nous avons défini un vocabulaire conceptuel, une *ontologie descriptive* qui couvre l'*univers de discours ASA* (voir pour plus d'explications voir la *Partie IV* de ce livre).

Une structure thématique qui est représentée par un seul terme conceptuel est dite *condensée* ; une structure thématique qui est représentée par une configuration de termes conceptuels est dite *expansée* (ou, plutôt, plus ou moins expansée). La structure thématique maximalement expansée est celle qui est constituée par une agrégation de configurations locales définissant, chacune, une des cinq séquences

majeures du modèle de description servant à la réalisation du type d'analyse spécifique appelée *description thématique (du contenu)** (voir supra, chapitre 4).

La *structure thématique** qui est maximalement condensée est celle qui est réduite à *deux termes conceptuels* :

– le premier représente un *objet d'analyse** ;
– le second représente l'*activité de description** à employer pour renseigner l'objet d'analyse.

La figure 5.1 nous montre les configurations locales principales qui forment la *structure thématique** à l'aide de laquelle nous saisissons un *sujet** développé (verbalement, visuellement ou par des moyens purement acoustiques) dans un texte audiovisuel. Elle nous montre également le fait que seule la configuration topique fait partie intégrale de la structure thématique (d'où d'ailleurs l'usage synonymique fréquent des deux expressions « topique » et « thématique »).

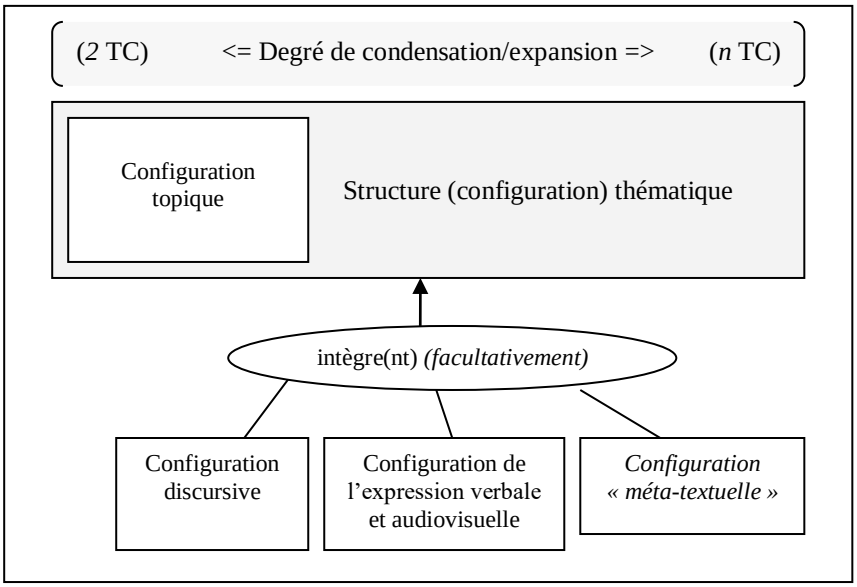

Figure 5.1. *La configuration thématique et les principales configurations locales qui la composent*

Revenons aux *objets d'analyse**. Dans [STO 11a], nous avons distingué cinq grandes catégories que nous utilisons pour définir les *modèles thématiques* (c'est-à-dire de description thématique ou du contenu) à l'aide desquels nous décrivons un sujet thématisé et développé dans un texte ou, plutôt, dans un corpus de textes audiovisuels :

1. les *objets référentiels* (c'est-à-dire faisant partie de tel ou tel domaine empirique documenté par un corpus audiovisuel) ;
2. les *objets contextuels* (de localisation spatiale, temporelle, etc.) ;
3. les *objets discursifs* (c'est-à-dire de mise en discours) ;
4. les *objets de l'expression audiovisuelle* (c'est-à-dire de l'expression visuelle, de l'expression acoustique, de l'expression syncrétique, etc.) et *verbale* ;
5. les *objets* dits *réflexifs* (c'est-à-dire servant à produire un métadiscours dont l'objet est l'analyse elle-même).

Un modèle de description thématique (du contenu) d'un texte ou d'un corpus de textes audiovisuels « pioche », pour parler ainsi et comme le montre la figure 5.1, dans chacune de ces cinq catégories d'objets pour en sélectionner ceux dont elle a besoin pour rendre compte de son objet textuel.

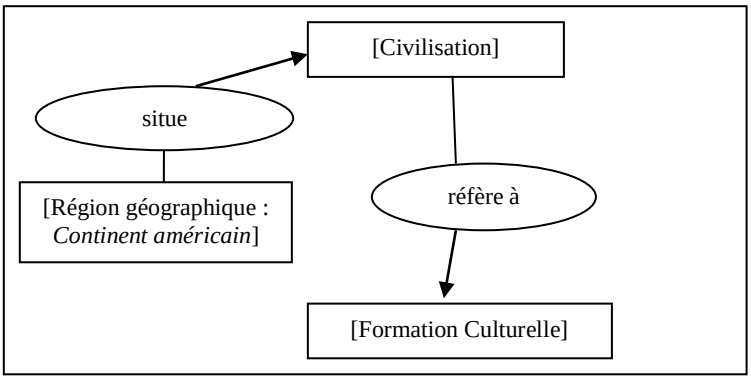

Figure 5.2. *Configuration de termes conceptuels représentant une partie de l'objet de connaissance « Les civilisations et leurs cultures sur le sol américain »*

Considérons maintenant de plus près les objets référentiels – la première des cinq catégories d'objets d'analyse que nous venons d'identifier. Les objets référentiels définissent en effet une *sous-partie* de la structure thématique qui est la topique ou, plutôt, la *structure topique*. La topique est, simplement dit, ce dont « parle » un texte et une structure topique est le modèle à l'aide duquel nous pouvons décrire tous les textes (audiovisuels) dont le sujet (ou un des sujets) correspond (plus ou moins bien) à la structure topique sous-jacente.

Ainsi, pour reprendre notre exemple introduit au troisième chapitre, le sujet, la *topique* du texte audiovisuel soumis à l'analyse est *la culture technique de la civilisation chavin pendant la période de son apothéose telle qu'elle se manifeste sur le lieu du village andin Huantar au Pérou* (voir chapitre 3, figure 3.2).

La *structure topique* dont le formulaire de travail représenté par la figure 3.2 constitue l'interface de travail forme, par contre, une *configuration générique* qui permet d'analyser en fait pas seulement (heureusement !) le texte audiovisuel en question mais bien tous textes audiovisuels qui traitent des *civilisations et de leurs cultures sur le sol du continent américain*.

La figure 5.2 montre une représentation (cognitive) d'une partie de la structure topique qui a permis l'analyse, entre autres, du texte audiovisuel parlant de la civilisation *chavin* et qui nous sert d'exemple ici.

Tout type d'objets d'analyse est représenté par un *concept* ou, comme nous préférons le dire, par un *terme conceptuel* (ou TC) qui est mis entre crochet et qui commence avec une lettre majuscule. Ainsi, [Civilisation] est un terme conceptuel tandis que <Civilisation> est une expression nominale générée par l'analyste ou prédéfinie dans un thesaurus ;

Considérons maintenant de plus près la figure 5.2. Elle représente un petit graphe [SOW 84] orienté entre deux termes conceptuels mis en rapport à l'aide d'une relation conceptuelle. Le graphe est ainsi le produit :

– d'une sélection de *deux termes conceptuels* (parmi les termes conceptuels faisant partie du vocabulaire, c'est-à-dire du méta-lexique de termes ASA, qui représente les objets d'analyse du l'*univers de discours** ASA) et d'une *relation conceptuelle* ;
– d'un processus de positionnement réciproque des deux termes à l'aide de la relation {TC_i réfère à TC_j} avec comme résultat une configuration thématique dans laquelle le terme conceptuel [Civilisation] est le terme *ab quo* (c'est-à-dire d'origine) et le terme conceptuel [Formation culturelle] le terme *ad quem* (c'est-à-dire d'aboutissement).

Retenons donc que toute *description référentielle** ou *topique* d'un corpus de textes audiovisuels (ou autres) s'appuie, se base, sur la sélection d'un ou d'un ensemble de termes conceptuels (TC) représentant la première catégorie d'objets d'analyse, c'est-à-dire la catégorie des objets dits référentiels. Ils – les termes conceptuels sélectionnés – spécifient également, parmi toutes les *séquences** composant un modèle de description du contenu audiovisuel, une classe fonctionnelle particulière qui sert à la *description du domaine* (ou, plutôt, des *objets* du domaine) *thématisé dans un texte* ou un corpus de textes (audiovisuels). Autrement dit, cette classe de séquences sert à la description de ce *de quoi parle* un texte (voir chapitre 1.6). Rappelons encore que cette classe fonctionnellement spécialisée de séquences occupe toujours la *première position* dans la *structure syntagmatique canonique* (voir chapitre 4) d'un modèle de description thématique. Dans le cadre d'une analyse du contenu audiovisuel, c'est aussi la *seule* séquence dans le formulaire interactif à renseigner obligatoirement par l'analyste.

5.4. Une bibliothèque de séquences pour la description référentielle

La figure 5.2 est une représentation de la partie référentielle du modèle de description du contenu audiovisuel qui sous-tend le formulaire interactif de travail représenté par la figure 3.2. La figure 5.3 montre, en revanche, comment on peut rendre compte du même modèle du point de vue de l'organisation du formulaire interactif qui sert d'interface de travail entre le modèle de description et l'analyste.

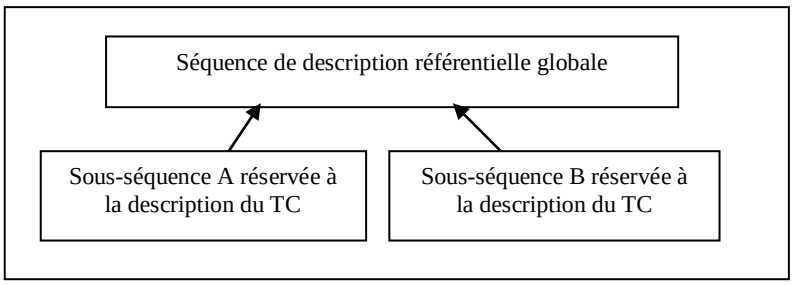

Figure 5.3. *Organisation interne – hiérarchique et fonctionnelle – d'une séquence référentielle*

Le choix ici est de construire, pour chaque terme conceptuel composant la configuration topique représentée dans la figure 5.1, des sous-séquences particulières. Chaque sous-séquence est, dans cette optique, réservée à la description d'un terme conceptuel.

Considérons la figure 5.4. Elle nous montre un extrait de la *bibliothèque des séquences* utilisées pour la description du contenu des textes audiovisuels composant le fonds des archives A la Rencontre des Cultures (ARC). Cette bibliothèque des séquences est définie et réalisée par une personne (ou un groupe de personnes) jouant le rôle de *modélisateur* (de l'*ingénieur de connaissances*) dont la mission est de développer et de gérer le métalangage de description pour l'analyse des textes et corpus audiovisuels des archives ARC. Ce travail spécifique – qui précède obligatoirement tout travail d'analyse – se réalise à l'aide de l'*Atelier de Modélisation ASA** et à l'aide d'un éditeur xml qui s'intitule OntoEditeur[38].

[38] OntoEditeur est un outil conçu et développé à l'ESCoM par Francis Lemaitre dans le cadre de plusieurs projets de R&D européens (notamment le projet LOGOS) et français (notamment les projets SAPHIR et ASA-SHS). Cet outil sert aujourd'hui à la spécification et réalisation des ressources métalinguistiques ASA composées, comme déjà expliqué au deuxième chapitre, de *deux méta-lexiques des termes conceptuels* représentant d'une part les objets d'analyse et d'autre part les activités d'analyse ; d'un *thesaurus* composé d'un ensemble de facettes et de plusieurs milliers d'expressions prédéfinies ; de *briques de modèles de*

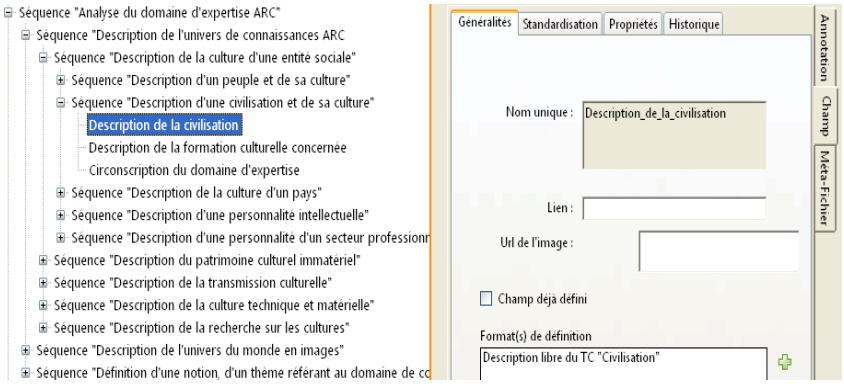

Figure 5.4. *Extrait de la bibliothèque des séquences définissant l'univers de discours des archives ARC*

Les séquences composant la bibliothèque des séquences ARC sont organisées sous forme de collections de séquences par grands thèmes ou *sujets* qui circonscrivent l'*univers de discours* des archives* ARC. On y trouve, par exemple, des collections de séquences ou des séquences isolées définissant la structure topique de thèmes ou sujets tels que :

– formations culturelles d'un peuple ;
– formations culturelles d'un pays ;
– patrimoine culturel immatériel ;
– culture technique et matérielle ;
– recherche sur les cultures ;
– etc.

Remarquons que ces séquences peuvent être réutilisées telles quelles ou, moyennant certaines modifications, pour définir la partie référentielle des modèles de description du contenu audiovisuel qui font partie des bibliothèques d'autres archives[39].

Parmi les collections de séquences définissant l'univers de discours des archives ARC (figure 5.4) se trouve donc la collection de séquences *Description d'une civilisation et de sa culture* – collection qui contient nos deux sous-séquences

---

description (sous forme de *schémas de définition* et de séquences de *description fonctionnellement spécialisées*) ; de *modèles de description* ; et, enfin ; de *bibliothèques de modèles de description*. Nous le présenterons rapidement dans le chapitre 11.

[39] On voit bien ici que production et suivi de modèles de description peuvent devenir une activité professionnelle et économique tout à fait centrale dans le contexte de l'industrie et de l'économie de la connaissance.

Description de la civilisation et *Description de la formation culturelle concernée*. Elles sont accompagnées d'une troisième sous-séquence intitulée *Circonscription du domaine d'expertise* qui établit *a priori* les référents à respecter par l'analyste lorsqu'il procède à une analyse concrète. Ces référents définissent en quelque sorte ce que le grand sociologue Alfred Schütz [SCH 03] a appelé, en se référant à la phénoménologie husserlienne, l'horizon du sens de la structuration épistémique du monde de vie.

Rappelons qu'une séquence est obligatoirement définie par un ou plusieurs *schémas de définition**. Dans la figure 5.4, on voit (en bas à droite de l'image) un champ qui s'intitule « Champ de définition » réservé justement à la sélection du ou des schémas appropriés à la spécificité structurale et fonctionnelle d'une séquence (fonctionnellement spécialisée soit dans la description référentielle d'une topique, soit dans la description de la mise en discours d'une topique, soit dans la description de l'expression verbale ou audiovisuelle d'une topique, etc.).

Chaque schéma de définition sélectionné ou simplement sélectionnable pour spécifier une séquence fait partie lui-même d'une *bibliothèque de schémas conceptuels* de l'univers de discours des archives ARC. Cette bibliothèque définit des micro-configurations génériques soit d'*objets d'analyse**, soit d'*activités d'analyse**.

Un schéma de définition lui-même est caractérisé soit par un terme conceptuel soit par une réunion de deux ou plusieurs *termes conceptuels* et/ou d'un ou de plusieurs *référents*. Les termes conceptuels font partie, comme nous le savons déjà, du *méta-lexique des termes conceptuels* représentant le domaine d'expertise ASA. Les référents pouvant définir un schéma conceptuel, proviennent du thesaurus ASA. Dans la partie IV de ce livre, nous reviendrons plus en détail sur l'organisation assez complexe d'un modèle de description sous forme de séquences, schémas, termes conceptuels et référents.

5.5. Des architectures fonctionnelles alternatives pour définir des séquences de description référentielle

Ceci dit, la mise en forme d'un formulaire interactif selon le modèle représenté dans la figure 5.3 implique un choix conceptuel important.

En effet, l'usage d'un tel formulaire oblige l'analyste de décrire la structure topique *pas à pas*, d'une manière analytique, en renseignant, un par un, chaque terme conceptuel qui la compose. Comme le montre notre exemple portant sur l'extrait du formulaire d'analyse du contenu (chapitre 3, figure 3.2), l'analyste décrit d'abord librement, dans la *première* sous-séquence, le terme conceptuel [Civilisation] en l'indexant à l'aide de l'expression <Civilisation chavin> ; il décrit

ensuite d'une manière contrôlée, dans la *deuxième* sous-séquence, le terme conceptuel [Formation culturelle] à l'aide d'un thesaurus spécialisé.

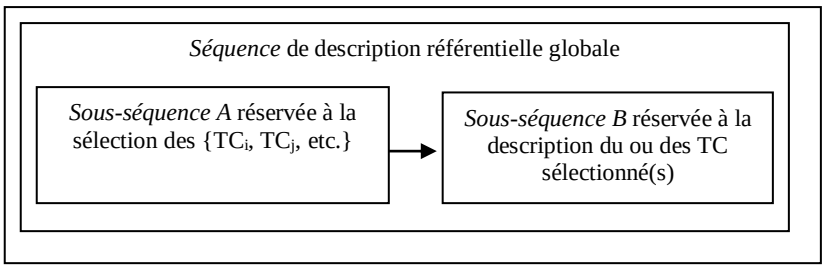

Figure 5.5. *Architecture fonctionnelle d'un modèle de description alternatif à celui représenté par la figure 3.2*

En revanche, la figure 5.5 décrit l'architecture fonctionnelle d'un *modèle alternatif* à celui dont on voit l'interface dans figure 3.2 (chapitre 3). Tout en gardant la distinction entre séquence globale (réservée à la description référentielle d'un texte audiovisuel), l'architecture décrite dans la figure 5.5 réserve :

– la *première sous-séquence* à la *sélection* par l'analyste du ou des termes conceptuels appropriés pour définir le sujet ou la topique qu'il envisage de décrire dans le texte audiovisuel soumis à son analyse ;

– et la *deuxième sous-séquence* aux *procédures de description* (libres et/ou contrôlées) à proprement parler de la configuration des termes conceptuels sélectionnés.

La figure 5.6 montre à son tour l'interface qui est une réalisation de l'architecture fonctionnelle alternative représentée dans la figure 5.5 et qui tient également compte (comme celle représentée dans la figure 5.3) de la configuration générique définissant la partie du domaine de connaissances *Civilisation et sa culture* (voir figure 5.2).

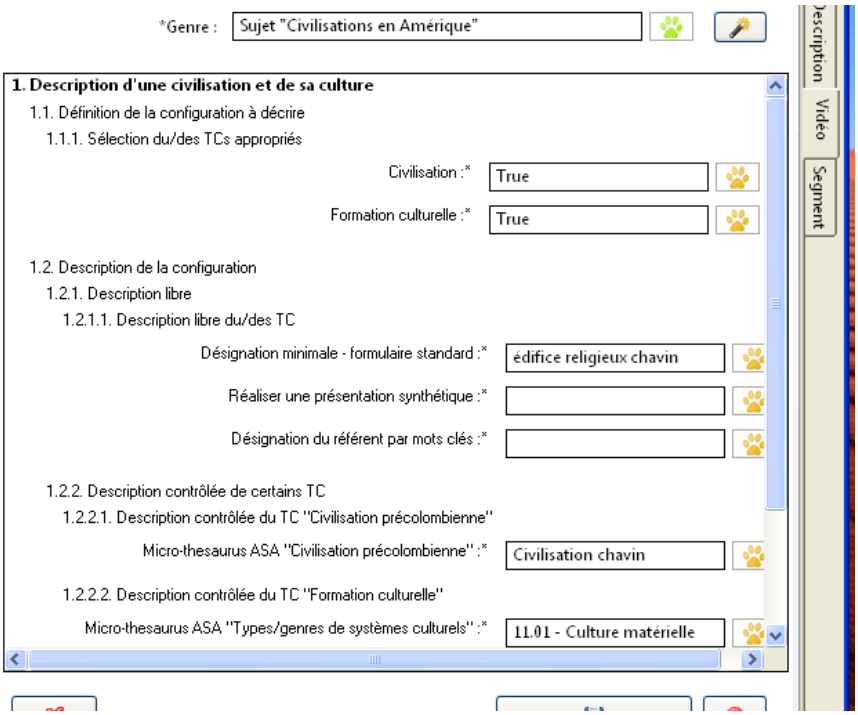

Figure 5.6. *Interface du modèle de description alternatif à celui représenté par la figure 3.2*

Cette interface alternative est en effet bien plus puissante que celle représentée par la figure 3.2 (chapitre 3) mais elle exige également, de la part de l'analyste, plus de connaissances « techniques », plus d'aisance à manipuler des concepts métalinguistiques.

Si l'interface représentée par la figure 3.2 (chapitre 3) n'exige, de la part de l'analyste, que la description à proprement parler de chaque terme conceptuel, l'interface représentée par la figure 5.6 demande en revanche à l'analyste :

1. de *configurer d'abord son domaine* d'analyse ou d'expertise,
2. *avant de le décrire* à proprement parler.

L'exemple montré dans la figure 5.6 est encore assez simple : elle montre deux termes conceptuels qui forment déjà une configuration générique très contraignante définissant (une partie de) la structure topique pour thématiser un objet de connaissance dans le texte. Dans ce cas concret, l'analyste peut sélectionner soit l'un soit l'autre terme conceptuel soit les deux à la fois.

Dans ce dernier cas (c'est-à-dire en sélectionnant les deux termes conceptuels), l'analyste signifie bien par là que la topique identifiée et décrite par lui dans le texte audiovisuel concerne la *configuration en elle-même des deux termes conceptuels* et moins l'un ou l'autre terme conceptuel pris isolément. Dans notre exemple (figure 5.6), l'analyste décide que l'expression verbale librement saisie <Edifice religieux chavin> indexe les deux termes conceptuels [Civilisation] et [Formation culturelle] *ensemble !*

L'analyste ne peut, bien sûr, sélectionner qu'un seul terme – soit le terme [Civilisation], soit le terme [Formation culturelle]. Dans ce cas là, il réalise un travail de description qui est similaire à celui qu'il réalise avec l'interface représenté par la figure 3.2 (chapitre 3).

Tout en prenant en considération le fait qu'il s'agisse ici d'un exemple bien simple (c'est-à-dire d'une configuration définitionnelle composée de deux termes conceptuels seulement), on peut se rendre compte assez aisément que le fait de laisser à l'analyste une (certaine) liberté de configurer son objet de connaissance (c'est-à-dire de sélectionner parmi un ensemble de termes conceptuels) un *sous-ensemble de termes pertinents*, voire à la limite *un seul terme conceptuel*), ouvre des possibilités très intéressantes dans le développement de formulaires d'analyse du contenu pour des *fonds audiovisuels à thèmes très variés*.

Nous pensons ici plus particulièrement au fonds audiovisuel qui compose le Programme Archives Audiovisuelles de la Recherche. C'est un fonds qui couvre une grande diversité de disciplines en sciences humaines et sociales. Difficile, dans ce contexte d'envisager l'élaboration de formulaires interactifs d'analyse qui ressemblent à celui représenté par la figure 3.2. Cela impliquerait très certainement le développement de plusieurs centaines de formulaires difficilement maîtrisables par les différentes parties impliquées dans le processus de la modélisation (et du suivi) des formulaires, de l'analyse des corpus audiovisuels et de la publication de ces analyses sous forme de portails thématiques ou d'accès spécialisés.

Ainsi, la possibilité évoquée ci-dessus, c'est-à-dire d'afficher des ensembles raisonnés de termes conceptuels dans un formulaire interactif de travail et de laisser à l'analyste le choix d'en sélectionner ceux qui sont pertinents pour son travail, semble constituer une solution possible pour limiter assez drastiquement le nombre de formulaires interactifs dédiés à l'analyse du contenu audiovisuel. C'est en tout cas la voie que nous avons choisi pour développer une bibliothèque de formulaires de base pour l'analyse du contenu audiovisuel du fonds du programme AAR.

Revenons encore une fois sur notre exemple représenté dans la figure 5.6. Nous avons dit que l'analyste peut définir son objet de connaissance par la sélection conjointe [Civilisation] et [Formation culturelle]. Mais il peut également se contenter de ne choisir qu'un seul terme conceptuel – soit le terme [Civilisation], soit le terme [Formation culturelle]. En optant pour l'une ou pour l'autre de ces deux

solutions, l'analyste signifie bien par là qu'il procède, selon le cas, soit à la description d'une civilisation sans référence à une formation culturelle particulière, soit à la description d'une formation culturelle sans référence à une civilisation particulière.

Du point de vue de la modélisation du domaine référentiel de connaissances de l'univers de discours d'une archive (dans notre cas : celui des archives ARC), cela peut poser cependant des problèmes.

En ne sélectionnant que le terme conceptuel [Civilisation] ou que le terme conceptuel [Formation culturelle], l'analyste se débarrasse en quelque sorte de la structure très particulière qui définit l'objet de connaissance du formulaire représenté dans la figure 3.2, c'est-à-dire la description référentielle de sujets ou de topiques qui traitent des civilisations (amérindiennes) *en rapport* à une formation culturelle particulière.

Ainsi l'analyste peut *a priori* utiliser le formulaire représenté par la figure 4.2 pour décrire tout texte audiovisuel qui traite des civilisations (sur le sol américain) « en tant que telles » et sans rapport aucun avec une formation culturelle, voire même, au contraire, en relation avec des objets de connaissance totalement différents de ceux pour lesquels le formulaire en question devrait être utilisé et pour l'analyse desquels il existe, peut-être, d'autres formulaires dans la bibliothèque des formulaires d'analyse du contenu audiovisuel.

De même, l'analyste peut opter dans l'utilisation dudit formulaire pour ne décrire qu'une formation culturelle (sur le sol américain) sans rapport avec une civilisation particulière. Il peut, par exemple, utiliser le formulaire pour décrire un texte audiovisuel qui thématise la culture technique à une certaine époque donnée dans telle ou telle région du continent américain.

Cette ambiguïté partage aussi bien l'architecture fonctionnelle sous-tendant le formulaire représenté par la figure 5.6 que celle qui sous-tend le formulaire représenté par la figure 3.2. Si nécessaire, elle peut être levée si on précise davantage les exigences propres à une configuration définissant un objet de connaissance tel que celle qui représente l'objet *Les civilisations et leurs cultures sur le continent américain* (figure 5.2).

Ainsi, dans la figure 5.2, rien n'est dit si, par exemple, l'un ou l'autre ou les deux termes conceptuels sont nécessaires pour la définition de l'objet en question, rien n'est dit si l'un des deux termes dépend « crucialement » de l'autre (dans ce sens que le terme dépendant ne peut être sélectionné que si le terme dont il dépend est déjà asserté, sélectionné), rien n'est dit non plus si le choix de tel ou tel terme exclut tel ou tel autre terme, et ainsi de suite.

Tout cela relève de l'*analyse conceptuelle du domaine de référence* de l'*univers de discours d'une archive*. La configuration générique telle qu'elle se présente dans la figure 5.2 est muette par rapport à toutes ces contraintes possibles. En son état actuel elle ne fournit pas les consignes nécessaires pour pouvoir décider de l'architecture fonctionnelle la plus optimale du formulaire interactif de travail qui est supposé « matérialiser » aussi précisément que possible le modèle de description qui le sous-tend.

Dans le chapitre 9, nous reviendrons sur cette question très importante d'une modélisation conceptuelle plus contraignante d'un domaine de connaissances qui tient compte des exigences que nous venons d'énumérer : dépendance d'un (ensemble) de terme(s) conceptuel(s) d'un autre (ensemble) de terme(s) conceptuel(s), choix exclusif ou non d'un terme conceptuel, etc.

Chapitre 6

La tâche d'analyse « contextualisation du domaine de connaissances »

6.1. Introduction

Les figures 3.3 à 3.5 (chapitre 3) montrent une deuxième étape dans la description thématique (c'est-à-dire du contenu) d'un texte audiovisuel ou d'un passage spécifique dans un texte audiovisuel : celle de la *contextualisation de l'objet de connaissance* identifié au préalable.

Le terme de la contextualisation est ici utilisé au sens de *contextualisation référentielle*. Il s'agit de la procédure qui explicite le *cadre référentiel* – le *périmètre* – de l'objet de connaissance. Dans ce sens, la contextualisation désigne une *tâche de description* qui constitue, avec celle de la description de l'objet de connaissance (voir chapitre 6), la tâche de la *description référentielle**. Rappelons que la tâche de la description référentielle, à son tour, forme l'une des quatre principales tâches d'un type spécifique d'analyse du texte audiovisuel – type d'analyse appelé *description thématique** (ou du contenu) – les trois autres tâches étant celles de la description de la *mise en discours* d'une topique*, celle de l'*expression audiovisuelle** et/ou

verbale de la topique et, enfin, celle du commentaire méta-textuel (permettant, par exemple, l'*explicitation du cadre* – du « point de vue » – de l'analyse du contenu elle-même).

La contextualisation référentielle se compose, en premier lieu, de la *contextualisation spatiale* et de la *contextualisation temporelle*. Dans nos recherches sur l'analyse de corpus audiovisuels, nous avons traité surtout ces deux formes de contextualisation afin de pouvoir expérimenter la localisation sur des cartes géographiques et la localisation chronologique sur des frises temporelles de sujets et thèmes développés dans des corpus audiovisuels soumis à l'analyse. Il s'agit ici de modalités actuellement assez populaires d'accès à des contenus textuels d'archives et de bibliothèques numériques [GUE 11, LEM 11b, DES 11d].

Mais il existe également une troisième forme de contextualisation que nous réunissons sous le terme général de *contextualisation thématiqu*e. Cette troisième forme explicite le cadre social ou institutionnel, historique ou culturel, épistémique, etc. de l'objet de connaissance préalablement décrit. Nous n'avons pas systématiquement développé cette troisième forme de contextualisation. Nous l'utilisons seulement, pour le moment, dans le cadre de la description des sujets qui traitent les domaines ou objets de recherche en sciences humaines et sociales. Par exemple, la littérature autrichienne de la fin du XIXe et de la première moitié du XXe siècle constitue, bien évidemment, l'objet d'un grand nombre de chercheurs et d'activités concrètes de recherche. Or, la contextualisation thématique de l'objet de recherche permet ici d'expliciter l'intérêt d'un chercheur pour cet objet, ses objectifs, ses références théoriques. Il s'agit donc d'une forme de contextualisation extrêmement importante qui pourra permettre, à terme, de mieux profiler, de mieux hiérarchiser, les résultats obtenus suite à des requêtes d'information (*à la* Google) de types *« identifiez moi toutes les ressources textuelles ou audiovisuelles qui traitent de la littérature autrichienne de la fin du 19ème et du de la première moitié du 20ème siècle en tant qu'objet de recherche (historique, esthétiques, sémiologiques, etc.) »*.

La section 6.2 décrit brièvement la problématique de la contextualisation par localisation spatiale d'un objet de connaissance traité dans un texte audiovisuel. La section 6.3 s'occupe de la localisation géopolitique et, plus précisément, de la localisation par pays qui en est la forme la plus usuelle. Dans la section 6.4, nous traitons encore brièvement la localisation géographique (physique) à proprement parler, c'est-à-dire la localisation dans un espace dit naturel.

Les sections 6.5 à 6.7 traitent de la localisation temporelle mais aussi historique des sujets ou thèmes abordés dans un texte audiovisuel.

Enfin, la section 6.8 propose une discussion rapide portant sur exemple concret de la contextualisation thématique.

6.2. La contextualisation par localisation spatiale

L'une des premières formes de contextualisation d'un objet de connaissance thématisé dans un texte audiovisuel servant d'objet de description à l'analyste, est celle de sa localisation dans l'espace. Etant donné la nature de nos corpus audiovisuels de travail (c'est-à-dire ressources audiovisuelles documentant des domaines de connaissances qui correspondent aux différents ateliers d'expérimentation du projet ASA-SHS – domaines tels que la diversité culturelle, le patrimoine et la recherche archéologiques ou encore le patrimoine littéraire), certaines formes de localisation spatiale sont très présentes tandis que certaines autres ne le sont pas. Ainsi, la contextualisation « géographique » au sens large du terme est évidemment omniprésente dans nos corpus de travail tandis que la contextualisation d'un objet de connaissance dans un espace abstrait (par exemple : géométrique[40]) ne l'est pas.

Une première forme de localisation de l'objet de connaissance est celle que nous appelons *géographique au sens large du terme*. Celle-ci inclut l'espace physique comme support de localisation aussi bien au sens d'un *espace social* (d'un espace « socialement construit ») qu'au sens d'un *espace naturel*.

Les figures 3.3 et 3.4 du chapitre 3 montrent l'exemple du traitement de cette forme de localisation s'appuyant sur l'espace physique (social ou naturel). Nous y voyons à l'œuvre plus précisément :

– une forme de localisation que nous désignons par l'expression *géopolitique* intégrant une forme de localisation plus spécifique encore, que nous appelons (faute de mieux) *administrative* (ou *territoriale*) ;
– ainsi qu'une forme de localisation *géographique* dite *physique*.

C'est la séquence 2 (voir chapitre 3, figure 3.3) du formulaire permettant de décrire des corpus audiovisuels consacrés à la topique [Civilisation en Amérique] qui est consacrée à la localisation *géographique* dans un espace naturel. La séquence

---

[40] Avec notre collègue Dominique Flament, chercheur en histoire des mathématiques au CNRS (Centre National de la Recherche Scientifique), nous sommes engagés dans un projet portant sur l'analyse du contenu audiovisuel d'un fonds important d'enregistrements de séminaires de recherche, de colloques et d'entretiens avec des chercheurs en (histoire des) mathématiques en vue de la mise en place d'une archive audiovisuelle dédiée à l'histoire des mathématiques et, plus particulièrement, des géométries (selon l'expression de D. Flament). Dans ce contexte, la localisation d'objets de connaissances dans l'espace abstrait devient naturellement une problématique majeure. Pour le métalangage ASA, ce projet constitue donc une excellente opportunité d'enrichir son vocabulaire conceptuel (c'est-à-dire son méta-lexique de termes conceptuels) ainsi que son thesaurus et d'élargir son champ empirique d'application.

3 (voir chapitre 3, figure 3.3) est, en revanche, consacrée à la description que nous appelons *géopolitique* et *territoriale*.

La localisation dite géopolitique et, plus particulièrement, territoriale, utilise l'environnement spatial *socialement construit et informé* comme moyen de localisation, tandis que la localisation géographique dite physique s'appuie sur l'*environnement naturel*. Dans le premier cas, ce sont des étendues et lieux tels que les agglomérations, les habitats, les territoires administratifs, les zones et régions à valeur culturelle, économique, militaire, etc. qui servent de « localisateurs ». Dans le deuxième cas, c'est le relief terrestre et les formations naturelles qui servent pour localiser un objet de connaissance.

En ce qui concerne la localisation dite géopolitique, on voit bien que celle-ci constitue un *cas particulier* de la localisation qui s'appuie sur l'espace social ou sur l'espace possédant un *sens* pour un acteur humain (une personne, un groupe, une société, un état, etc.)[41]. A côté de cette forme de localisation, on trouve bien entendu toute une diversité d'autres formes de localisation qui s'appuient sur l'espace social. C'est la spécificité du corpus d'analyse et de l'univers du discours d'une archive qui décide de l'opportunité d'expliciter telle ou telle forme particulière de localisation d'un objet de connaissance.

Notons seulement en passant qu'une forme de localisation est *toujours* pourvue d'une double spécification :

1. une *spécification thématique* : celle-ci nous précise le type d'espace social sollicité pour localiser un objet de connaissance – espace de la vie quotidienne, espace organisant le monde de travail, etc.
2. et une *spécification fonctionnelle* : celle-ci nous précise le rôle que joue et la valeur que possède tel ou tel lieu pour l'objet de connaissance localisé (c'est l'espace dans sa fonction narrative, selon Greimas [GRE 76] : lieux de qualification, lieux de reconnaissance, lieux de mise à l'épreuve, mais aussi lieux euphoriques ou dysphoriques, lieux à fonction véridictoire et ainsi de suite).

6.3. Localisation et contextualisation par pays

Une des formes les plus répandues de localisation appelée *géopolitique* de l'objet de connaissance thématisés dans un texte audiovisuel, est la *localisation par pays*. La figure 3.4 (chapitre 3) nous en fournit un exemple simple. Elle nous montre que

[41] Ce sont des formes de localisation qui s'appuient sur une *géographie humaine* qui, elle-même ne constitue, en fin de compte, un type particulier de localisation s'appuyant sur un espace de sens (un *espace « sémiotique »* général), ensemble avec des formes et procédures de localisations à l'aide des espaces anthropomorphes (c'est-à-dire imaginaires, de fiction, etc.) ou encore à l'aide des espaces tels qu'ils organisent la vie des espèces naturelles.

cette forme de localisation procède par trois activités qui se succèdent. La réalisation de chacune de ces activités est confiée à une sous-séquence thématiquement et fonctionnellement spécialisée (voir aussi figure 6.1).

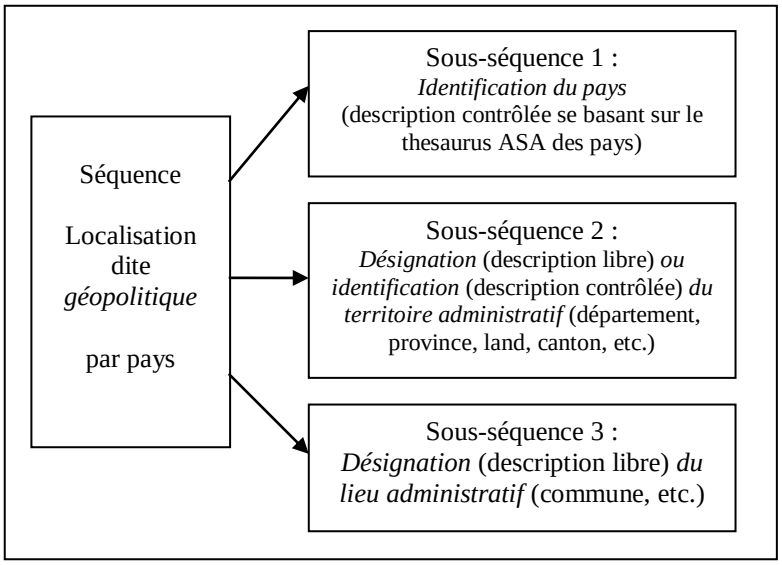

Figure 6.1. *Structure canonique pour procéder à la tâche de la localisation géopolitique par pays d'un objet de connaissance*

Dans une *première sous-séquence* (figure 6.1), l'analyste identifie le pays pertinent dans le micro-thesaurus ASA des *pays contemporains*. Il s'agit donc ici d'une procédure de *description contrôlée** que l'analyste peut compléter, s'il le souhaite, en produisant encore une petite présentation explicative.

Dans une *deuxième sous-séquence* (figure 6.1), l'analyste décrit éventuellement (ce n'est pas obligatoire !) le territoire au sens administratif du terme (en France, le *département* ou la *région* ; en Allemagne ou en Autriche, le *land* ; en Suisse, le *canton*, etc.). Pour certains pays, le thesaurus ASA dispose des listes des territoires administratifs. C'est le cas, bien sûr, de la France mais aussi du Pérou et de la Bolivie étant donné que ces deux pays forment le monde géographique, politique et culturel d'un des ateliers d'expérimentation du projet ASA-SHS qui portent sur le patrimoine culturel immatériel des communautés andines (voir la figure 6.2 qui montre un extrait du micro-thesaurus ASA relatif au régions administratives du Pérou). Si les territoires administratifs d'un pays se trouvent saisis dans le thesaurus ASA, alors leur identification pour localiser un objet de connaissance se fait via une procédure de description contrôlée, sinon, l'analyste saisit librement celui qui est pertinent pour son propos.

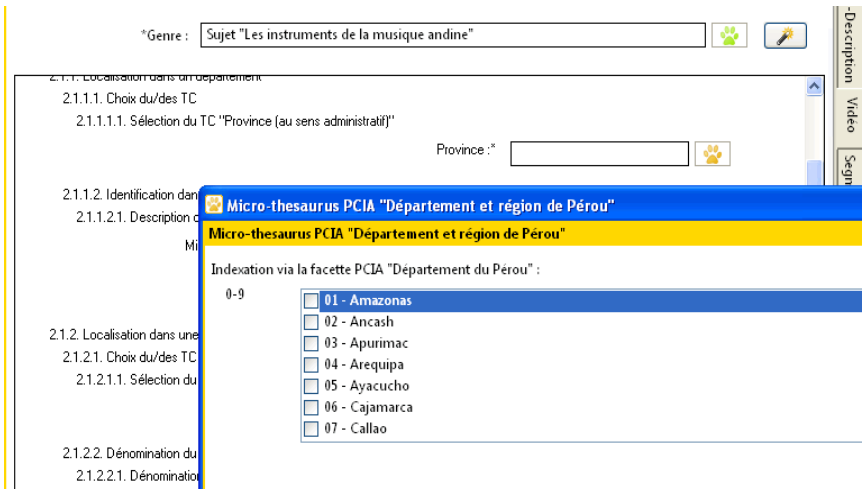

Figure 6.2. *Extrait du micro-thesaurus ASA des régions administratives du Pérou – micro-thésaurus réalisé par V. Legrand [LEG 11a]*

Dans une *troisième sous-séquence*, l'analyste décrit librement, s'il le souhaite, le lieu administratif plus précis qui sert de localisation de l'objet de connaissance thématisé dans un texte audiovisuel. Nous avons identifié, pour les corpus identifiés, un type particulier de lieux administratifs représenté par le terme conceptuel [Commune]. Si l'analyste le souhaite, il saisit librement le nom de la commune, en fournit une présentation synthétique complémentaire et, ensuite, la classifie à l'aide d'un petit micro-thesaurus répertoriant toute une diversité culturelle de types de communes (<Bourg>, <Lieu-dit>, <Ville>, <Village>, <Hameau>, <Commune agricole>, etc.).

On voit que la succession de ces trois sous-séquences définissant la structure de la localisation dite géopolitique par pays, est motivée par un mouvement *du global vers le local* : d'une *localisation globale* de l'objet de connaissance thématisé dans un texte audiovisuel vers sa *localisation locale* et plus *circonscrite*. Cela signifie plus particulièrement que la première sous-séquence est présupposée par la deuxième et la troisième. Pour renseigner un territoire administratif et/ou une commune, l'analyste doit d'abord renseigner le pays !

La localisation dite géopolitique par pays (voir chapitre 3, figure 3.4 ainsi que la figure 6.1 dans ce chapitre) ne constitue qu'une variante thématique de cette forme de contextualisation spatiale. Mais elle est la plus répandue et, pour nos propres corpus de travail, la plus productive.

Cependant, il est tout à fait possible des construire, sur la base des ressources métalinguistiques ASA existantes, d'autres formes de localisation géopolitique tenant compte, par exemple, de territoires géostratégiques, d'espaces communs formés par des groupements de pays, d'espaces linguistiques, religieux, d'activités professionnelles et économiques, etc.

En ce qui concerne les procédures de description à utiliser pour procéder à la localisation par pays d'un objet de connaissance thématisé dans un texte audiovisuel, nous utilisons de nouveau les deux procédures que nous avons déjà rencontrées dans le chapitre précédent, à savoir la procédure de la description libre et celle de la description contrôlée. Dans notre exemple (voir figure 3.4), l'analyste procède comme suit :

– (procédure de description contrôlée :) il sélectionne d'abord <Pérou> dans le micro-thesaurus ASA des pays contemporains – micro-thesaurus associé à la tâche de description *Localisation par pays* (chapitre3, figure 3.4) ;
– (procédure de description contrôlée :) il sélectionne ensuite <Ancash> dans le micro-thesaurus ASA des régions (administratives) du Pérou d'aujourd'hui (chapitre3, figure 3.4) ;
– (procédure de description mixte – libre et contrôlée :) enfin, il saisit l'expression <Chavin de Huantar> pour dénommer le terme conceptuel [Commune] et le classe – à l'aide du micro-thesaurus ASA des toponymes administratifs sous le terme de <Village> (chapitre3, figure 3.4).

La procédure, telle que nous venons de l'exposer, nous suffit pour procéder d'une manière satisfaisante à la localisation par pays des sujets traités dans les différents corpus audiovisuels. Ajoutons, cependant qu'il est fort facile de compléter cette approche par diverses extensions permettant d'enrichir considérablement les résultats concrets d'une description de la localisation par pays telle que nous venons de la présenter.

Dans cette optique citons l'ontologie géopolitique de la FAO[42] qui constitue en effet une extension potentielle très enrichissante pour notre approche. L'ontologie géopolitique de la FAO opère avec un canon de termes conceptuels sensiblement similaires à celui que nous utilisons pour localiser géographiquement un objet de connaissance traité dans un texte ou un corpus de textes audiovisuels. Il s'agit de termes conceptuels tels que [Zone], [Territoire], [Pays], etc. Cependant, l'ontologie géopolitique de la FAO va bien plus loin, bien sûr, dans la définition de modèles pour la description d'un pays ou d'un territoire en y incluant, par exemple, des concepts pour déterminer :

– les antécédents historiques d'un pays (l'URSS en tant qu'antécédent de la Russie) ;
– les frontière(s) d'un pays ;

[42] Pour plus d'informations : http://www.fao.org/countryprofiles/geoinfo.asp

– les appartenances d'un pays à un groupement de pays et/ou une organisation ;
– le PIB d'un pays ;
– les divers codes d'appellation d'un pays, etc.

Toutes ces informations peuvent être intégrées automatiquement dans une procédure de base de localisation géographique telle que celle que nous venons de présenter et rendues disponibles à un public intéressé[43]. L'exemple de l'ontologie géopolitique de la FAO nous montre par ailleurs l'intérêt de veiller, constamment et aussi systématiquement que possible, à ce qu'on appelle l'*interopérabilité* des métadonnées descriptives produites à travers telle ou telle approche d'analyse ou de description de corpus de données.

Enfin, il nous paraît encore important de souligner un autre fait, à savoir celui qu'une *séquence de localisation géopolitique* utilisée dans le contexte d'une analyse de corpus audiovisuels peut être *réutilisée* telle quelle ou moyennant certaines modifications et adaptations dans des contextes d'utilisation les plus diverses – par exemple, dans le contexte de l'analyse de données géopolitiques relatives à un pays, un groupement de pays, etc.

On voit s'ouvrir ici en effet des très nouvelles problématiques en ingénierie de connaissances telles que celles de la mise en place de bibliothèques de *modèles* et de *briques** (*schémas* ou *séquences*) de modèles de description totalement inter-définis et partagés par des utilisateurs les plus divers travaillant dans le cadre de la constitution, conservation et diffusion/appropriation de patrimoines numériques de connaissances.

6.4. Localisation et contextualisation géographique physique

A côté de la localisation dite géopolitique et administrative, il existe également, comme déjà dit, une deuxième forme de localisation spatiale qui s'occupe de la *localisation géographique* à proprement parler d'un objet de connaissance thématisé par un texte audiovisuel, c'est-à-dire de sa localisation (éventuelle) dans un *espace physique naturel*.

---

43 Techniquement parlant, cette intégration se fait par une mise en correspondance soit des termes conceptuels soit des schémas conceptuels ou séquences d'un modèle de description ASA avec l'ontologie ou la ou les ressources externes à ASA. Cette mise en correspondance se fait à l'aide de l'outil OntoEditeur dont nous nous servons pour définir les ressources métalinguistiques pour l'analyse des corpus et archives audiovisuels.

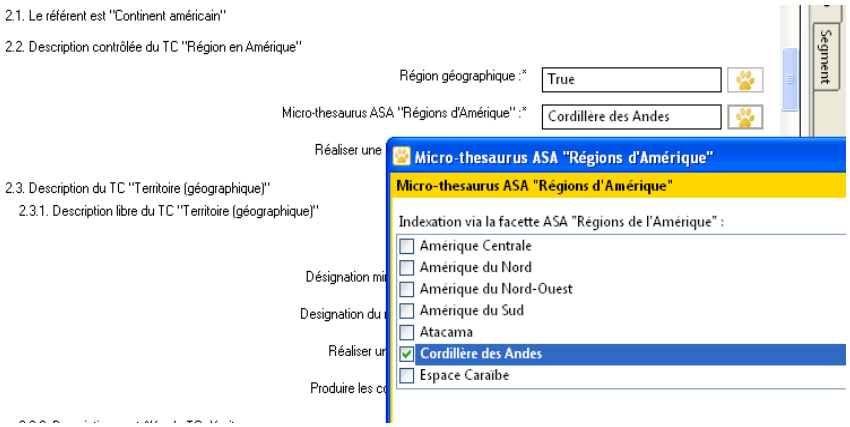

Figure 6.3. *Extrait du micro-thesaurus ASA des régions sur le continent américain*

La figure 3.3 dans le chapitre 3 nous montre un exemple simple d'une telle localisation géographique physique. L'exemple porte sur la localisation géographique de vestiges de la *culture technique de la civilisation chavin* dans l'*altiplano* des *Cordillères Blanches* au Pérou. Comme on peut le voir dans la figure 3.3 (chapitre 3), la tâche de la localisation physique (qui est une tâche spécialisée de la contextualisation spatiale d'un objet de connaissance thématisé dans un texte audiovisuel) est prise en charge par une séquence spécifique dans le modèle de description. Dans notre exemple, il s'agit de la deuxième séquence dans le modèle de description à l'aide duquel nous pouvons décrire des corpus audiovisuels dans lesquels sont développés des sujets relatifs à une ou plusieurs civilisations du passé ou du présent sur le sol américain. La procédure de description elle-même, composée de plusieurs activités de description spécifiques, se fait en deux temps :

1. Dans un premier temps, il s'agit de procéder à une *macro-localisation de l'objet de connaissance* thématisé dans le texte audiovisuel (ou tel ou tel passage du texte audiovisuel). Dans notre cas, cette macro-localisation renvoie à la chaîne des montagnes <Cordillère des Andes> qui traverse le continent sud-américain du nord au sud. La macro-localisation d'un objet de connaissance se fait via la première sous-séquence de la séquence réservée à la localisation géographique (voir figure 3.3 ; chapitre 3).
2. Dans un deuxième temps, la macro-localisation est complétée par une localisation plus précise qui identifie l'endroit, le lieu spécifique où se trouve physiquement l'objet thématisé dans le texte audiovisuel. Dans notre exemple, il s'agit de la <Cordillère Blanche> qui se situe dans la région péruvienne d'<Ancash>. Cette localisation plus précise est prise en charge par la deuxième sous-séquence de la séquence ayant en charge la description de la localisation géographique physique. A l'aide d'un micro-thesaurus

spécialisé, l'analyste peut encore, comme le montre la figure 6.4, déterminer le type ou le genre du lieu physique servant à la localisation plus précise de l'objet thématisé dans le texte audiovisuel. Dans notre cas, le lieu en question est une <Plaine d'altitude> ou <Altiplano>.

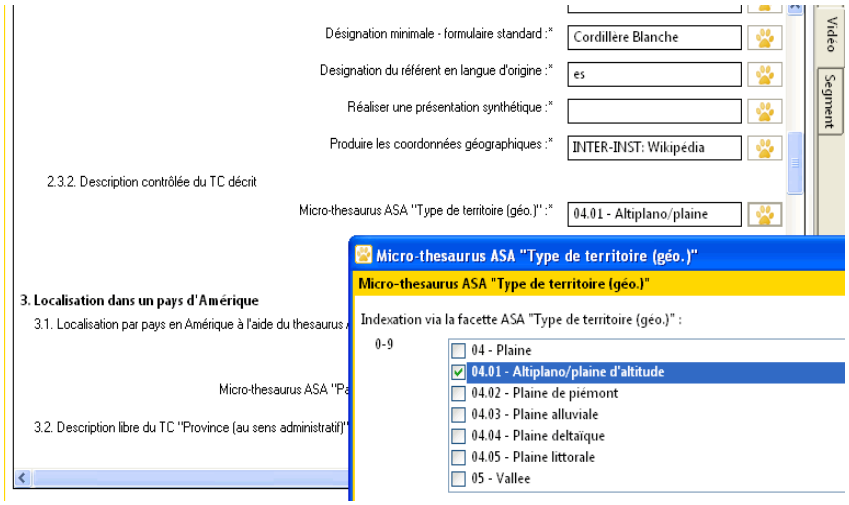

Figure 6.4. *Extrait du micro-thesaurus ASA des types géographiques physiques de territoires*

Ce processus de localisations successives et de plus en plus précises peut être affiné si on souhaite produire des localisations physiques encore plus détaillée. Mais, en règle générale la distinction entre ces deux niveaux suffit amplement. Le premier niveau de localisation, la macro-localisation, est une sorte de cadrage général tandis que le deuxième niveau correspond à la circonscription du lieu, de la région réellement significative pour la localisation de l'objet de connaissance thématisés. Remarquons en passant qu'on retrouve la même distinction fonctionnelle entre *deux* (voir, le cas échéant, *plusieurs*) *niveaux* de localisation aussi dans le contexte de la localisation administrative d'un objet de connaissance.

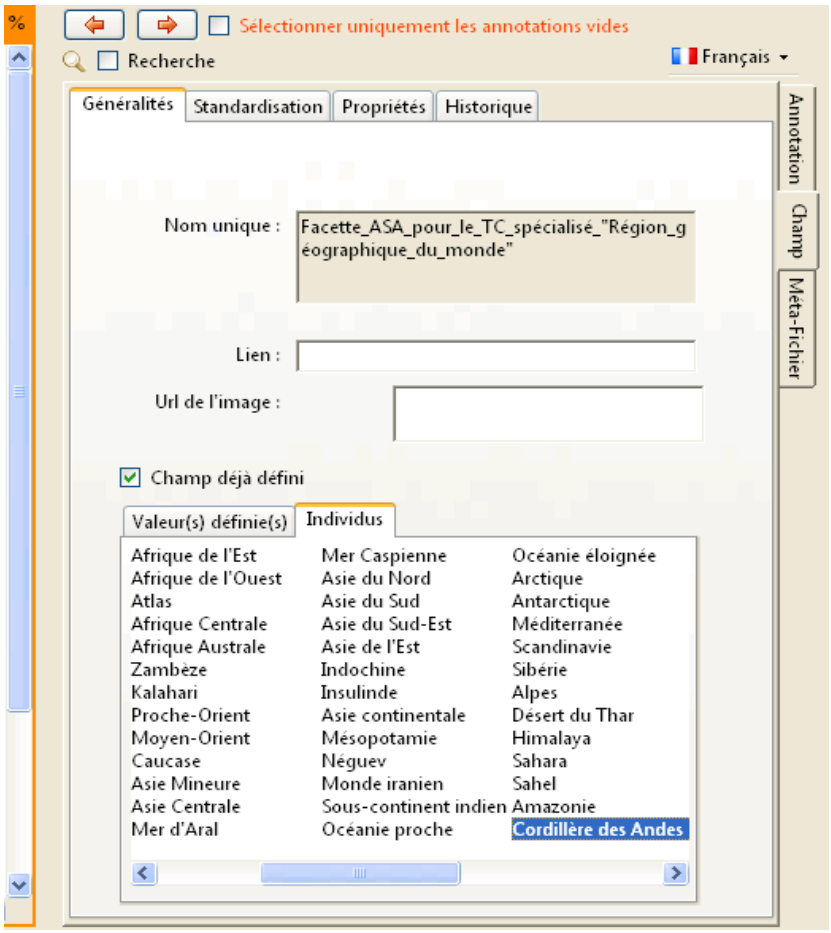

Figure 6.5. *Extrait du micro-thesaurus ASA « Régions du monde » (vue OntoEditeur)*

Dans le cadre de nos recherches nous avons pu identifier et réunir sous forme d'un *micro-thesaurus* toute une série d'étendues physiques pouvant jouer le rôle de « macro-localisateurs », voire des lieux physiques de localisation plus circonscrits des objets de connaissance thématisés dans les corpus audiovisuels qui proviennent, comme déjà dit à plusieurs reprises, du fonds collecté par notre programme AAR (Archives Audiovisuelles de la Recherche). On y trouve les continents, les grandes régions par continent, les cours d'eau, les océans, les mers, les chaînes de montagne, les grandes plaines, les grandes forêts, les déserts, etc. La figure 6.5 en montre un extrait. A côté de ce micro-thesaurus « mondial », nous avons également défini une série de micro-thesaurus de grandes étendues physiques par continent et/ou « partie du monde ». Comme le montre la figure 6.3, pour la description de la macro-

localisation de l'objet de connaissance *culture technique de la civilisation chavin*, nous avons utilisé le *micro-thesaurus ASA des étendues physiques du continent américain*. On trouve certaines étendues physiques dans l'un et l'autre micro-thesaurus (c'est le cas, justement, de la <Cordillère des Andes>, de l'<Espace caraïbe> ou encore du désert <Atacama>). Cela montre seulement qu'une expression (en général, nominale) faisant partie du thesaurus ASA peut être caractérisée par plusieurs *facettes* (de sens)* et donc appartenir à plusieurs *micro-thesaurus* (un micro-thesaurus représentant une *facette* particulière *de sens*).

Précisons, cependant, que le choix d'interpréter une certaine catégorie d'étendues physiques comme des *macro-lieux*, est entièrement déterminé par la nature du corpus audiovisuel avec et sur lequel on travaille. Ainsi, si l'univers de discours d'une archive est intrinsèquement limité, par exemple, à un pays ou une région, il va de soi que la liste des macro-lieux pertinents devra être adaptée au contexte spécifique de l'univers de discours en question. Cette adaptation est assez facilement à réaliser, à condition, bien sûr, qu'on connaît les macro-lieux pertinents :

– d'abord constitution d'un micro-thesaurus ASA des macro-lieux pertinents (pour l'univers de discours de telle ou telle archive) ;
– et ensuite remplacement, dans la séquence 2 (chapitre 3, figure 3.3), du micro-thesaurus ASA « Régions d'Amérique » par le « bon » micro-thesaurus.

Si l'univers du discours d'une archive n'a pas besoin de ces macro-lieux ou si les macro-lieux physiques ne sont pas pertinents pour contextualiser les objets de connaissance thématisés dans un corpus audiovisuels (comme c'est le cas, par exemple, pour un corpus portant sur des cours en histoire des mathématiques), la séquence 2, alors, sera tout simplement supprimée.

En ce qui concerne maintenant l'identification du lieu géographique plus circonscrit (voir la *deuxième phase* de la tâche de la localisation géographique), nous avons laissé à l'analyste la responsabilité de l'identifier et de le désigner « librement ». Ceci dit, si l'univers du discours d'une archive est suffisamment circonscrit et si on connaît *a priori* tous les lieux qui peuvent servir à une localisation plus précise d'un objet de connaissance thématisé dans un texte audiovisuel, alors la procédure de la description libre peut être remplacée par celle d'une description contrôlée. Dans ce cas là :

1. il faut d'abord *élaborer le thesaurus* des lieux de localisation, c'est-à-dire le thesaurus qui définit *a priori* le rang des valeurs du terme conceptuel [TERRITOIRE] (voir figure 4.3) ;
2. *supprimer* ensuite les activités de description composant la *procédure libre* de description d'un territoire physique ;

3. et enfin les *remplacer par* une *procédure de description contrôlée* composée d'un *micro-thesaurus de lieux plus circonscrits* dans lequel l'analyste choisit le lieu approprié lors de son travail de description d'un texte audiovisuel.

On voit ici bien l'intérêt de l'usage de la procédure de description contrôlée : le surplus de travail en amont (lié à la définition et la constitution d'un ou de plusieurs micro-thesaurus) est largement compensé « en aval » par la réduction du temps d'analyse (qui s'allonge considérablement dans le cadre d'une procédure de description libre).

La description contrôlée permet également de maintenir un niveau maximal d'homogénéité dans les résultats des analyses produites par différents analystes étant donné que les analystes sont obligés d'utiliser des termes prédéfinis pour traiter leurs objets textuels (ce qui n'est évidemment pas le cas si on utilise la procédure de description libre).

Par ailleurs, les termes prédéfinis dans un micro-thesaurus peuvent être *a priori* pourvues d'informations supplémentaires qui enrichissent une analyse concrète. Ainsi, par exemple, toute région saisie dans un micro-thesaurus peut être *a priori* pourvue de coordonnées de localisation (libérant ainsi l'analyste de faire lui-même ce travail qui peut devenir rapidement fastidieux). Elle peut être référencée dans des ressources extérieures du système métalinguistique ASA. Par exemple, telle ou telle région du monde peut être référencée dans la base de données textuelles Wikipédia. Enfin, toute étendue, tout lieu figurant dans un micro-thesaurus ASA peut être classé *a priori* et faire partie d'une classification telle que celle présentée par la figure 6.5 sans que l'analyste soit obligé de procéder par lui-même à une telle classification à chaque fois qu'il saisit un nouveau lieu, une nouvelle étendue.

6.5. La contextualisation par localisation temporelle

Comme déjà dit au début de ce chapitre, la contextualisation par localisation temporelle est une deuxième forme de contextualisation d'un objet de connaissance thématisé dans un texte audiovisuel. Nous distinguons ici deux formes plus spécialisées :

1. la première est celle de la *localisation chronologique* à proprement parler d'un objet de connaissance,
2. la seconde est celle de la *localisation historique*, c'est-à-dire la localisation d'un objet de connaissance à l'aide d'époques composant telle ou telle histoire (histoire humaine, histoire nationale, histoire culturelle, histoire scientifique, etc.).

La figure 3.5 dans le chapitre 3 montre un exemple très simple d'une séquence fonctionnellement spécialisée d'un formulaire d'analyse du contenu audiovisuel qui

propose une analyse succincte de la localisation à la fois chronologique et historique d'un objet de connaissance. Dans notre exemple, il s'agit de la *culture technique* de la *civilisation chavin* au *Pérou* du *4ème siècle avant J.C.* – période qui est appelée parfois <Ere pré-moche> (c'est-à-dire époque historique se situant d'avant l'ère de la civilisation Moche qui apparaît aux alentours du dernier siècle avant J.C. dans une vallée du même nom au Pérou).

Techniquement parlant, cette séquence se compose de trois activités à effectuer par l'analyste :

1. La première activité (obligatoire) se réduit à la simple sélection de l'unique terme conceptuel de la séquence, c'est-à-dire du terme conceptuel [Siècle].
2. La deuxième activité (également obligatoire) consiste en l'identification du ou des siècles pertinent(s) dans une liste chronologique de siècles (liste qui fait partie du thesaurus ASA). Dans notre exemple (chapitre 3, figure 3.5), il s'agit du <4ème siècle> et du <3ème siècle> d'avant J.C.
3. La troisième activité (facultative) consiste en l'éventuelle désignation chrononymique du ou des siècles sélectionnés qui, selon un certain point de vue historique, forment une ère, une époque ou encore une période. Dans notre cas, les deux siècles sélectionnés font partie de l'époque que l'analyste à désigné par l'expression <Ere pré-moche>.

La séquence représentée dans la figure 3.5 (chapitre 3) forme en effet une séquence simple et robuste qui suffit souvent pour localiser avec une certaine précision un objet de connaissance thématisé dans un texte audiovisuel. Mais il existe toute une série de cas où elle ne suffit pas, parmi lesquels les quatre suivants :

1. *chronologie précise* : date précise (d'un événement), période précise (début, fin, durée), échelles de datation, etc. ;
2. *chronologie approximative, approchée* : dates pas connues, dates connues partiellement, indications approximatives de dates, etc. ;
3. *périodisation historique* : ères/époques/périodes, etc. périodisation variable selon tradition historiographique et/ou selon domaine (histoire nationale, etc.) ;
4. enfin, *terme chronologique* qui est en même temps aussi un *terme historique* (par exemple, le *18ème siècle français* peut être interprété à la fois comme une époque historique et une période temporelle, etc.).

Nous avons essayé de spécifier et de développer une collection de séquences de description du contexte temporel d'un objet de connaissance qui permettent de tenir compte de chacun de ces trois cas de figure. Cette collection de séquences distingue, par ailleurs, entre des séquences représentant une configuration simple et prédéfinie (telle que celle représentée par la séquence dans la figure 3.5 au chapitre 3) et des séquences plus complexes qui permettent à l'analyste de procéder à des localisations historiques relativement fines.

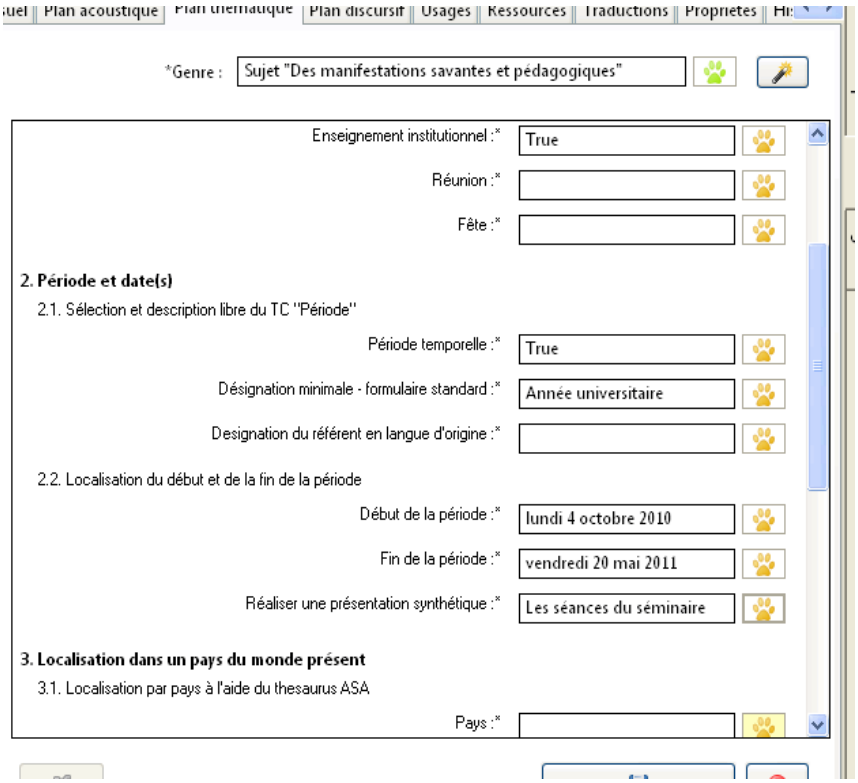

Figure 6.6. *Localisation temporelle par période d'un objet de connaissance thématisé dans un corpus audiovisuel*

Bien évidemment, c'est l'*objectif* d'une analyse d'un corpus audiovisuel qui détermine laquelle ou lesquelles des séquences en question sont les plus appropriées pour répondre aux besoins concrets de l'analyste.

La figure 6.6 montre l'extrait d'un *modèle de description** que nous utilisons pour identifier *périodes* et *dates* précises d'un événement savant (colloque, séminaire, etc.). La séquence de la localisation temporelle intitulée « Période et dates » est ici placée exceptionnellement pour des raisons de démonstration en deuxième position, juste après la séquence de description de l'objet de connaissance « Manifestation savante ou pédagogique ». La séquence en question possède une *structure générale* (elle peut être employée pour la localisation temporelle de tout événement, de toute activité dont on connaît la date du début et la date de la fin) et est d'un emploi parfaitement simple. Cependant, son emploi présuppose que

l'analyste dispose des dates précises d'un événement thématisé dans un texte audiovisuel.

6.6. La contextualisation par époque historique

Le deuxième cas de figure cité ci-dessus est celui de la localisation temporelle approximative. Celle-ci ne se base pas nécessairement sur des connaissances partielles mais souvent sur des connaissances qui, dans le texte audiovisuel soumis à l'analyse, ne sont pas exprimées sous forme d'une date précise. Ainsi, peut-on contextualiser en référence à une époque, à un règne, à un événement historique de référence, à des événements historiques concomitants, etc. Ces « stratégies » de *contextualisation historico-temporelle* peuvent se faire en absence d'une référence chronologique *stricto sensu* (année, siècle, etc.) ou sous une forme mixte : tantôt sous forme d'une référence à une date précise, tantôt sous forme d'une référence qualitative, non-numérique mais compréhensible pour tous ceux qui possèdent les compétences historiques nécessaires.

C'est, par exemple, le cas des dates importantes dans la vie d'un chercheur : une première expédition, une découverte, une rencontre décisive pour son évolution intellectuelle, des échecs et des reconnaissances académiques. Tous ces genres d'événements (qui s'organise, on l'aura compris, selon le fameux schéma narratif de Greimas [GRE 79]), peuvent être accompagnés de références chronologiques mais ne le sont pas obligatoirement – bien loin de là.

Nous avons essayé de tenir compte de ce type qualitatif (et narratif) de contextualisation temporelle dans le cadre de la description des objets de connaissance relevant des domaines d'expertise du projet ASA-SHS. La figure 6.7 montre l'extrait d'un modèle de description de corpus audiovisuels dans lesquels on traite d'une ou de plusieurs langues (familles de langues, dialectes, protolangues, etc.). L'extrait montre la séquence 4 réservée à la contextualisation historico-temporelle de l'objet de connaissance thématisé (dans notre cas, il s'agit de l'évolution des langues du domaine ouest-sémitique dans la période comprise entre le 12 siècle avant J.C. et le 1er siècle après J.C.[44]).

Cette séquence est construite sur une structure assez complexe qui connaît deux sous-séquences (représentées par les deux figures 6.6 et 6.7) dont la deuxième présuppose la première. Cela veut dire que l'analyste doit d'abord renseigner la première sous-séquence avant de renseigner la deuxième. La deuxième sous-séquence à son tour est suivie d'une troisième sous-séquence qui, elle, présuppose de nouveau la première sous-séquence. Autrement dit, seule la première sous-séquence doit être renseignée.

---

[44] Ce sujet a été abordé par le philologue français André Lemaire (Ecole Pratique des Hautes Etudes) lors d'un entretien réalisé en 2005 dans le cadre du programme AAR

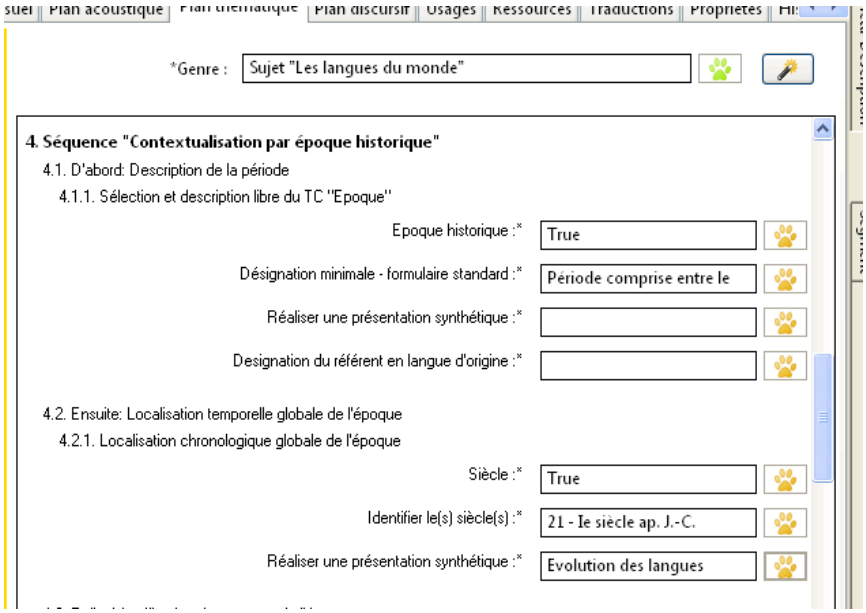

Figure 6.7. *Identification et localisation temporelle d'une époque historique (première partie)*

Dans la première sous-séquence (figure 6.7), l'analyste est invité de désigner l'époque historique dans laquelle se situe l'objet de connaissance thématisé dans le texte audiovisuel soumis à l'analyse. En d'autres termes, l'analyste décide lui-même de l'appellation de l'époque historique. Il s'agit donc d'une procédure dite de *description libre** (voir chapitre 9) du contexte historique pertinent.

La figure 6.9 par contre montre un exemple d'une description contrôlée du contexte historique pertinent. Dans l'exemple concret représenté par la figure 6.9, il s'agit de localiser historiquement un site archéologique français dans l'époque <La France au temps des Celtes> (cette époque, fait partie du micro-thesaurus ASA dédié aux époques nommées de l'histoire de la France).

Revenons à notre exemple représenté par les deux figures 6.7 et 6.8. Une fois que l'analyste a renseigné la première sous-séquence (figure 6.7), c'est-à-dire le fait que son objet de connaissance (dans notre cas concret, *l'évolution des langues de la branche ouest-sémitique*) se situe dans la période historique du dernier millénaire avant J.C., il peut enfin procéder à une localisation chronologique à proprement parler (voir figure 6.8) : dans le micro-thesaurus ASA réunissant une liste (ouverte)

de siècles, l'analyste peut sélectionner le ou les siècles pertinents pour localiser « son » époque.

Dans notre cas, il pourra ainsi sélectionner tous les siècles compris entre le 12ème siècle avant J.C. et le 1er après J.C. mais, il pourra aussi sélectionner seulement les siècles qui sont effectivement pertinents pour son sujet et/ou nommément cités dans le texte audiovisuel soumis à l'analyse.

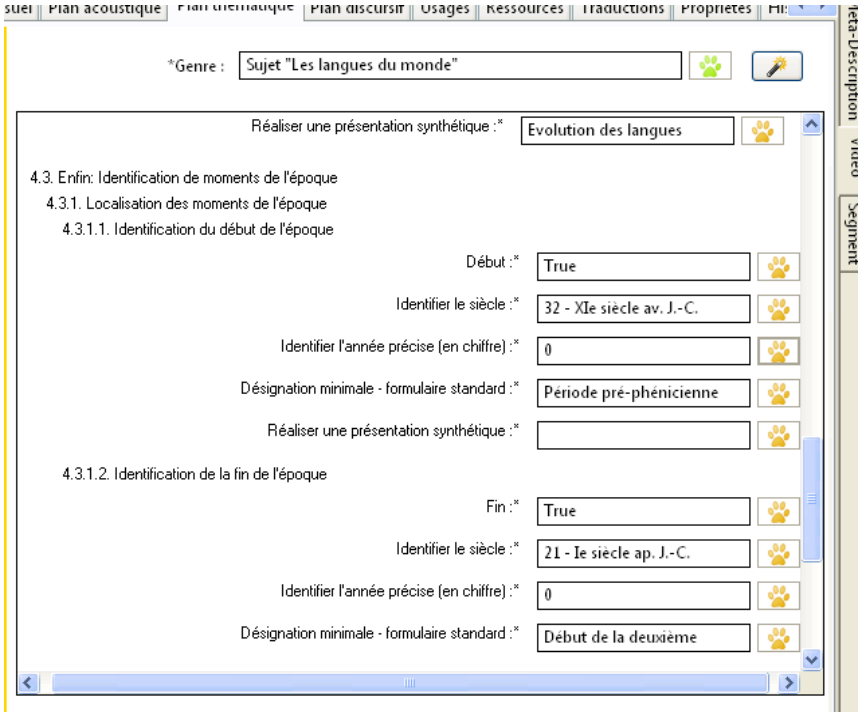

Figure 6.8. *Identification et localisation temporelle d'une époque historique (deuxième partie)*

Enfin, l'analyste peut encore renseigner, s'il le souhaite, *certaines dates importantes* dans une époque identifiée préalablement. La figure 6.8 ne retient que deux grandes dates – le début d'une époque et la fin d'une époque. Mais la figure 6.10 montre une séquence permettant l'identification et l'explicitation d'une ou de plusieurs dates (moments, périodes, etc.) faisant partie d'une époque historique choisie.

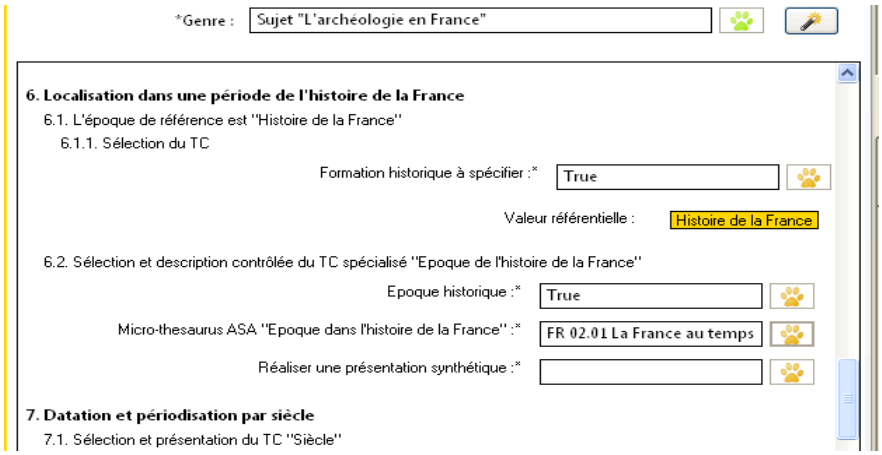

Figure 6.9. *Identification et localisation temporelle d'une époque historique à l'aide de la procédure de description dite contrôlée*

Dans notre cas, l'analyste identifie le <Règne de Louis XIV> comme une période particulière de l'époque appelée <Âge classique français>. Ce règne lui sert de localisateur historique du sujet *Culture française –cultures de France* traité dans le texte soumis à l'analyse (voir figure 6.10).

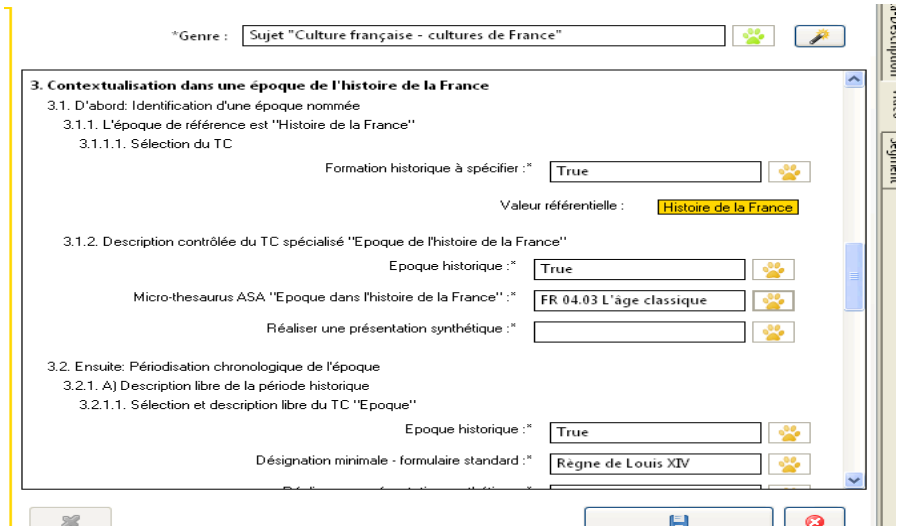

Figure 7.10. *Identification d'une période spécifique faisant partie d'une époque historique*

L'expression <Age classique français> fait partie d'un micro-thesaurus consacré aux époques de l'histoire française (figure 6.11). Bien sûr, pour d'autres histoires nationales ou de pays, des micro-thesaurus appropriés d'époques historiques doivent être utilisés. Ainsi, la figure 6.12 nous montre un micro-thesaurus que nous utilisons pour contextualiser les objets de connaissance dans l'histoire de la Chine.

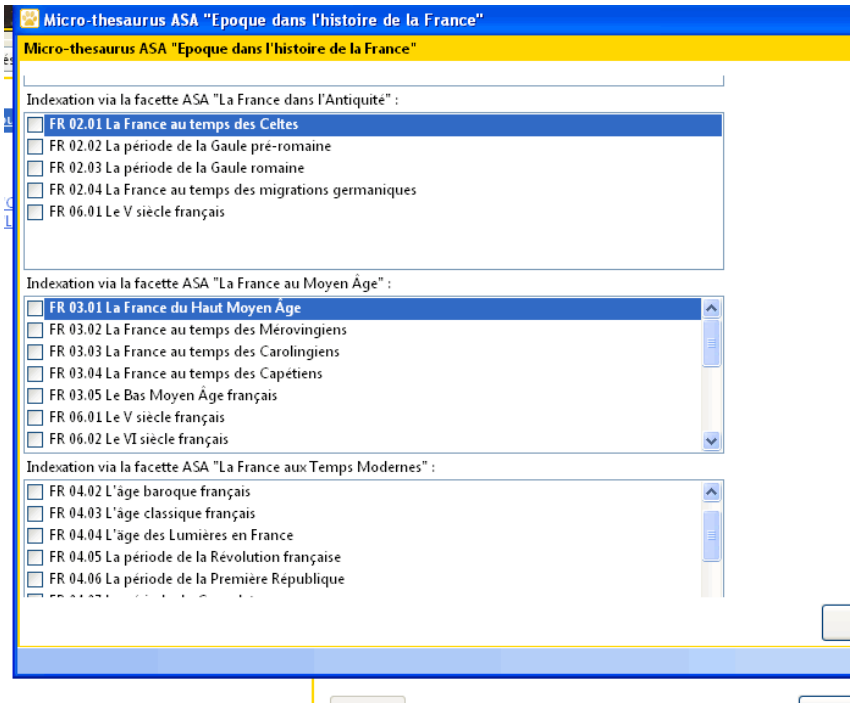

Figure 6.11. *Extrait du micro-thesaurus ASA réunissant les différentes époques dans l'histoire de la France*

6.7. Contextualisation historique et périodisation

La figure 6.10 montre un autre cas de contextualisation chronologique et historique. Il s'agit ici de produire une périodisation sur la base d'une connaissance historique existante qui sert de référence à l'analyste. Dans notre exemple, il s'agit de l'histoire de la France dont nous acceptons une périodisation par époques qui semble prévaloir parmi les spécialistes en la matière. Ces époques sont identifiées dans un micro-thesaurus spécifique qui fait partie du thesaurus ASA (figure 6.11). Bien évidemment, ce micro-thesaurus peut être adapté, voir remplacé ou simplement utilisé d'une manière concurrente avec d'autres micro-thesaurus consacrés aux époques de l'histoire de la France.

Dans l'exemple que nous montre la figure 6.10, le travail de l'analyste est organisé comme suit :

– 1ère étape : l'analyste identifie l'histoire, la formation historique pour laquelle la périodisation existante sous forme d'époques est valide. Dans notre cas, il s'agit de l'histoire de la France.
– 2ème étape : l'analyste identifie la ou les époques qui sont pertinentes pour son travail d'analyse qui consiste à trouver le contexte historique approprié, pertinent pour son travail. Ici, il s'agit de l'époque de l'Âge classique français.

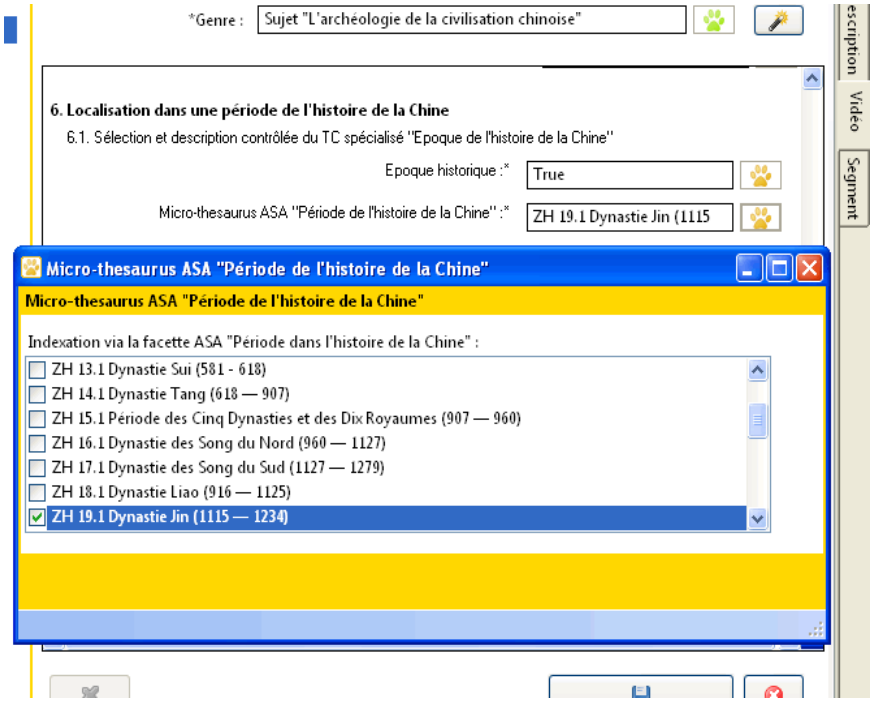

Figure 6.12. *Extrait du micro-thesaurus ASA réunissant les différentes époques dans l'histoire de la Chine*

Ces deux étapes forment une sous-séquence qui est présupposée par toutes les autres sous-séquences composant la séquence 3 destinée à la contextualisation (d'un objet de connaissance thématisé dans le texte audiovisuel soumis à l'analyse) dans l'histoire de la France.

– 3ème étape : l'analyste peut maintenant entreprendre une périodisation plus fine qui est adaptée à son objet d'analyse. Il commence à décrire la période particulière qui est concernée et qui se situe soit à l'intérieur d'une époque soit entre deux ou plusieurs époques figurant parmi les époques nommées retenues dans le micro-thesaurus ASA. Dans notre cas, la période concernée est le règne de Louis XIV qui constitue une période spécifique de l'époque <Âge classique français>.

– 4ème étape : l'analyste procède ensuite comme déjà décrit ci-dessus : il peut – s'il le souhaite – positionner chronologiquement ou simplement d'une manière qualitative la période considérée (c'est-à-dire le règne de Louis XIV) et/ou telle ou telle date particulièrement importante de celle-ci (ici ne sont retenues que les dates du début d'une période et de sa fin mais le modèle de référence sous-jacent peut intégrer facilement d'autres types fonctionnels de dates).

Même si la tâche d'analyse de la contextualisation historico-temporelle semble être assez complexe, la construction des modèles pouvant supporter ce genre d'expertise ne l'est pas. En fin de compte, seulement la partie 3.1 dans la figure 6.10 doit être adaptée : choix du type d'histoire et, puis, choix du micro-thesaurus comportant les époques nommées pour le type d'histoire en question. La figure 6.12 montre, par exemple, l'usage d'un micro-thesaurus contenant les principales époques (reconnues) de l'histoire de la Chine. Il remplace le micro-thesaurus contenant les époques nommées de l'histoire de la France. Le reste du modèle de description demeure identique.

Le but de cette section a été de montrer de voies possibles pour dépasser la simple localisation chronologique des objets thématisés dans un texte audiovisuel en essayant de fournir des outils pour les positionner également dans « leurs » contextes historiques appropriés tout en tenant compte du fait que la périodisation dépend du type de l'histoire (et de l'approche historiographique) et qu'il n'y a pas une et une seule périodisation « vraie » de l'histoire.

6.2. La contextualisation thématique

Cette forme de contextualisation permet, comme déjà dit dans l'Introduction (section 6.1) à ce sixième chapitre, d'expliciter le cadre social ou institutionnel, culturel ou historique, épistémique ou mental d'un objet de connaissance traité dans un corpus audiovisuel. Ce type de cadre est souvent appelé *cadre thématique*. Nous préférons utiliser le terme *cadre axiologique* pour insister sur le fait qu'il s'agit ici de l'*environnement pertinent* selon lequel un objet de connaissance prend son sens, c'est-à-dire son intérêt, son enjeu pour le texte ou le corpus de textes dans lequel il en est question. En parlant de *cadre axiologique*, nous souhaitons ainsi renforcer

cette dimension de valeur qui est intrinsèquement attachée à tout objet thématisé dans un texte.

Comme déjà dit dans l'introduction à ce chapitre, cette forme de contextualisation n'a pas été systématiquement développée dans les limites de nos recherches. Nous l'avons surtout utilisé dans le cadre de la description de corpus audiovisuels qui font partie du fonds des archives audiovisuelles de la Fondation Maison des Sciences de l'Homme à Paris – fonds composé essentiellement d'entretiens avec des chercheurs, d'enregistrements de séminaires de recherche ou encore de colloque ou autres manifestations scientifiques.

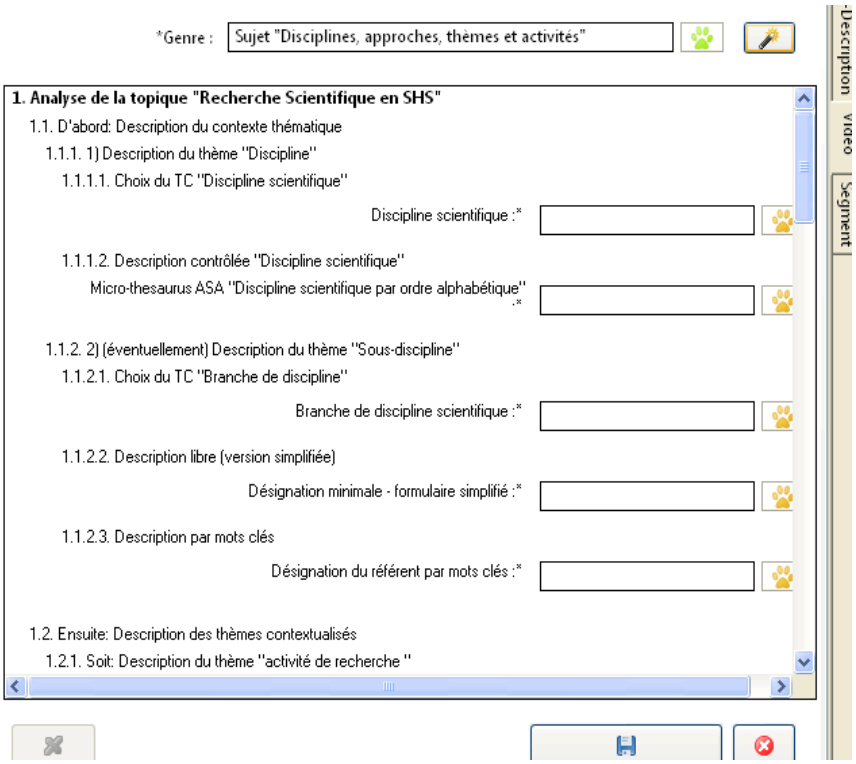

Figure 6.13. *Le thème [Discipline] dans sa fonction de contexte axiologique (thématique)*

Dans ce grands fonds de quelques milliers d'heures de vidéos numériques, l'un des sujets les plus récurrents est, bien sûr, la présentation d'une discipline ou d'une approche disciplinaire comme, par exemple, l'anthropologie historique, la sociologie cognitive, l'histoire des sciences, l'urbanisme ou encore la muséologie qui forment – comme tant d'autres disciplines plus ou moins reconnues faisant partie des sciences

sociales et humaines – le *cadre épistémologique et théorique* pour les activités de recherche de tel ou tel chercheur ou chercheur-enseignant, l'identification des problématiques pertinentes à la recherche, la définition des enjeux et des objectifs de la recherche, l'approche et le traitement de l'objet de la recherche, etc.

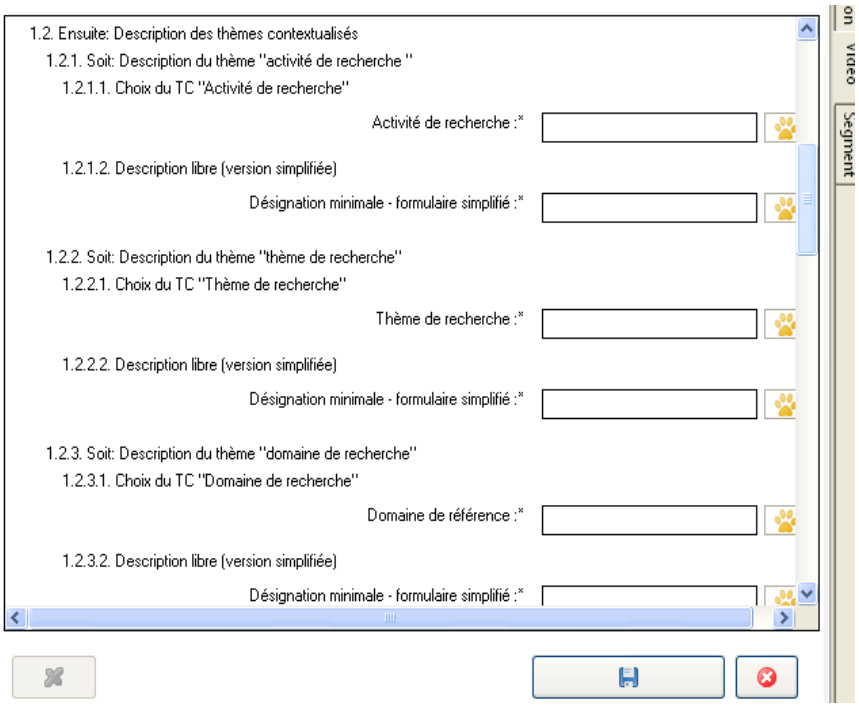

Figure 6.14. *Les thèmes [Activités de recherche], [Thèmes de recherche], etc. dans leur fonction de thèmes contextualisés*

Les figures 6.13 et 6.14 montrent comment nous traitons cette question dans notre approche de l'analyse de corpus audiovisuels dans laquelle est thématisée ce type de sujet. Dans le cadre, par exemple, de la description du contenu d'un entretien avec un chercheur où il est question des thèmes, domaines, objectifs, activités, résultats, conséquences, etc. de recherche d'un chercheur (chercheur-enseignant, chercheur-étudiant, etc.), l'analyste est invité à préciser d'abord le cadre thématique qui constitue le contexte de référence pour ces questions (figure 6.13). Dans notre cas, le cadre thématique est constitué par le terme conceptuel [Discipline] que l'analyste doit préciser en sélectionnant une ou plusieurs disciplines réunies au sein d'un micro-thesaurus. Il peut éventuellement déterminer davantage ce cadre axiologique en explicitant – via une procédure de description libre – une ou plusieurs « sous-disciplines » concernées. Bien évidemment, cette sous-séquence

servant à la description du cadre thématique selon ou « dans » laquelle on traite un ensemble de questions, peut être enrichie par une structure topique bien plus développée – tout dépend ici de nouveau de la finalité de l'analyse.

Une fois le cadre thématique ou axiologique (dans notre cas : l'approche disciplinaire) renseigné, l'analyste, dans un deuxième temps, doit informer les différents sujets traités en référence au dit cadre thématique. Comme on peut le voir dans la figure 6.14, il s'agit de sujets représentés par des termes conceptuels tels que [Activité de recherche], [Thème de recherche], [Domaine de recherche], [Objectifs de recherche], etc. Selon le texte audiovisuel à décrire et la finalité de la description, l'analyste renseignera l'un ou l'autre ou tous ces termes conceptuels formant une structure topique simple mais assez robuste.

La distinction entre un cadre thématique composé, au minimum, par le terme conceptuel [Discipline] et les thèmes ou sujets contextualisés en référence à ce cadre (tels que ceux représentés par les termes conceptuels comme [Activité de recherche], [Thème de recherche], etc.) acquiert toute sa pertinence dès lors qu'il s'agit d'analyser des corpus audiovisuels dans lesquels les cadres thématiques ou axiologiques eux-mêmes semblent être très proches, voire partiellement similaires. C'est le cas, par exemple, lorsque l'on traite un corpus d'entretiens de chercheurs travaillant sur un objet commun tel que celui de la *culture* mais l'abordant des *points de vue* (parfois seulement légèrement) différents : point de vue de l'histoire culturelle, point de vue de l'anthropologie culturelle, point de vue de la sociologie des cultures, point de vue de la sémiotique des cultures ou encore point de vue de la philosophie de la culture.

La figure 6.13 représente, comme déjà dit, une structure topique très simple qui définit un type de cadre thématique ou axiologique. Nous savons très bien que le point de vue « disciplinaire » n'est jamais « théoriquement neutre » mais dépend toujours d'une certaine vision (théorique, philosophique, etc.) sous-jacente. Ainsi, en parlant du point de vue de l'histoire culturelle, on suppose que ce point de vue est, *in fine*, déterminé par une référence à un auteur, à une école de pensée, un paradigme scientifique, etc.

Telle quelle, la structure topique représentée par la figure 6.13 ne permet pas de rendre compte de ces nuances très importantes dans une expertise scientifique. Cependant, grâce à la structure modulaire, il est relativement simple d'ajouter à la séquence 1.1 (figure 6.13) une troisième sous-séquence dédiée à l'explicitation de références intellectuelles ou théoriques qui déterminent la position du chercheur et de son travail dans telle ou telle discipline.

Pour terminer ces quelques réflexions au sujet de la distinction entre *cadre thématique* et *thème contextualisé* – une troisième forme donc de contextualisation d'un objet thématisé dans un texte soumis à l'analyse – ajoutons que celle-ci est, à notre avis, d'une grande importance pour la bonne conduite d'une activité telle que

celle de la veille, de l'expertise et de la restitution de l'information (audiovisuelle ou autre). En effet, il nous semble que la *valeur ajoutée* d'une information quelconque réside davantage dans ses *références à un cadre axiologique* que dans sa dimension dénotative ou référentielle.

Or, c'est une des pistes de recherche que nous voudrions explorer davantage dans les années à venir, notamment en référence à une *sémiotique des cultures*, c'est-à-dire à une sémiotique du *cadre des références* [TAY 98] qui sert de *standard* [STO 88, STO 89, STO 05] à un personne ou un groupe social pour « lire », « interpréter », « comprendre » et, enfin, « exploiter » (ou « consommer ») un texte audiovisuel comme un *bien*, comme une ressource cognitive.

Chapitre 7

La tâche d'analyse de la mise en discours d'un sujet

7.1 Introduction

Voici maintenant le cas d'une tâche de description qui, techniquement parlant, repose sur une classe de configurations qui peuvent devenir très complexes. Il s'agit ici d'une tâche de description très importante, voir même en fin de compte *cruciale* dans le cadre d'une *description thématique** (ou du contenu) d'un *texte audiovisuel**. Elle constitue un des moments décisifs de l'analyse au sens d'une *expertise du contenu*, c'est-à-dire de l'évaluation structurée et explicite de la *valeur discursive* d'une information (voir aussi [STO 99, STO 01]), de la façon dont une topique est abordée dans un texte, un document audiovisuel.

Il existe une différence évidente entre le fait d'affirmer que dans deux textes audiovisuels composant un corpus d'analyse on parle d'une fouille archéologique concrète, dans tel lieu physique consacré à telle époque et à tel type de vestiges, et le fait de comment le « même » sujet est abordé dans les deux textes – la première affirmation relève en effet de la *description référentielle**, la seconde relève de la *description discursive* à proprement parler.

Ce chapitre est consacré à une brève présentation de la description de la mise en discours d'un sujet. Comme déjà expliqué au chapitre 5, il s'agit de comprendre et d'expliciter comment un thème est pris en charge par l'*auteur* ou, plutôt, par son *énonciateur*, c'est-à-dire par le rôle qui signe responsable de ce qui est dit à propos d'un objet de connaissance donné. Tout comme la description référentielle d'un thème (c'est-à-dire tout comme la description d'une structure topique), la

description de la mise en discours repose également sur des modèles de description. Ces modèles font partie de la classe des configurations dites discursives.

Rappelons-nous de la figure 5.1 (chapitre 5) qui nous montre que la description thématique (du contenu) d'un corpus de textes audiovisuels repose sur l'explicitation de sa *structure thématique** (dans ce sens on parle ici aussi de la *description thématique**). Cette structure thématique est constituée, d'une part, de la *configuration topique** permettant de procéder à la description référentielle du contenu (à l'analyse de ce dont parle le texte) et, d'autre part, de *configurations discursive*, d'expression verbale et/ou audiovisuelle* et *méta-textuelle*. La prise en compte des configurations restantes permet donc d'affiner et de compléter la description référentielle, selon le besoin et les objectifs à réaliser, par une description de la mise en discours (de la *discursivisation* [GRE 79]) du ou des thèmes abordés, par la description des modalités et stratégies d'expressions verbales ou audiovisuelles et encore par l'explicitation du cadre de référence de l'analyse elle-même.

La section 7.2 est consacrée à une brève présentation synthétique des principales procédures de mise en discours que nous avons prise en considération dans notre travail de recherche.

Dans la section 7.3, nous discuterons plus en avant la tâche de la description de la mise en discours d'une topique et de sa place dans l'analyse thématique (du contenu) d'un texte ou d'un corpus de textes audiovisuels.

La section 7.4 est réservée à la présentation de quelques exemples concrets de la description de la mise en discours d'une topique.

Enfin, dans la section 7.5, nous développerons encore quelques idées prospectives au sujet de l'importance de la prise en compte d'une analyse du discours (ou, plutôt, de la mise en discours d'une topique) pour la veille d'information et, plus particulièrement, pour pouvoir *tracer* l'information.

7.2. Procédures de mise en discours

La tâche de la description de la mise en discours d'une topique (voir aussi [STO 01]) s'occupe du discours qui traite et « porte » un sujet, en fonction des intentions de son auteur, des connaissances de son objet et, enfin, de son public.

En dehors du fait de déterminer de quoi parle un texte (= *description référentielle** du sujet ou du thème), il s'agit ici de vouloir rendre compte de comment on en tient compte dans un *discours* donné et véhiculé par un texte ou un corpus de textes audiovisuels. La littérature scientifique au sujet de l'analyse du

discours est très importante et il est difficile d'en avoir une vision synthétique et appropriée. Les travaux des chercheurs en linguistique, sémiotique ou pragmatique sur le discours et la « prise en charge » du contenu référentiel par le discours montrent cependant le très grand intérêt que possède l'analyse du discours pour l'analyse du contenu de corpus audiovisuels qui, réduite à sa simple dimension référentielle, reste très souvent stérile et insatisfaisante.

Tour en étant conscient de la grande diversité et richesse de la recherche dans le domaine de l'analyse du discours, nous nous sommes concentrés, pour le moment, sur un petit ensemble de stratégies ou procédures dites de mise en discours d'un sujet. Les voici sous forme de six questions générales :

1) *Qui est l'auteur « véritable » (c'est-à-dire l'énonciateur) du sujet traité, développé dans le texte ?* Autrement dit, qui est la véritable source du contenu et qui en est responsable (pour sa *vérité*, par exemple, sa *nouveauté* ou son *originalité*). Etre capable de répondre à cette question est important notamment lorsqu'il s'agit de citations directes ou indirectes, de références à ce que « quelqu'un aurait dit », à des courants de pensée, à des textes publics ou non, etc., dans le contexte de citations. L'explicitation de la structure de l'énonciation du sujet traité dans le texte audiovisuel est particulièrement enrichissante dans le cadre de l'analyse d'entretiens, de témoignages, de récits de vie ou encore de récits d'expériences (voir aussi nos explications dans [STO 01]). L'analyse de l'énonciation inclut également l'explicitation du destinataire ou encore du public auquel le texte est adressé. Ici, le destinataire ou le public est un rôle qui fait partie du discours (parfois, on l'appelle d'ailleurs l'*énonciataire* et on le considère comme un rôle complémentaire nécessaire à l'énonciateur). Autrement dit, ce sont des indices apparents dans le discours, ou à inférer à partir du discours, qui peuvent montrer à qui est adressé le sujet abordé par l'énonciateur. La prise en compte de l'énonciataire, du destinataire du contenu dans un texte permet d'élucider deux problèmes importants pour toute communication :

– le premier problème est celui de l'évaluation de la *congruence* entre le *public empirique, « réel »* d'un texte audiovisuel et le *destinataire visé* par le même texte (c'est-à-dire le « destinataire idéal », selon Eco [ECO 79]) ;
– le deuxième problème est celui de la prise en compte des *variations contextuelles* de l'*engagement de l'énonciateur* à propos de la vérité, de l'originalité, de la précision, etc. du sujet dont il parle. Comme on le sait, ces variations contextuelles peuvent jouer un rôle déterminant dans la construction du *discours de la vulgarisation scientifique* et aussi dans celle du *discours de groupe (social)*, du *discours dit communautaire*.

2) *Quel est le point de vue selon lequel l'auteur aborde son sujet ?* Cette deuxième question est intimement liée à la première mais désigne une problématique différente. Il s'agit ici de rendre explicite la *position (cognitive)* à

partir de laquelle l'auteur d'un texte audiovisuel développe son sujet. Par exemple, l'auteur peut développer son sujet à partir de sa propre position, qu'il assume telle quelle (il s'assume à la fois dans son rôle d'auteur et dans celui d'énonciateur). Mais comme nous venons de le voir (voir *supra*, première question), il peut également se référer à un autre énonciateur. Et dans ce deuxième cas, l'auteur peut exprimer une attitude fort variable face à la position défendue par l'énonciateur-autre-que-lui : de l'accord complet jusqu'à l'opposition extrême en passant par toutes les formes intermédiaires de neutralité, de réserve, etc. (voir à ce propos le travail de recherche très instructif de B.-N. Grunig et R. Grunig [GRU 98]). L'explicitation de la position de l'auteur peut s'avérer importante pour comprendre, par exemple, des *points de vue en compétition* sur un sujet donné ou encore pour distinguer, dans un entretien avec un chercheur d'une part le discours du scientifique et d'autre part le discours du militant. En effet, il arrive assez régulièrement que, dans le cadre d'une même discussion ou d'un même entretien, la ou les personnes interviewée(s) *change(nt) de posture* à plusieurs reprises et, cela, d'une manière plus ou moins radicale et sans se soucier de la cohérence globale de leur discours.

Ainsi par exemple, dans son rôle d'expert ou de spécialiste d'un groupe ou d'un mouvement social, le chercheur peut évoluer à la fois comme *sujet autonome qui sait ou pense savoir*, comme s*ujet délégué d'une instance supérieure de vérité* qui tire sa légitimité scientifique ou d'expert de son appartenance à un courant de pensée ou à une théorie scientifique, comme *sujet de conviction* qui s'octroie le droit de passer d'un discours descriptif à un discours moralisateur, comme *sujet empathique* qui fait sien la position du groupe ou du mouvement social étudié pour militer en sa faveur ou défaveur, et ainsi de suite.

3) *Quel est le type ou genre de discours utilisé par l'auteur pour développer son sujet ?* Il s'agit ici (dans la mesure du possible) d'expliciter si le traitement et le développement d'un sujet dans un corpus audiovisuel se fait sous forme, par exemple, de descriptions, de narrations, d'explications ou encore sous forme d'exposés, de témoignages, voire sous forme de débats, d'enseignements, de mises en scène ludiques, de simulations, etc. Il faut ici déterminer l'acte caractérisant le discours que consacre l'auteur à son sujet. Cet acte peut appartenir aussi bien à tel ou tel type d'*actes élémentaires de langage* identifiés et traités par la pragmatique linguistique ou la linguistique du discours [AUS 70, SEA 72] ou à un genre social et culturel plus complexe de discours ou de textes (voir par exemple [ADA 90, ADP 89, STO 01]).

En tout cas, tenir compte de cette troisième question est important dans la mesure où elle permet de préciser si un sujet concret est développé sous forme d'une narration, d'un exposé argumenté ou encore sous forme d'une série d'exemples illustratifs, voire d'un jeu dialogique composé de questions et de réponses pédagogiquement motivées. On voit bien que l'intérêt – la valeur ajoutée – d'un sujet peut être très fortement déterminé par le type ou le genre de discours qui le

« porte » : d'un côté, le genre de discours peut le prédisposer à des usages plutôt scientifiques ou experts, de l'autre à des usages plutôt éducatifs.

4) Comment l'auteur (l'énonciateur) *cadre-il* le domaine référentiel qui forme le sujet de son discours ? C'est la question du *focus d'attention* qui est posée ici, c'est-à-dire celle qui pose le problème de l'identification et de l'explicitation des aspects particuliers de l'objet thématisé dans le discours que l'auteur (l'énonciateur) souhaite mettre en avant, souligner, traiter en priorité et, enfin, transmettre à, partager avec, faire accepter par son public. Le *cadre discursif* d'une topique organise, d'une manière analogue au cadre visuel, une topique en différentes régions spécifiques tels que l'*avant-scène* (c'est-à-dire les aspects sur lesquels portent l'attention de l'auteur ou de l'énonciateur du discours) et l'*arrière-fond* (c'est-à-dire tous les autres aspects qui constituent un réservoir que l'auteur-énonciateur peut exploiter tout au long de son discours). Une autre distinction, cruciale ici, est celle entre le *thème* et le *rhème* qui joue en rôle central dans le développement, la progression du discours : le *thème* étant la partie (ou l'aspect) de la topique qui est *posée* par l'auteur-énonciateur du discours et le rhème étant la partie qui forme le propos que l'auteur-énonciateur souhaite développer (le rhème est ce qui est considéré comme *nouveau* pour un public, ce que l'auteur-énonciateur veut faire apprendre ou comprendre au public et pour lequel il cherche son adhésion).

Si, par exemple, dans deux textes audiovisuels différents, il est question de la même personne, les informations fournies à propos de cette dernière ne sont pas obligatoirement identiques. Très simplement dit, dans le premier texte, l'auteur peut se concentrer sur les relations sociales de la personne servant de référence à son discours, et dans le deuxième texte, l'auteur peut se concentrer, en revanche, sur la carrière professionnelle de la même personne. Une même structure topique sous-jacente (un même modèle conceptuel du domaine référentiel qui est ici l'objet de connaissance « personne ») est exploitée différemment dans les deux textes. Cette exploitation consiste dans le choix du ou des thèmes qui, d'une part, constituent l'avant-scène d'un discours sur une personne donnée et qui, d'autre part, servent à l'auteur-énonciateur du discours pour développer son propos, son « message ». Ainsi, même si deux textes audiovisuels produisent des informations sur la carrière professionnelle d'une même personne, le propos (techniquement parlant, le *rhème*) peut varier très considérablement : l'aspect de la topique « carrière d'une personne » constituant le *thème discursif* sélectionné dans deux discours différents, il peut servir de thème à l'auteur du premier discours pour développer un propos (un rhème) à *contenu moralisateur* et de thème à l'auteur du deuxième discours pour développer un *propos purement chronologique ou biographique*. C'est un exemple, qui simplifie une réalité textuelle autrement plus riche mais il nous montre de quoi il s'agit lorsque l'on parle du *cadrage discursif :* c'est l'auteur (l'énonciateur) qui sélectionne (d'une manière consciente ou non) certains aspects (au détriment d'autres) d'une structure topique pour aborder son sujet afin de réaliser son intention de communication en fonction du contexte de communication dans lequel il est engagé, et en fonction de sa compétence d'auteur. La prise en compte de cette

procédure peut nous aider à mieux cerner la place spécifique occupée par un texte dans un corpus de textes audiovisuels traitant – *grosso modo* – le même sujet.

5) *Quelle est le niveau de précision (de « granularité ») du traitement du sujet abordé par l'auteur ?* Tout en étant indépendante, cette cinquième question est fortement liée à la quatrième (la sélection discursive des facettes pour aborder un sujet) et à la sixième (voir ci-après, la question de la pondération discursive). Expliciter le niveau de précision d'un texte est un moyen de lui assigner une position pragmatique spécifique à l'intérieur d'un *corpus de textes audiovisuels* traitant d'un même sujet : certains textes peuvent être considérés comme des textes spécialisés, d'autres comme des textes généralistes, d'autres textes encore comme des textes possédant un profil de granularité circonscrite du point de vue référentiel.

6) *Comment le sujet est-il développé tout au long du texte audiovisuel analysé ?* Tout en renvoyant de nouveau à la quatrième question, à celle du *cadrage discursif* d'une topique, celle-ci aborde plus particulièrement la question de la logique et des stratégies de la *progression* syntagmatique d'un thème et de son propos dans un discours donné. Comme on le sait, le développement purement linéaire d'un propos consistant à ajouter à un thème sélectionné de plus en plus d'informations « nouvelles » qui l'enrichissent selon les intentions de l'auteur du discours et en fonction de son public, est plutôt un cas limite.

Le développement d'un thème dans son discours peut constituer pour l'auteur le « prétexte » pour aborder d'autres thèmes ou sujets ; le développement lui-même peut prendre la forme d'un « approfondissement d'un domaine de référence » mais ce n'est qu'une forme de développement discursif parmi bien d'autres. Par exemple, même dans le discours dit scientifique, on rencontre régulièrement le cas où la spécialisation et l'enrichissement cognitifs d'un thème posé sont remplacés ou « complétés » – un peu comme dans le discours *politique* ou *didactique* – par des *figures de répétition* et d'insistance de toute sorte (de reprises pures et simples, de variations terminologiques, des variations autour d'un thème posé, de résumés de ce qui a été dit, d'amplifications, d'exemples servant à démontrer l'intérêt ou l'importance du thème posé, etc.).

Dans le cadre de nos recherches sur l'*univers du discours** des archives audiovisuelles, nous avons commencé à intégrer, dans le processus d'analyse de corpus audiovisuels, la problématique de la description de la mise en discours d'une topique. Celle-ci constitue sans aucun doute un enrichissement très considérable aux pratiques de description/indexation habituelles de corpus de textes (numériques). L'apport de l'analyse du discours se fera ressentir surtout, à notre avis, dans le cadre de la veille d'informations ainsi que dans celui de la republication de corpus audiovisuels afin de les adapter à des contextes d'usages spécifiques.

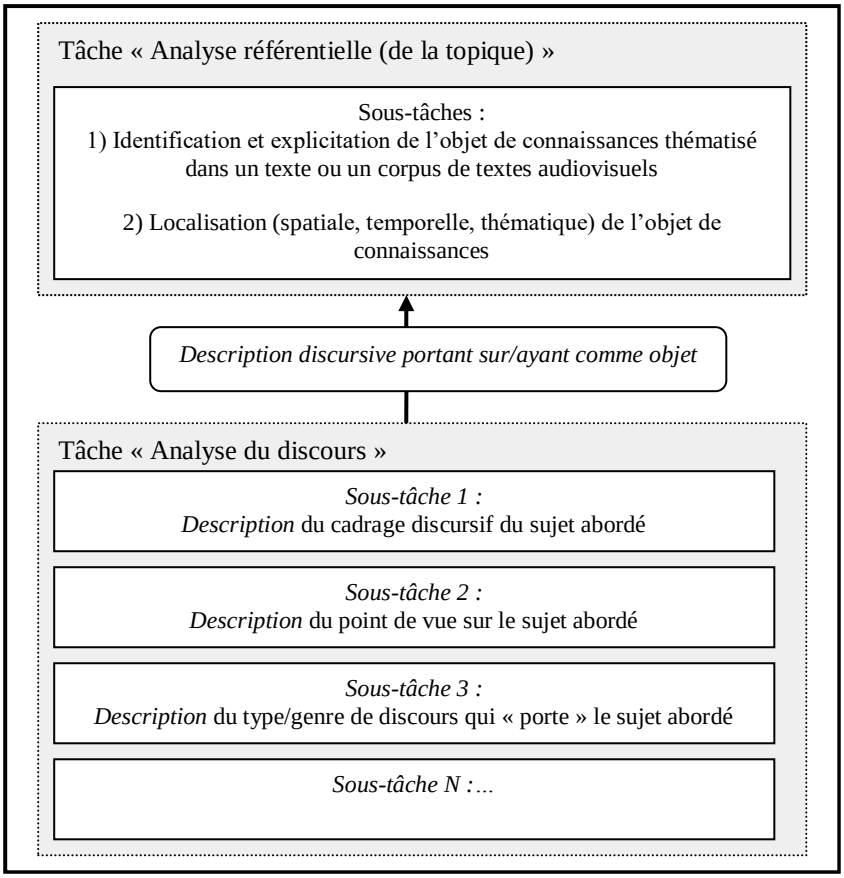

Figure 7.1. *Vue synthétique montrant la place de l'analyse discursive dans l'analyse du contenu d'un corpus ainsi que certaines de ses sous-tâches*

Ceci dit, la prise en compte de cette strate dans l'organisation structurale générale d'un texte exige des connaissances réelles en analyse du discours et des compétences professionnelles difficiles à trouver sur le « marché du travail » car peu enseignées dans les établissements devant former les analystes de corpus (audiovisuels ou autres).

7.3 Anatomie de la description de la mise en discours d'un sujet

La figure 7.1 montre comment s'intègre, dans le processus de l'analyse du contenu, l'analyse de la mise en discours d'un sujet. Une fois la description référentielle du domaine de référence ou d'expertise achevée, l'analyste peut – si cela est pertinent pour lui – compléter son analyse purement référentielle du contenu par une analyse du discours lui-même, c'est-à-dire par l'analyse du traitement du sujet de l'auteur (l'énonciateur) du discours.

En tout cas et comme le montre la figure7.1, la description de la mise en discours d'une topique *présuppose obligatoirement* sa description référentielle. La description de la mise en discours constitue une tâche spécifique de l'analyse thématique d'un texte ou d'un corpus de textes audiovisuels et se décompose elle-même en plusieurs tâches plus spécialisées. La figure 7.1 en identifie les trois suivantes : 1) description du *cadrage discursif*, 2) description *du point de vue* et enfin 3) description *du genre de discours*.

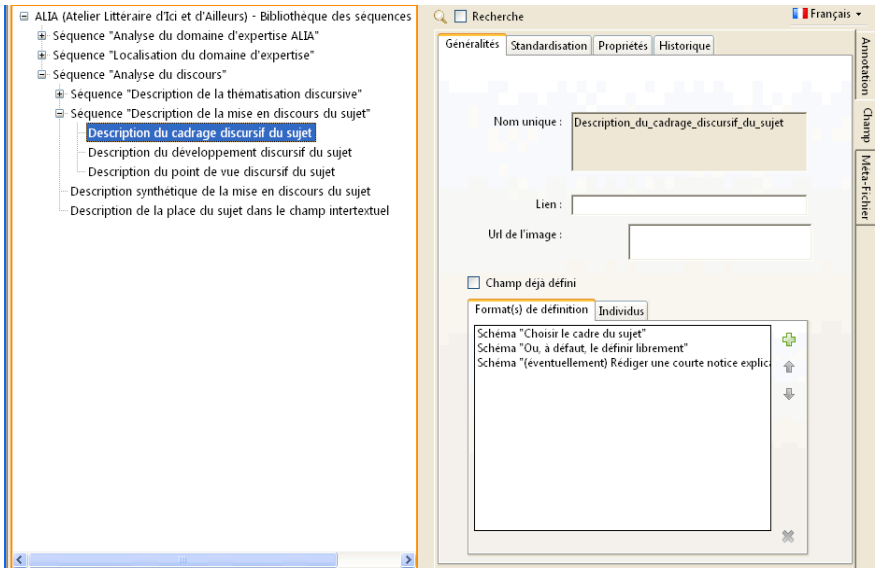

Figure 7.2. *La bibliothèque des séquences pour l'analyse de l'univers du discours de l'archive ALIA – focus sur la collection spécialisée des séquences « Analyse du discours »*

Ces trois sous-tâches correspondent aux questions 2, 3 et 4 identifiées ci-dessus (section 7.2). Elles ont été définies et implémentées dans le cadre du projet ASA-SHS (voir [LEM 11a]) avec d'autres sous-tâches propres à l'analyse du discours (notamment celles de la *thématisation discursive* et du *développement discursif*). Ces sous-tâches forment une petite bibliothèque de séquences spécialisées intégrée

dans la bibliothèque des séquences d'analyse dont dispose l'analyste d'une archive donnée. Rappelons-nous, qu'une *séquence** constitue une *brique** d'un modèle de description d'un texte ou d'un corpus de textes audiovisuels. Ici, il s'agit donc de séquences fonctionnellement spécialisées en l'analyse de la mise en discours d'une topique.

La figure 7.2 nous montre plus particulièrement la bibliothèque des séquences « Analyse du discours » réservée à l'analyse de l'univers du discours de l'archive ALIA développé dans le cadre du projet ASA-SHS. Une sous-tâche spécifique de l'analyse du discours correspond, dans des cas simples, à une séquence. Certaines sous-tâches exigent, cependant le concours de plusieurs séquences. Comme le montre l'exemple *Description du cadrage discursif du sujet* (figure 7.2), une séquence définissant une sous-tâche (ou une partie d'une sous-tâche) de l'analyse du discours, est elle-même composée d'un ou de plusieurs schémas conceptuels (rappelons nous qu'un *schéma conceptuel** est composé d'un ou de plusieurs termes conceptuels faisant partie du méta-lexique ASA[45]). En ce qui concerne notre exemple, trois schémas définissent la séquence *Description du cadrage discursif du sujet*. Il s'agit :

1. du schéma *Choisir le cadre du sujet :* ce schéma propose à l'analyste sous forme d'un micro-thesaurus spécialisé, une liste de cadres discursifs prédéfinis parmi lesquels il choisit celui qui se rapproche le plus au cadre discursif déployé dans le texte soumis à l'analyse ;
2. du schéma *Ou, à défaut le définir librement :* ce schéma offre à l'analyste de définir lui-même le cadre discursif selon lequel le sujet est traité ;
3. et du schéma *(Eventuellement) Rédiger une courte notice explicative* : ce schéma permet à l'analyste d'ajouter une remarque explicative.

Précisons que la définition – très modulaire – de la séquence *Description du cadrage discursif* peut être adaptée au contexte et à la spécificité de l'univers du discours d'une archive audiovisuelle. Autrement dit, les trois schémas utilisés pour définir la séquence *Description du cadrage discursif* dans le cadre de l'analyse de textes audiovisuels pour l'archive ALIA, peuvent être, dans le cadre d'une analyse de textes audiovisuels faisant partie d'une *autre* archive, partiellement remplacés par d'autres schémas ou encore complétés par des schémas non retenus ici. Aussi la définition de tel ou tel schéma conceptuel concerné, voire même la définition du ou des micro-thesaurus spécialisés et utilisés dans la séquence ci-dessus qui nous sert d'exemple, peut varier d'une archive à une autre, c'est-à-dire de l'analyse d'un univers du discours à un autre.

Cet exemple nous montre la grande adaptabilité de l'approche modulaire proposée ici pour rendre compte à la fois de la spécificité intrinsèque d'un corpus

[45] Une présentation détaillée du méta-lexique ASA est proposée dans les chapitres 11, 12 et 13 de ce livre

audiovisuel donné et des attentes spécifiques des publics potentiels face au même corpus.

7.4 Exemples illustrant l'analyse de la mise en discours

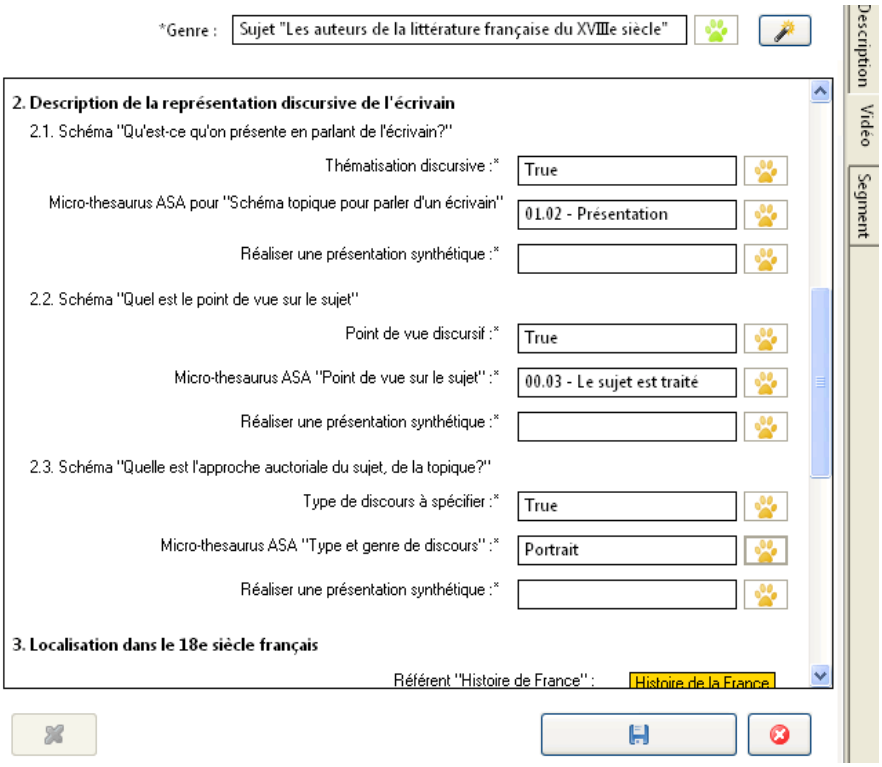

Figure 7.3. *La séquence réservée à la description discursive de la topique* Auteurs de la littérature française du XVIII siècle

Une façon assez simple de prendre en considération le discours consacré à un sujet spécifique nous est présentée par la figure 3.6 (chapitre 3). La figure nous montre l'extrait de l'interface d'un modèle de description de corpus audiovisuels faisant partie des archives ARC coordonnées par Elisabeth de Pablo [DEP 11a]. L'analyste peut, s'il le souhaite, affiner son analyse purement référentielle (portant sur la culture technique de la *civilisation chavin* qui est une civilisation précolombienne) par une analyse du discours qui se concrétise, dans notre cas, sous forme de deux sous-tâches spécialisées :

– sous forme de la sous-tâche *Thématisation discursive* et,

– sous forme de la sous-tâche *Point de vue discursif*.

Chacune des deux sous-tâches peut être réalisée d'une manière simple à l'aide d'une sélection d'une ou de plusieurs valeurs appropriées dans les micro-thesaurus correspondants et intégrés dans leurs interfaces de travail respectifs (voir chapitre 3, figure 3.6).

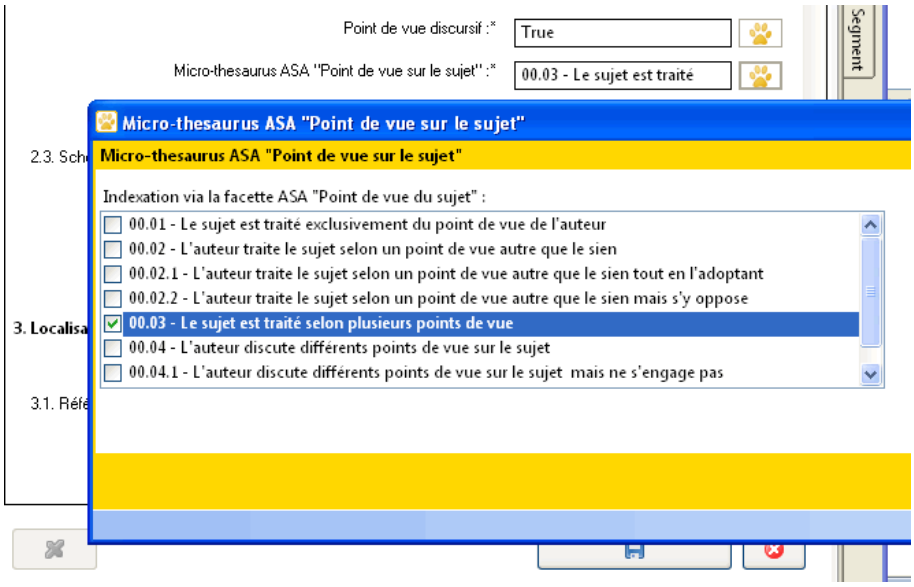

Figure 7.4. *Extrait du micro-thesaurus consacré à l'analyse du point de vue sur le sujet thématisé dans un texte audiovisuel*

La figure 7.3 montre un autre exemple de l'analyse de la mise en discours d'une topique. Elle représente l'interface de travail d'un modèle réservé à la description de corpus audiovisuels dans lesquels on « parle » d'auteurs de la littérature française du XVIIIe siècle. Ce modèle fait partie de la bibliothèque des modèles des archives ALIA coordonné par Muriel Chemouny [CHE 11a]. Contrairement à l'exemple représenté par la figure 3.6 (chapitre 3) et contrairement aussi à l'exemple de la figure 7.2, la séquence, ici, réunit trois sous-tâches spécialisées dans l'analyse du discours. Cette fois-ci, nous n'utilisons donc pas une séquence par tâche spécialisée en analyse du discours (voire, plusieurs séquences pour une tâche spécialisée) mais, au contraire, *une seule séquence* réunissant *toutes* les sous-tâches spécialisés. L'avantage est, bien sûr, une économie aussi bien en efforts conceptuels qu'en temps de travail à investir lors du développement et du suivi d'une bibliothèque de tels modèles d'analyse. On utilise, dans le cas présent, pour tous les textes

composant le fonds d'une archive *une et une seule séquence* pour la description de la mise en discours d'une topique.

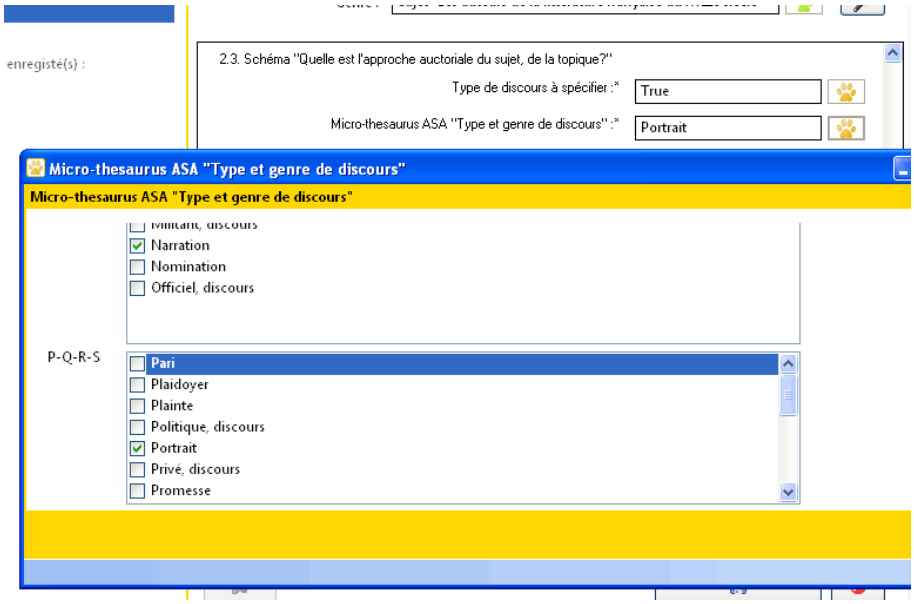

Figure 7.5. *Extrait du micro-thesaurus consacré à l'analyse du type ou genre de discours choisi par l'auteur pour développer le sujet*

La séquence réservée à l'analyse du discours développée dans les textes audiovisuels qui font partie du fonds des archives ALIA recouvre en effet trois sous-tâches spécialisées :

1. L'analyse de la *thématisation discursive* de la topique *Auteur (de la littérature française du XVIIIe siècle)*. Ici, il s'agit de préciser la ou les facettes, le ou les aspects qui sont mis en avant dans un texte audiovisuel pour parler, décrire, présenter, etc. un écrivain.

2. L'analyse du *point de vue* emprunté par l'auteur du discours pour parler d'un auteur de la littérature française du XVIIIe siècle. A l'aide de cette sous-tâche, l'analyste peut – s'il le souhaite et si cela est pertinent – préciser si les informations produites dans un texte audiovisuel sur un écrivain proviennent de l'auteur du texte, d'une autre source (citée par l'auteur du texte), de plusieurs autres sources, etc.

3. Et, enfin, l'analyse du *genre* ou du *type de discours* choisi par l'auteur du discours pour parler de « son » écrivain. A l'aide de cette sous-tâche, l'analyste peut encore préciser si les informations produites relatives à un écrivain le sont sous forme d'une narration, d'une description, d'une série d'exemples, de témoignages, etc.

La procédure de description en tant que telle pour réaliser « techniquement » l'une ou l'autre ou les trois sous-tâches, est de nouveau assez simple. La réalisation de chacune des trois sous-tâches repose sur une procédure de *description contrôlée**, c'est-à-dire sur l'usage de micro-thesaurus spécialisés qui proposent à l'analyste de choisir une ou plusieurs valeurs appropriées pour renseigner le terme conceptuel définissant une « stratégie » de mise en discours.

Ainsi, dans notre exemple (figure 7.3), l'analyste renseigne le point de vue selon lequel le sujet <Culture technique de la civilisation chavin au IVe siècle avant J.C.> est traité dans son texte. En utilisant le micro-thesaurus approprié pour expliciter cette stratégie de mise en discours, l'analyste sélectionne l'affirmation <Le sujet est traité de plusieurs points de vue> (figure 7.4). Si l'analyste souhaite affiner davantage sa description en explicitant, par exemple, chacun des différents points de vue ou encore la position de l'auteur par rapport à ces différents points de vue, il peut encore ajouter une petite notice explicative (figure 7.3).

On voit bien que nous atteignons ici une des limites actuelles de notre approche de description, à savoir celle de ne pas pouvoir produire un cadre métalinguistique permettant de structurer et d'expliciter davantage l'identité d'un point de vue, l'« orchestration » entre différents points de vue, les rapports possibles entre l'auteur du discours et les différents points de vue qu'il sollicite, etc.

Enfin, la figure 7.5 nous montre l'extrait du micro-thesaurus mis à la disposition de l'analyste afin de lui permettre d'expliciter le type ou genre de discours employé par l'auteur pour développer son sujet. Dans notre cas, l'analyste a coché le genre <Narration> et le genre descriptif <Portrait> en indiquant ainsi que le sujet <Culture technique de la civilisation chavin … > est développé sous forme d'un récit (ici : chronologique) et d'une description (d'un « portrait ») exposant les aspects les plus saillants de la culture technique de cette civilisation.

On voit bien ici une autre limite de notre approche de description de corpus audiovisuels qui consiste dans le fait de ne pas pouvoir offrir à l'analyste la possibilité de renseigner l'organisation structurale elle-même d'un récit ou d'un portrait [ADA 90, ADP 89] telle qu'elle se manifeste dans tel ou tel passage d'un texte audiovisuel. C'est une chose d'identifier un passage audiovisuel comme étant un récit ou un portrait, c'en est une autre d'expliciter leurs spécificités caractéristiques.

7.5 De l'expertise textuelle et discursive

L'approche de description de corpus audiovisuels présentée ici connaît, sans aucun doute, toute une série de limites notamment dans la prise en compte de la spécificité linguistique et sémiotique du discours qui forme un domaine de recherche très actif et innovant en France et ailleurs (voir par exemple [CHM 02, GRU 98, MAI 96, SAR 07]).

La prise en compte des résultats de ces recherches et leur intégration au moins partielle dans les ressources métalinguistiques ASA dépend de la compatibilité théorique entre ces différentes contributions (et avec l'approche présentée ici) et de l'opportunité, de l' « intérêt » de procéder dans ce sens. Concrètement, celle-ci se manifestera sous forme de la mise en place d'une bibliothèque bien plus importante de séquences et schémas de description spécialisée dans l'analyse du discours (bibliothèque de séquences à l'image de celle représentée par la figure 7.2) – bibliothèque de séquences et schémas qui tiendra compte des limites énoncées ci-dessus et des solutions éventuelles proposées dans la recherche spécialisée sur le fonctionnement du discours.

L'intérêt de l'élaboration d'une telle bibliothèque de séquences et schémas spécialisés en l'analyse du discours, nous ramène à la question des genres ou types plus spécialisés d'expertise d'informations :

– expertises qui se contentent d'une analyse du contenu purement référentiel d'un corpus de textes (audiovisuels) ;
– ou expertises qui tiennent également compte du discours et des différentes stratégies de la mise en discours d'une information (voir supra, chapitre 7.2).

Notons, en passant, qu'un des problèmes certainement les plus centraux dans l'actuelle économie des connaissances, est celle de la *traçabilité* de l'information (dans le cadre ou *traçabilité* signifie notamment *fiabilité*). Or, celle-ci est simplement impossible à mettre en place sans la prise en compte du discours de l'auteur d'une information (nouvelle, reprise, citée, retravaillée, etc.).

On le sait bien, ce n'est pas la quantité d'informations disponibles (dans une bibliothèque, dans une archive, sur le web, dans les différents réseaux sociaux, etc.) qui pose problème (bien au contraire) mais bien la *fiabilité*, la *qualité* ou encore la *valeur ajoutée* d'une information pour un public donné. L'analyse discursive ne résoudra pas à elle seule cette question essentielle à tous les métiers de « connaissances » mais sans elle, cette question ne saurait être résolue.

Troisième partie

Procédures de description

Chapitre 8

Définition du domaine de connaissances et configuration de la structure topique

8.1. Introduction

Toute tâche d'analyse commence, normalement, par l'identification et la circonscription du type d'objet de connaissance dont est question dans un texte audiovisuel. En d'autres termes, il s'agit de définir d'abord la *structure (configuration) topique** appropriée pour traiter un sujet donné dans un texte audiovisuel qui constitue l'objet de l'analyse.

La structure topique elle-même, comme nous l'avons déjà vu, tient compte de deux types d'*objets d'analyse** – les *objets référentiels* et les *objets de localisation référentielle*. En ce qui concerne plus particulièrement la partie *objets référentiels*, la structure topique peut se présenter, dans le cas le plus simple, sous forme d'un seul terme conceptuel que l'analyste doit confirmer avant de le décrire (voir ci-après, la section 8.2). Mais la structure topique peut, bien évidemment, se déployer sous forme d'une configuration positionnant plusieurs termes conceptuels les uns par rapport aux autres. C'est alors à l'analyste de définir – en sélectionnant les termes conceptuels appropriés – la structure précise dont il a besoin pour analyser un sujet particulier développé dans « son » texte ou « son » corpus de textes audiovisuels.

En définissant la structure topique appropriée pour son travail d'analyse, l'analyste *configure* (ou, plutôt, *reconfigure*) la structure topique qui est proposée dans un formulaire interactif de description. Dans la section 8.2, nous discuterons le travail de configuration/reconfiguration le plus simple qui consiste en la sélection du terme conceptuel représentant à lui seul un domaine de connaissances. Ensuite, dans les sections 8.3 et 8.4, nous discuterons des cas plus complexes de définition (de reconfiguration) d'une structure topique par l'analyste afin de l'adapter à l'objet de son analyse.

Les exemples développés dans les sections 8.3 et 8.4 ouvrent également la voie à des problématiques propres en *ingénierie sémiotique* ou *conceptuel* de modèles de description. En effet, il ne faut pas se leurrer : lorsque l'on parle de « configuration » ou encore de « réseau (sémantique, conceptuel) », etc. il ne s'agit pas seulement d'« interconnecter » des termes conceptuels pour en faire des structures. Comme nous le verrons ci-après, une *configuration de termes conceptuels* (une configuration conceptuelle) peut non seulement se *contracter* en un seul terme ou se *dilater* en un ensemble de termes conceptuels différents, mais également se former à partir de *configurations locales* qui se positionnent les unes par rapport aux autres. Or, une des relations de positionnement les plus centrales – déjà bien présente dans la théorie sémantique structurale de Greimas [GRE 67] – est celle de la *présupposition*. Partant de ce constat, toute une série de relations plus complexes peuvent être sollicitées pour élaborer des configurations topiques de plus en plus sophistiquées rendant compte d'une manière fine des spécificités internes d'un sujet donné dans un texte ou un corpus de textes audiovisuels. C'est ici où la grammaire narrative de Greimas (certainement une des plus abouties dans son genre) entre en jeu en « fournissant » au cogniticien ou « concept designer » toute une variété de relations dites narratives explicites pour l'élaboration de modèles topiques plus sophistiqués [GRE 79, STO 83, STO 85, STO 87]. Même si – pour des raisons surtout pratiques – nous n'avons pas intégré systématiquement toute la richesse de la *sémiotique** narrative greimasienne dans les modèles actuellement utilisés pour l'analyse des corpus audiovisuels, insistons néanmoins sur le fait que c'est – comme déjà expliqué dans le premier chapitre – cette approche, cette « vision » qui oriente et en quelque sorte « nourrit » notre démarche de modélisation, c'est-à-dire d'élaboration de modèles de description de contenus (audiovisuels).

8.2. Quelques rappels et précisions

Rappelons-nous de la distinction que nous avons introduite, dans le chapitre 5.3, entre le *terme conceptuel** (TC) et l'objet ou le domaine de connaissances. Un *terme conceptuel* est l'expression lexicale d'un *concept* et fait partie du *vocabulaire* du métalangage ASA qui nous sert à l'élaboration des modèles pour la description de corpus audiovisuels. En revanche, un *objet* ou *domaine de connaissances** est un objet qui est thématisé (d'une manière ou d'une autre et selon les objectifs propres à une communication) dans un texte ou corpus de textes audiovisuels et qui est décrit

par l'analyste à l'aide d'un modèle faisant partie du *métalangage de description** ASA.

La *structure topique* est une *configuration* de *termes conceptuels (TC)* qui font partie du méta-lexique ASA (voir chapitres 12 et 13). Elle définit un type d'objets ou de domaines de connaissances de l'*univers du discours** d'une archive qui est thématisé (traité, développé) dans un texte audiovisuel particulier faisant partie du fonds de l'archive en question.

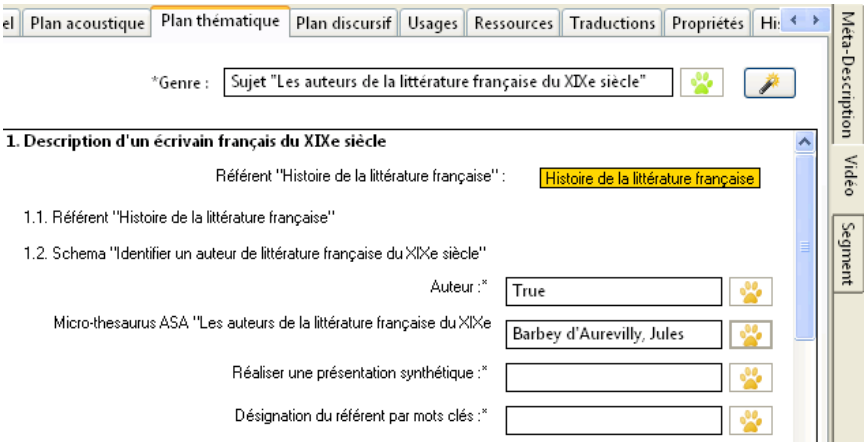

Figure 8.1. *L'opération de sélection d'un terme conceptuel*

Avant toute analyse concrète du contenu d'un texte audiovisuel, il faut donc définir la structure topique à l'aide de laquelle on veut appréhender la thématisation d'un objet de connaissance. Deux cas de figure peuvent se présenter :

1. soit la structure topique appropriée pour analyser le contenu d'un texte audiovisuel existe déjà et fait partie d'une bibliothèque de structures topiques définissant les objets et domaines de connaissance de l'univers de discours d'une archive ;
2. soit la structure topique n'existe pas encore ou, en tout cas, ne fait pas partie de la bibliothèque des structures topiques de l'archive pour laquelle une analyse d'un texte ou d'un corpus de textes audiovisuels doit être réalisée.

Le deuxième cas de figure requiert l'intervention d'un *modélisateur** (d'un *concept designer*), c'est-à-dire d'une personne ou d'un groupe de personnes spécialisé(e) en la définition et l'élaboration de modèles de description nécessaires pour procéder à l'analyse et l'indexation de ressources numériques. Quant au premier cas de figure, il faut de nouveau distinguer entre deux possibilités :

1. La première possibilité est celle où la structure topique doit être utilisée telle quelle par l'analyste. Le rôle de l'analyste se « réduit » à son métier principal qui est celui de localiser, de décrire, d'expliciter, d'interpréter ou encore de classer et d'indexer les informations pertinentes dans le ou les textes qui constituent son objet de travail ;

2. La deuxième possibilité est celle où l'analyste peut *adapter* une structure topique existante à ses besoins propres. Autrement dit, pour utiliser la structure topique qui forme une *séquence** de travail dans un formulaire de description du contenu d'un texte audiovisuel, l'analyste peut être invité à sélectionner parmi les termes conceptuels constituant la structure topique ceux qui sont effectivement pertinents pour son analyse. L'analyste intègre ici une partie des compétences du modélisateur (du concept designer) dans la mesure où il reconfigure une structure topique déjà existante afin de pouvoir utiliser une variante appropriée à son objet, c'est-à-dire au texte audiovisuel étudié. Le cas extrême est ici, bien sûr, celui où la structure topique d'un formulaire de description se confond avec le vocabulaire entier des termes conceptuels composant l'ontologie des *objets d'analyse** de l'*univers du discours** ASA.

Dans ce chapitre, nous allons étudier rapidement les deux possibilités citées, à savoir celle de l'*utilisation directe* (sans modification) d'une structure topique existante pour l'analyse d'un contenu audiovisuel et celle – *indirecte* – exigeant de la part de l'analyste un travail de reconfiguration plus ou moins importante d'une structure topique existante afin d'en faire un outil approprié pour ses besoins d'analyse.

Une structure topique ressemble à un graphe sémantique ou conceptuel [SOW 84, STO 85, STO 87, STO 92] ou encore à un « script » [SCA 77, STO 94]. Une différence notable ici est celle que les termes (les « concepts ») composant une structure topique sont définis dans un *vocabulaire de termes conceptuels* (dans une *ontologie descriptive*) tandis que les réseaux conceptuels ou sémantiques ne disposent guère de plus que de *quelques primitifs conceptuels* (appelés parfois aussi par abus de langage, *cognitifs*) rendant difficile leur emploi pour des analyses concrètes et empiriquement riches de ressources textuelles.

Le cas d'une structure topique la plus simple est celui où il se réduit à *un seul terme conceptuel* représentant en elle seule l'objet de connaissance thématisé dans un texte ou corpus de textes audiovisuels. La figure 8.1 nous montre un exemple concret provenant de la *bibliothèque des modèles de description** que l'analyste utilise pour décrire des corpus audiovisuels composant les archives ALIA. Dans ce cas concret, il s'agit d'un texte audiovisuel (d'un entretien) qui traite de Jules Barbey d'Aurevilly. Comme le montre la figure 8.1, pour identifier et décrire le contenu du texte en question (c'est-à-dire le fait qu'il traite de <Jules Barbey d'Aurevilly>), l'analyste doit d'abord sélectionner le terme conceptuel approprié,

c'est-à-dire dans notre cas, le terme [Auteur] qui fait partie de la structure topique définissant l'objet ou le domaine de connaissances « auteur de la littérature française du XIXe siècle ».

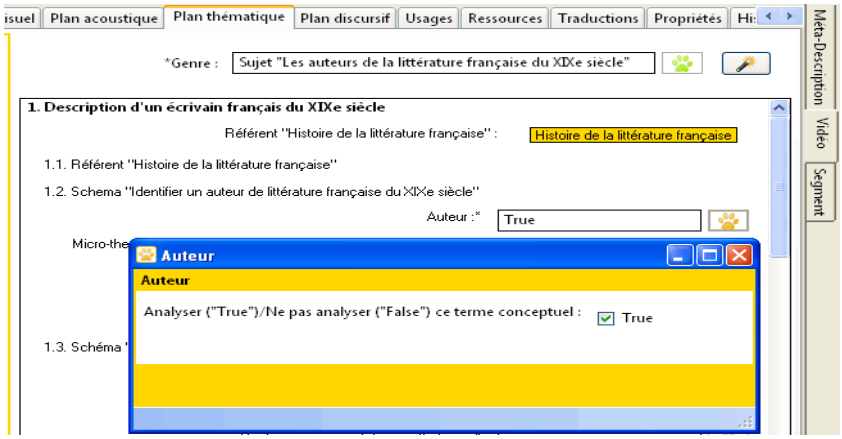

Figure 8.2. *Interface permettant la sélection d'un terme conceptuel*

Dans notre cas, la tâche de définition ou d'ajustement de la structure topique que l'analyste doit réaliser, se réduit à un geste technique simple :

1. l'analyste active le bouton qui se trouve à droite de l'expression métalinguistique [Auteur] (voir figure 8.1) ;
2. s'ouvre ensuite une boîte de dialogue qui invite l'analyste (voir figure 8.2) à préciser s'il envisage, oui ou non, d'analyser un objet de connaissance à l'aide du terme conceptuel en question ;
3. l'analyste choisit ensuite l'option *True* – choix qui exprime une affirmation du genre *Oui, il est vrai que je souhaite analyser un objet de connaissance qui est représenté par le terme conceptuel [Auteur]*.

Lorsque l'analyste sélectionne et ouvre un formulaire de description thématique, c'est-à-dire du contenu audiovisuel, la valeur par défaut est *False* (*Faux*, en français). *A priori*, on considère qu'aucune analyse n'est envisagée à l'aide d'un terme conceptuel. Pour analyser un objet de connaissance à l'aide d'un terme conceptuel ou d'une configuration (ajustée) de plusieurs termes conceptuels, l'analyste doit donc le signaler expressément.

Pour revenir à l'exemple représenté dans la figure 8.1, répétons qu'il s'agit ici de l'exemple le plus simple d'une utilisation directe de la structure topique : l'analyste ne doit que confirmer son intention d'analyser un objet de connaissance à l'aide d'un et d'un seul terme conceptuel. Dans notre exemple concret, il s'agit de l'auteur français du XIXe siècle Jules Barbey d'Aurevilly (1808 – 1889). Le nom de cet

auteur fait partie d'un *micro-thesaurus** ASA spécialisé sur les noms d'auteurs dont se sert l'analyste des textes audiovisuels faisant partie du fonds des archives ALIA.

On voit bien que l'analyse concrète à l'aide d'un formulaire tel que celui représenté par la figure 8.1 est extrêmement simple et rapide à réaliser. Une fois le passage pertinent identifié dans le texte audiovisuel (c'est-à-dire le passage où on parle « effectivement » de Jules Barbey d'Aurevilly), l'analyste doit seulement :

– sélectionner le formulaire approprié dans la bibliothèque des formulaires ALIA ;
– confirmer ensuite son choix de traiter un sujet représenté par le terme conceptuel [Auteur] et ;
– enfin, sélectionner le nom approprié dans la liste des noms d'auteurs du micro-thesaurus mis à sa disposition.

On ne peut guère s'imaginer d'analyses plus courtes et plus simples que celle-ci qui se basent, par ailleurs, sur un modèle qui peut être *réutilisé tel quel* pour une très grande variété de cas similaires. Cependant, atteindre un tel niveau de rapidité et de simplicité (pour l'analyste), présuppose un travail énorme et très complexe de définition, conception et développement de modèles (et donc de ressources métalinguistiques) de description.

Ceci étant, comme déjà dit, la réduction de la structure topique définissant un type d'objet de connaissance à un seul terme conceptuel, constitue vraiment un cas limite. Dans les sections suivantes de ce chapitre, nous allons voir des cas plus complexes.

8.3. (Re-)configurer et adapter une structure topique existante

Voyons maintenant la deuxième possibilité évoquée ci-dessus, à savoir celle qui exige de la part de l'analyste un travail d'adaptation d'une structure topique existante afin d'en faire un outil approprié pour son travail d'analyse. Ce travail d'adaptation, nous l'appelons *(re-)configuration d'une structure topique*.

Dans le chapitre 5.2, nous avons déjà eu l'occasion de développer le fait que l'analyste doit d'abord *configurer* (c'est-à-dire définir) sa structure topique et ensuite seulement la décrire. Si l'acte de la « configuration » se réduit au geste simple, décrit ci-dessus, de sélectionner un et un seul terme, il est évident et ne pose aucun problème.

Mais, comme déjà dit également à plusieurs reprises, les séquences organisant la description référentielle dans un formulaire de description peuvent réunir des pans entiers du méta-lexique (du vocabulaire conceptuel) ASA tout en laissant à

l'analyste la liberté mais aussi la responsabilité d'en sélectionner les termes conceptuels qui conviennent au mieux à son analyse.

Les figures 8.3 et 8.4 montre un exemple de cette éventualité. Il s'agit d'un formulaire de description servant à l'analyste pour décrire des textes audiovisuels fournissant des informations (visuelles ou sous forme de discours tenus) sur des objets à statut patrimonial tels que documents sonores ou filmiques et/ou sur des institutions (musées, archives, etc.) dans lesquelles on peut trouver ce genre d'objets.

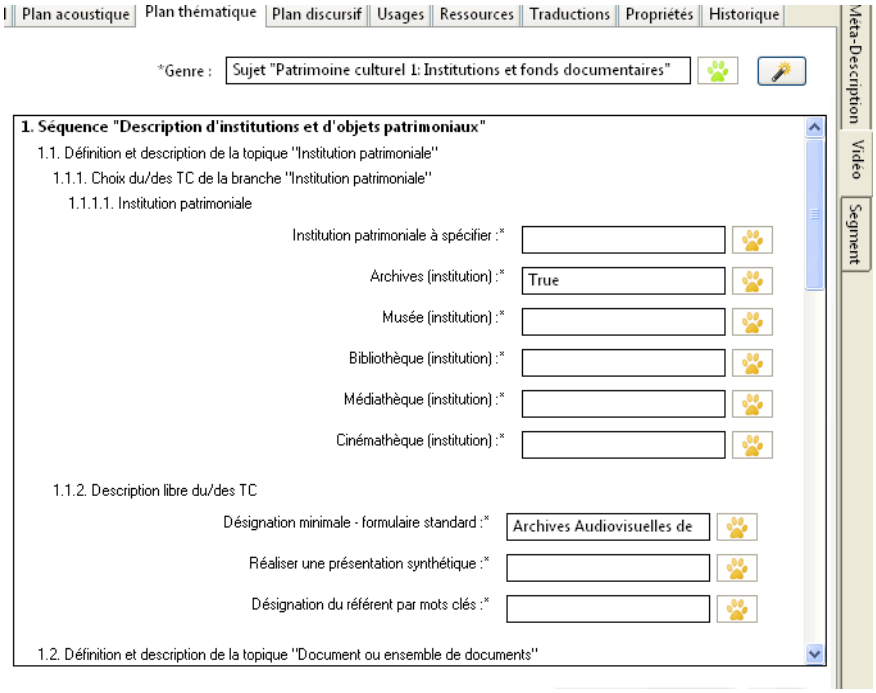

Figure 8.3. *Configuration et description de la structure topique – 1ère partie*

D'abord, l'analyste identifiera les « contours » du sujet à analyser : dans le texte audiovisuel soumis à l'analyse, le sujet est-il consacré uniquement aux objets à statut patrimonial (par exemple à telle ou telle documentation filmique, tel ou tel enregistrement sonore …) ou uniquement aux institutions patrimoniales (tel ou tel musée, telle ou telle archive, etc.) ou, traite-t-il, au contraire, des deux aspects à la fois (par exemple, telle ou telle collection d'enregistrements filmiques dans tel musée, tel ou tel dépôt d'enregistrements sonores dans telle archive, etc.).

Ensuite, l'analyste examinera la structure topique disponible pour analyser ce genre de contenu audiovisuel et sélectionner les termes conceptuels à solliciter pour configurer la structure topique de son sujet. Enfin, il décrira le sujet dans son texte audiovisuel à l'aide de la structure topique (re-)configurée.

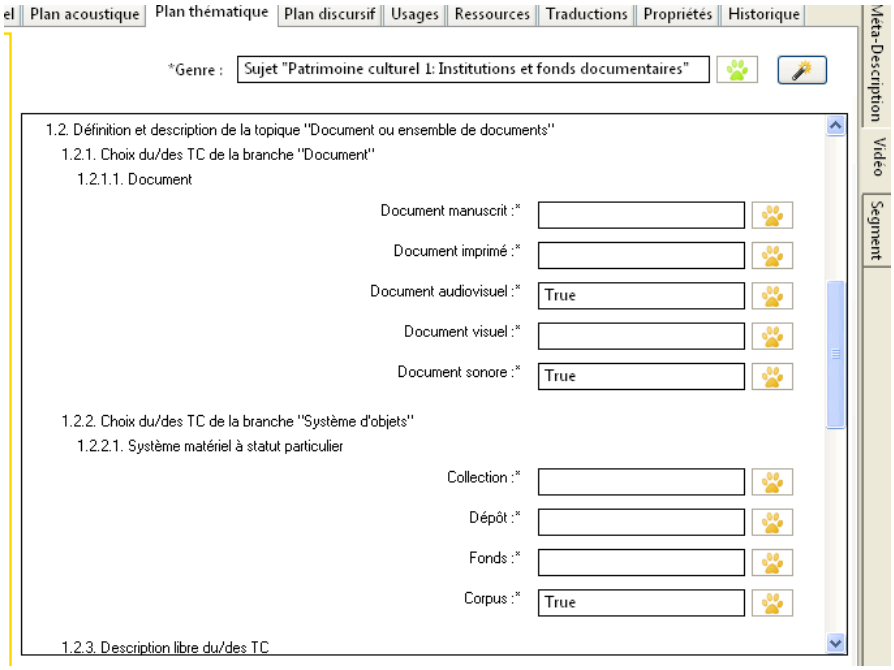

Figure 8.4. *Configuration de la structure topique – 2ème partie*

Si nous examinons les trois figures 8.3, 8.4 et 8.5 *ensemble*, la *description* du sujet du texte audiovisuel formant l'objet de l'analyse, stipule que ce dernier traite d'un corpus de documents filmiques consacrés à l'immigration allemande au Chili du XIXe siècle *qui fait partie des Archives Audiovisuelles de la Recherche (AAR) à Paris*. Pour arriver à ce résultat, l'analyste a procédé comme suit :

– sélection (figure 8.3) du terme conceptuel [Archives Audiovisuelles] dans une liste de termes conceptuels représentant de domaine de connaissances *Institutions patrimoniales* ;
– description libre (figure 8.3) du terme conceptuel [Archives] ;
– sélection (figure 8.4) des termes conceptuels [Document audiovisuel] et [Document sonore] dans une liste de termes conceptuels représentant le domaine des connaissances *Document au sens d'un objet à statut patrimonial ;*

– sélection (figure 8.4) du terme conceptuel [Corpus] dans une liste de termes conceptuels représentant le domaine des connaissances *Ensemble d'objets ;*
– description libre (figure 8.5) tenant compte de la *configuration [Corpus] de {[Document audiovisuel], [Document sonore]}.*

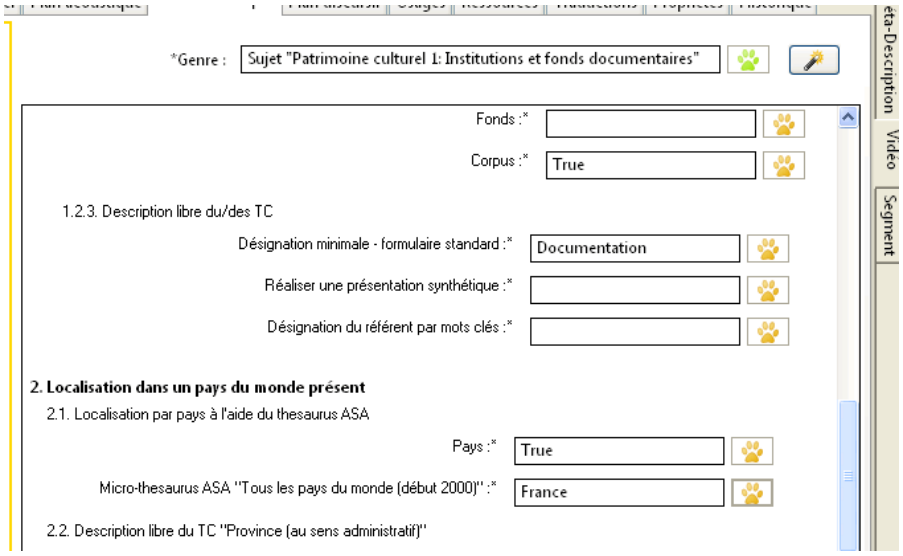

Figure 8.5. *Description (libre) de la deuxième partie de la structure topique (re-)configurée*

8.4. (Re-)configurer des structures topiques plus complexes

Revenons de nouveau à la figure 3.2 du chapitre 3. Elle nous montre une configuration très simple définissant le domaine de connaissance *Formation culturelle d'une civilisation*. Comme nous l'avons déjà fait remarquer, si par exemple, l'un ou l'autre, ou les deux termes conceptuels sont nécessaires pour la définition de l'objet en question, rien n'est dit si l'un des deux termes conceptuels dépend « crucialement » de l'autre (dans ce sens précis où le terme conceptuel dépendant ne peut être sélectionné que si le terme conceptuel dont il dépend est déjà asserté, sélectionné), rien n'est dit non plus si le choix de tel ou tel terme conceptuel exclut tel ou tel autre terme conceptuel, et ainsi de suite.

Ceci étant, l'approche développée ici permet de tenir compte, lors de la spécification et du développement de *modèles de description** du contenu audiovisuel, des contraintes spécifiques qui déterminent la structure d'une

configuration générique définissant un domaine de connaissances (ou une partie de celui-ci).

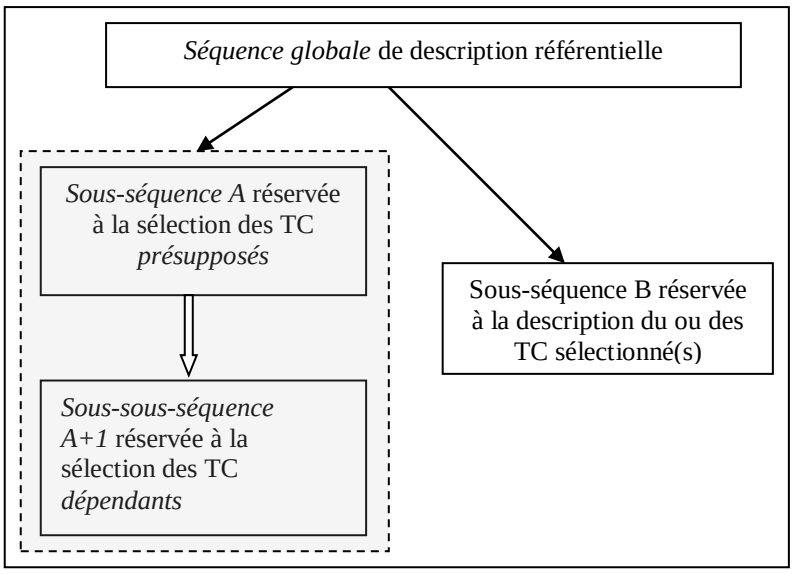

Figure 8.6. *Représentation de la structure fonctionnelle d'une interface intégrant une configuration complexe entre termes conceptuels présupposés et termes conceptuels dépendants*

En effet, à l'aide de la distinction que nous faisons du point de vue de l'*architecture fonctionnelle* d'un formulaire de travail entre *séquence globale* et *sous-séquences spécialisées*, il est tout à fait possible de réaliser des formulaires de description référentielle qui intègrent les contraintes telles que celles que nous venons d'énumérer ci-dessus.

Ainsi, la figure 8.6, nous montre l'exemple d'une architecture fonctionnelle de formulaires interactifs de travail qui intègre le fait que certains termes conceptuels (dit *termes conceptuels dépendants*) ne peuvent être sélectionnés par l'analyste que si certains autres termes conceptuels (dit *termes conceptuels présupposés*) ont déjà été sélectionnés.

Les figures 8.7 et 8.8 montrent l'extrait de l'interface d'un formulaire d'analyse du contenu audiovisuel portant sur le domaine de connaissances intitulé *Formation culturelle, époque et société*. Il permet l'analyse de tout texte audiovisuel qui traite de n'importe quelle formation culturelle (culture matérielle, culture immatérielle, culture politique, culture sociale, culture linguistique, etc.) :

– soit « en tant que telle » (par exemple en tant qu'objet *décontextualisé* de recherche ou de réflexion) ;
– soit en relation à une époque historique ou une civilisation donnée ;
– soit encore en relation à un acteur social collectif : groupe ou mouvement social, communauté, institution, etc.

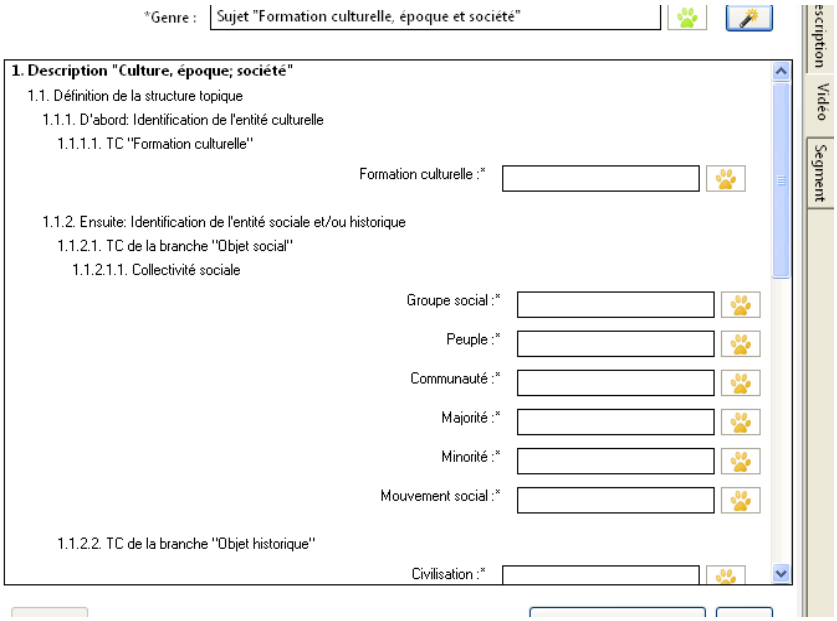

Figure 8.7. *Premier extrait de l'interface du formulaire « Formation culturelle, époque et société »*

Ce formulaire représente un *modèle de description** qui articule une configuration définitionnelle du domaine de connaissance qui est doublement remarquable :

– *premièrement*, ladite configuration définitionnelle impose une *contrainte de présupposition* entre deux groupes de termes conceptuels qui la composent : certains termes conceptuels (tels que ceux faisant partie de la branche [Objet social] dans la figure 8.7) présupposent d'autres termes conceptuels (dans notre exemple, le terme conceptuel *présupposé* est [Formation culturelle]) ;

– *deuxièmement*, la même configuration n'impose pas une et une seule structure topique (comme c'est le cas pour la configuration représentée par la figure 3.2 dans le chapitre 3). Le *modèle de description** sous-tendant le formulaire d'analyse du contenu qui est représenté par les deux figures 8.7 et 8.8 circonscrit ici plutôt une

certaine multiplicité de structures topiques potentielles parmi lesquelles l'analyste spécifie celle qu'il considère comme la plus appropriée pour décrire le sujet traité par le texte audiovisuel, objet de son analyse. Autrement dit, l'analyste configure (reconfigure) la structure topique qui lui est proposée par le formulaire de travail afin d'en faire un outil de travail approprié à son objet d'analyse.

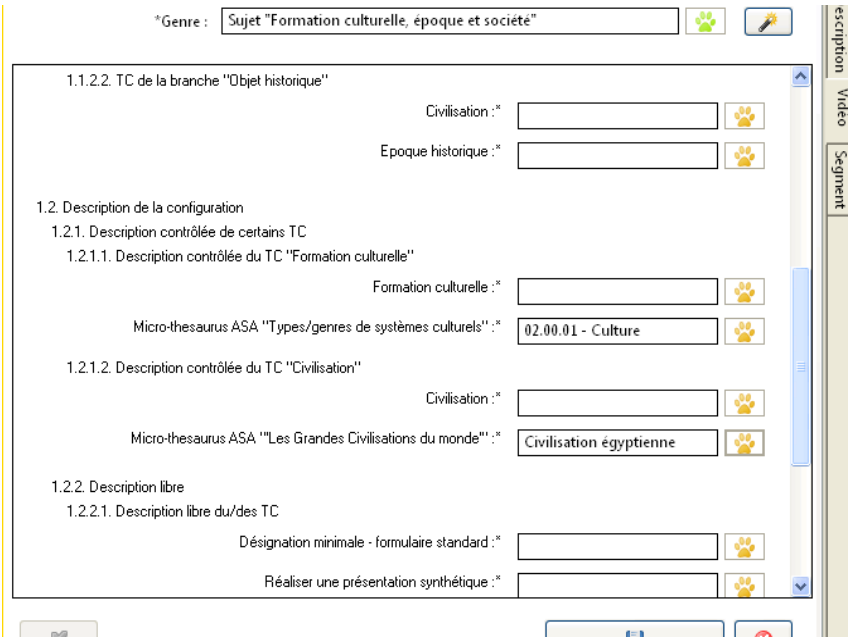

Figure 8.8. *Deuxième extrait de l'interface du formulaire « Formation culturelle, époque et société »*

On comprend aisément que cette deuxième propriété d'une configuration de termes conceptuels – la propriété de contenir une multiplicité de structures topiques – est d'une très grande importance pour l'activité de l'ingénierie cognitive ou sémiotique de modèles de description. En effet, la propriété en question montre qu'il est en soi tout à fait possible de *contenir le nombre* de formulaires de description (même pour des fonds thématiquement très variés) *à condition qu'on réussisse à spécifier des configurations définitionnelles permettant d'engendrer des structures topiques multiples et variées*.

Pour reprendre notre exemple, le formulaire représenté par les deux figures 8.7 et 8.8 permet de procéder à une description thématique simple et rapide de tous les textes (audiovisuels ou autres) qui traite un sujet relevant de ce vaste domaine

empirique, intéressant la plupart des disciplines en sciences humaines et sociales, qui est celui de la *culture en relation avec le monde social et historique*, etc. Mais, c'est de nouveau à *l'analyste* de *configurer la structure topique* dont il a besoin. La (re-)configuration de la structure topique s'effectue en deux étapes :

1) Dans une première étape, l'analyste sélectionne le terme conceptuel [Formation culturelle] (= partie de la première sous-séquence ayant comme intitulé *D'abord : Identification de l'entité culturelle*). Il signifie par là qu'il souhaite le décrire davantage (dans la deuxième sous-séquence). S'il souhaite décrire le terme conceptuel [Formation culturelle] d'une manière *décontextualisée* (c'est-à-dire en soi, pas en rapport avec une époque historique, une civilisation ou un groupe social particulier, etc.), il ne sélectionne plus aucun autre terme conceptuel. La structure topique dont il a besoin se réduit ainsi en un seul terme conceptuel. Mais celui-ci reste néanmoins *virtuellement* en relation avec tous les autres termes conceptuels composant la configuration générique qui définit le domaine de connaissances du modèle de description sous-tendant notre formulaire de travail.

2) Dans une deuxième étape, l'analyste sélectionne le ou les termes dépendants du terme présupposé qui est le terme [Formation culturelle] (= partie de la première sous-séquence ayant comme intitulé *Ensuite : Identification de l'entité sociale et/ou historique*). L'analyste peut décider de définir une structure topique à *dimension plutôt sociale* (en sélectionnant un, voire plusieurs termes de la branche [Objet social]) ou plutôt *historique* (en sélectionnant l'un ou l'autre terme conceptuel composant la branche [Objet historique]) ou encore une structure topique de nature *sociohistorique* (en sélectionnant, par exemple, [Epoque historique] et [Groupe social]). Tout dépend, bien sûr du contenu du texte audiovisuel qu'il doit décrire.

Une fois que l'analyste a configuré (c'est-à-dire défini) la structure topique appropriée pour décrire le contenu de son texte audiovisuel, il passe ensuite à la *description elle-même* de *« sa »* structure topique (= deuxième sous-séquence intitulée *Description de la configuration*).

Comme le montre la figure 8.8, il peut opter pour une procédure de *description dite libre* et/ou une *description dite contrôlée* (c'est-à-dire à l'aide d'un thesaurus). La description dite libre, en tout cas, est supposée renseigner *la structure topique telle qu'elle* a été définie d'abord (c'est-à-dire dans la sous-séquence *Définition de la structure topique*) par l'analyste :

– si la structure topique est réduite à un seul terme conceptuel sélectionné par l'analyste (dans notre cas, si elle est réduite au seul terme conceptuel [Formation culturelle]), la description ne renseigne que celui-ci ;

– si la structure topique est composée de deux ou plusieurs termes conceptuels, alors la description libre renseigne obligatoirement l'ensemble des termes sélectionnés et non pas tel ou tel terme pris isolément.

Par exemple, si la structure topique définie par l'analyste est composée des termes conceptuels {[Formation culturelle] et [Epoque historique]}, alors la description doit tenir compte de cette structure particulière. Ainsi, étant donné cette topique à dimension historique, une description libre du genre « culture politique du XVIIIe siècle » est appropriée ; en revanche une description libre du genre « culture politique andine » ne l'est pas.

Dans le cadre de nos recherches sur la description de corpus audiovisuels faisant partie des différents ateliers d'expérimentation du projet ASA-SHS[46], nous avons défini, développé et utilisé toute une série de modèles de description référentielle dont la particularité consiste justement dans le fait de laisser à l'analyste une certaine liberté de *configurer sa* structure topique au lieu de lui imposer une structure topique « tout faite ».

Cependant, par rapport au modèle qui définit le formulaire de travail représenté par les deux figures 8.7 et 8.8, ces modèles sont souvent plus simples et ne font qu'exploiter la nature taxinomique du méta-lexique (du vocabulaire) des termes conceptuels ASA (voir chapitres 12 et 13). Par exemple, la figure 8.7 nous montre la possibilité offerte à l'analyste de choisir dans la branche [Collectivité sociale] le ou les termes conceptuels les plus appropriés pour décrire le sujet spécifique dans un texte audiovisuel qui se réfère à tel groupe, tel acteur, tel mouvement, etc. Or, l'élaboration d'un formulaire de description du contenu audiovisuel qui n'intègre que la branche [Collectivité sociale] dans sa partie réservée à la description référentielle, est déjà d'une très grande utilité pour décrire et indexer des corpus audiovisuels (ou autres) composant une archive ou une bibliothèque de ressources sociologiques.

---

46 Rappelons-nous qu'il s'agit notamment des archives « A la Rencontre des Cultures » (http://semiolive.ext.msh-paris.fr/alia/), « Atelier Littéraire d'Ici et d'Ailleurs » (http://semiolive.ext.msh-paris.fr/alia/) et « Atelier des Arkéonautes » (http://semiolive.ext.msh-paris.fr/ada/) auxquels se sont ajoutés, au cours du projet plusieurs autres archives.

Chapitre 9

La procédure de la description libre d'un corpus audiovisuel

9.1. Introduction

L'analyse, la description d'un texte audiovisuel est, techniquement parlant, le processus de rendre conforme, selon le point de vue de l'analyste, un terme conceptuel et/ou une configuration de termes conceptuels aux spécificités du texte audiovisuel soumis à l'analyse.

Cette conformation consiste concrètement en des actions spécifiques telles que l'identification, dans un texte audiovisuel, d'un *objet filmique* lato sensu (incluant aussi bien des situations filmées, des activités filmées, des personnes et des groupes, des objets à proprement parler, etc.), sa catégorisation conceptuelle, sa désignation ou dénomination, sa présentation, son interprétation, sa mise en rapport avec d'autres objets filmiques, et ainsi de suite.

Dans ce chapitre, nous allons présenter l'une des deux principales procédures rendant possible cette conformation. Il s'agit de la procédure que nous appelons *description libre*. Comme nous le verrons, la description libre réunit toutes les activités qui doivent permettre à l'analyste d'identifier, d'expliciter, voire

d'expliquer ou encore d'indexer concrètement un objet de connaissance thématisé et mis en discours dans un texte audiovisuel concret.

La description libre est en effet l'une des deux procédures que nous rencontrons aussi bien dans la tâche de la *description référentielle** (c'est-à-dire dans la tâche de la description du plan topique d'un texte audiovisuel ; voir les chapitres 5 et 6) que dans la description de sa *mise en discours** (voir le chapitre 7) ou encore de sa *mise en scène audiovisuelle*. L'autre procédure principale de description que nous avons identifiée et développée est celle de la *description contrôlée** que nous allons présenter dans le chapitre 10.

Dans la section 9.2 nous allons présenter d'une manière globale la procédure de la description dite libre. Ensuite, dans les sections 9.3 à 9.7, nous discuterons plus en détail certaines des activités les plus récurrentes composant ladite procédure.

9.2. Anatomie de la procédure appelée « description libre »

La procédure de la description libre renvoie directement à la problématique de la production plus ou moins libre de métadonnées par l'analyste amateur ou professionnel (soit en tant qu'*utilisateur* individuel soit en tant que *communauté d'utilisateurs*) d'objets textuels (et, plus particulièrement, audiovisuels). Cette production plus ou moins libre de métadonnées par des utilisateurs ou groupes d'utilisateurs est néanmoins « contrôlée » dans la mesure où le processus concret de la production – l'*indexation* – s'exécute en référence à un modèle de description du contenu et suit une trame, un scénario imposé (pour plus de détails au sujet de la distinction entre modèle de description et scénario d'activités, voir [STO 11a].

Nous pensons, autrement dit, que l'explicitation des activités de la procédure de la description libre est une condition essentielle pour rendre plus sophistiquée ce processus essentiel qui est l'*analyse d'un texte ou d'un corpus audiovisuel* dans une perspective des médias sociaux numériques [DES 11a, DES 11b, DES 11c] et aussi pour mieux l'adapter aux attentes et aux intérêts des individus et groupes d'individus qui s'en servent nécessairement afin de pouvoir diffuser, partager, conserver et transmettre leurs contenus audiovisuels.

La procédure de la description libre ressemble en effet beaucoup à celle appelée traditionnellement, *indexation libre*. Cependant, contrairement à l'indexation libre au sens courant du terme, la description libre est comprise ici comme un faire qui est entièrement guidé par des *modèles* de description représentant un domaine d'expertise tel que ceux des ateliers d'expérimentation du projet ASA-SHS. *Uniquement les valeurs*, c'est-à-dire les *index* sous forme, par exemple, d'expressions nominales simples ou composées, sont librement générés par l'analyste (voir ci-après, la section 9.3).

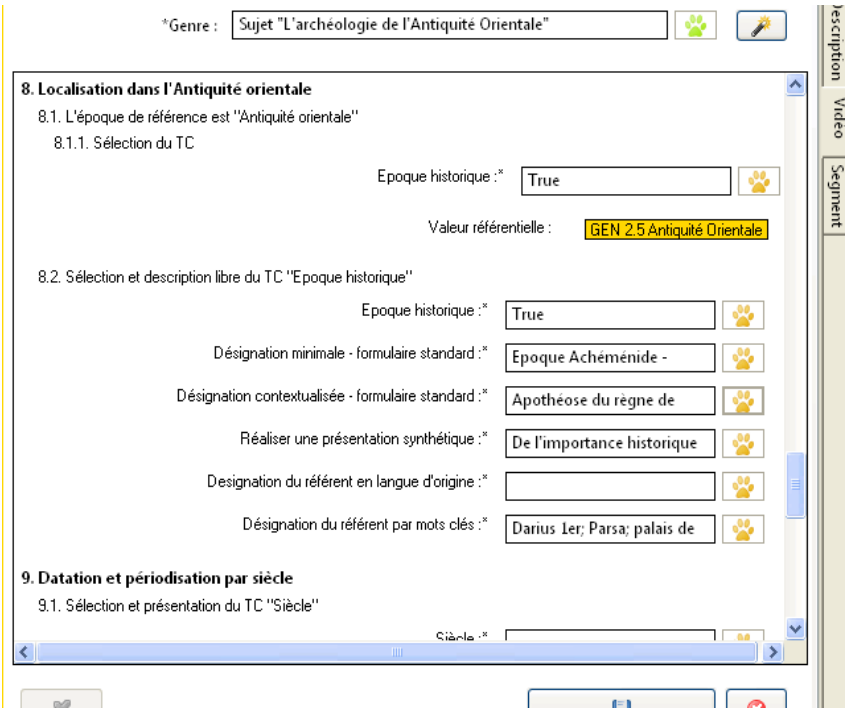

Figure 9.1. *La procédure de la description libre d'un terme conceptuel*

La figure 9.1 montre l'extrait d'un formulaire de description du contenu de corpus audiovisuels dont le sujet traite de l'archéologie de l'Antiquité Orientale. Il s'agit de la huitième séquence du modèle d'analyse consacrée à l'identification et la description de l'époque historique dans l'Antiquité Orientale concernée. Cette huitième séquence constitue, ensemble avec la neuvième séquence intitulée *Datation et périodisation par siècle*, la tâche de la *contextualisation référentielle* (ici : de la contextualisation chronologique ou historique) d'un objet de connaissance correspondant à la topique de la recherche archéologique consacrée à l'Antiquité Orientale (la topique de la recherche archéologique elle-même recouvre des objets de connaissance tels que *approche archéologique*, *domaine de la recherche*, *thème de la recherche* ainsi que *activités*, *lieux* et *objets de la recherche*).

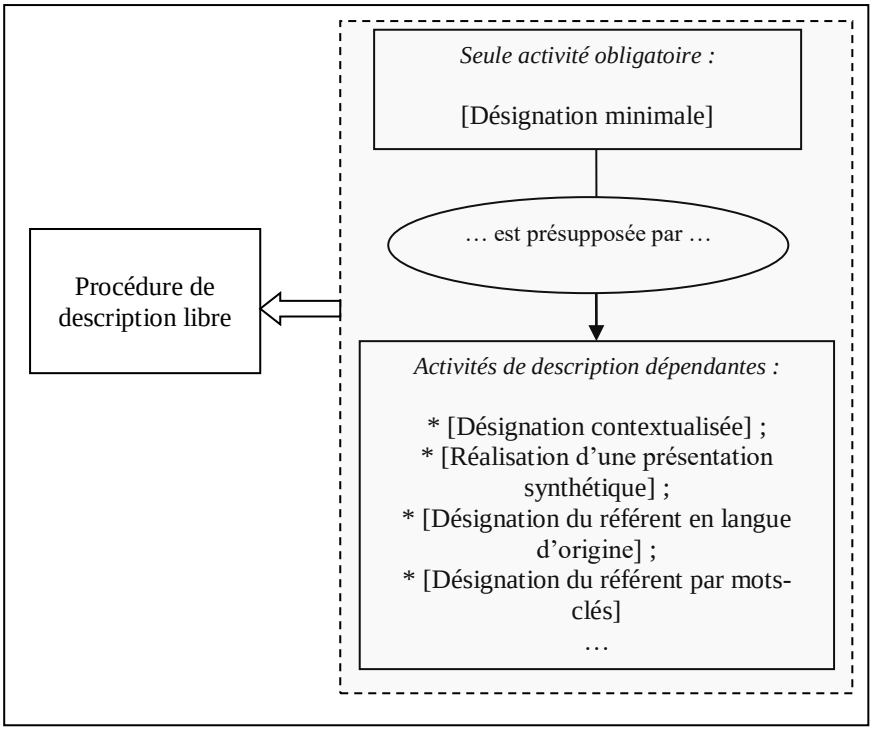

Figure 9.2. *Organigramme de la procédure de la description libre*

C'est plus particulièrement la *deuxième partie* de la huitième séquence (figure 9.1), c'est-à-dire la partie intitulée *Sélection et description libre du T(erme C(onceptuel) [Epoque historique]* qui nous intéresse ici. C'est cette partie qui permet à l'analyste de saisir *librement* le nom de l'époque et d'en fournir – si pertinent – des renseignements supplémentaires. En considérant de plus près cette troisième partie, nous remarquons qu'elle est composée :

1) du terme conceptuel [Epoque] qui représente à lui seul le domaine de connaissances historiques dont se sert l'auteur du texte audiovisuel pour situer dans le temps une activité de recherche en archéologie. Ce terme conceptuel fait partie du vocabulaire du méta-lexique ASA réservé à la représentation des objets d'analyse (c'est-à-dire des domaines de connaissance) de l'univers du discours ASA (pour plus de détails, voir chapitre 3) ;

2) de cinq termes conceptuels [Désignation minimale], [Désignation contextualisée], [Réalisation d'une présentation synthétique], [Désignation du référent en langue d'origine] et [Désignation du référent par mots-clés] qui représentent les activités de la procédure de la description libre. Ces termes

conceptuels font partie du vocabulaire du méta-lexique ASA réservé aux différentes procédures et activités d'analyse de l'objet « texte » (pour plus d'explications, voir chapitre 14)

D'une manière générale (voir figure 9.2), toute procédure de description libre est composée d'au moins une *activité spécifique* mais elle peut être aussi composée de deux ou plusieurs activités (la définition de la procédure dépend de l'objectif ou de la « politique » qui détermine le travail concret de l'analyse). Si la procédure de description libre est composée d'une seule activité, alors celle-ci est obligatoirement l'activité [Désignation minimale (de l'objet de connaissance)] (voir figure 9.2).

Si la procédure de description libre est composée de deux ou de plusieurs activités, alors celles-ci dépendent de l'activité [Désignation minimale] (voir figure 9.2). Ainsi, dans notre exemple concret (figure 9.1), les activités [Désignation contextualisée], [Réalisation d'une présentation synthétique], [Désignation du référent en langue d'origine] et [Désignation du référent par mots-clés] présupposent, toutes, l'activité [Désignation minimale] : l'analyste doit d'abord renseigner le formulaire associé à l'activité [Désignation minimale] et seulement ensuite les formulaires associés aux autres activités. Cependant, l'analyste est libre de réaliser l'une ou l'autre ou plusieurs des activités dépendantes (de l'activité [Désignation minimale].

Soulignons, en passant, que la procédure de description libre représentée par la figure 9.1 ne forme qu'un exemple concret d'une telle procédure. Autrement dit, une procédure de description libre peut recouvrir des activités spécifiques différentes de celles que nous voyons dans la figure 9.1 – tout dépend du contexte et de l'objectif de l'analyse à laquelle doit se conformer ladite procédure. La seule activité qui doit être obligatoirement présente est celle de la [Désignation minimale].

Voyons maintenant plus en détail les cinq activités qui composent la tâche de la description libre telle qu'elle est présentée dans la figure 9.1.

9.3. L'activité de description [Désignation minimale]

Parmi ces cinq activités qui composent la procédure de description libre donc *seule l'activité [Désignation minimale] est obligatoire*. Toutes les autres sont facultatives mais leur « bon usage » ajoute évidemment à la richesse d'une description concrète.

L'activité [Désignation minimale] veut dire : *fournir les* – du point de vue de l'analyste – informations *minimales nécessaires* pour identifier le contenu d'un objet audiovisuel qui fait référence (pour reprendre notre exemple représenté par la figure 9.1) à une époque historique spécifique de l'Antiquité Orientale. Pour réaliser cette activité, l'analyste accède, en cliquant sur l'icône représentant une patte de chat

(figure 9.1), à un formulaire *standard* (figure 9.3) qui encadre son activité de saisie d'une valeur appropriée.

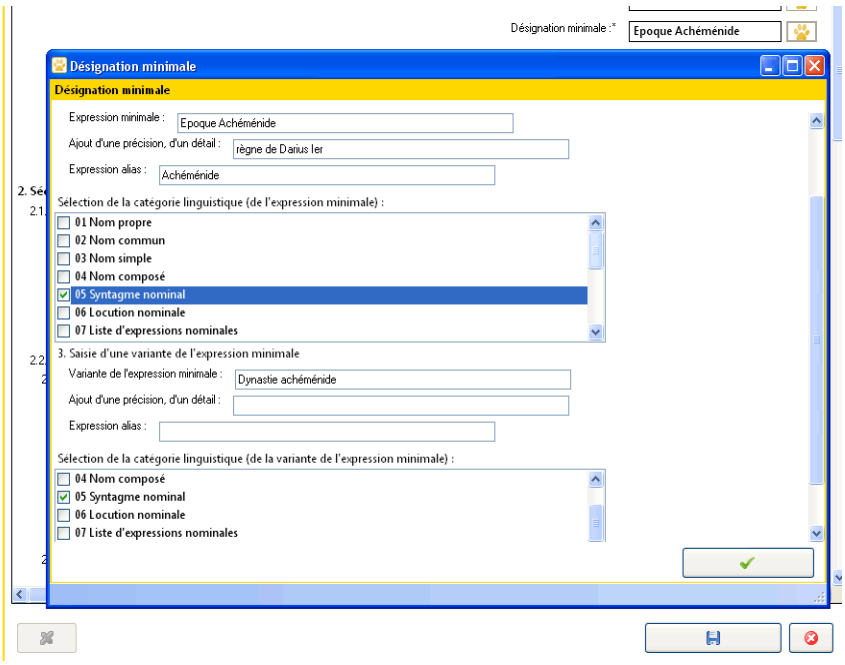

Figure 9.3. *Formulaire standard pour la réalisation de l'activité* [Désignation minimale], *seule activité obligatoire de la procédure de la description libre*

Comme nous le montre la figure 9.3, le formulaire standard qui encadre l'activité [Désignation minimale], est organisé en plusieurs parties. Une première partie est réservée à la saisie de l'*expression minimale* qui, selon l'analyste, est appropriée pour représenter – dans notre cas – le terme conceptuel [Epoque historique]. Dans notre exemple, l'analyste a choisi de le représenter par la valeur minimale <Epoque Achéménide> qui désigne la période de règne de la première dynastie des Empires Perses.

Le formulaire standard associé à l'activité [Désignation minimale] propose à l'analyste plusieurs autres options qu'il peut utiliser s'il souhaite rendre plus riche, plus détaillé les résultats de son activité de désignation d'un terme conceptuel ou d'une configuration de termes conceptuels.

Comme le montre la figure 9.3, l'analyste peut ajouter des mentions explicatives pour resserrer la signification de l'expression minimale, il peut introduire des

expressions alias de l'*expression minimale* et il peut également produire une, voire plusieurs *variantes de l'expression minimale*.

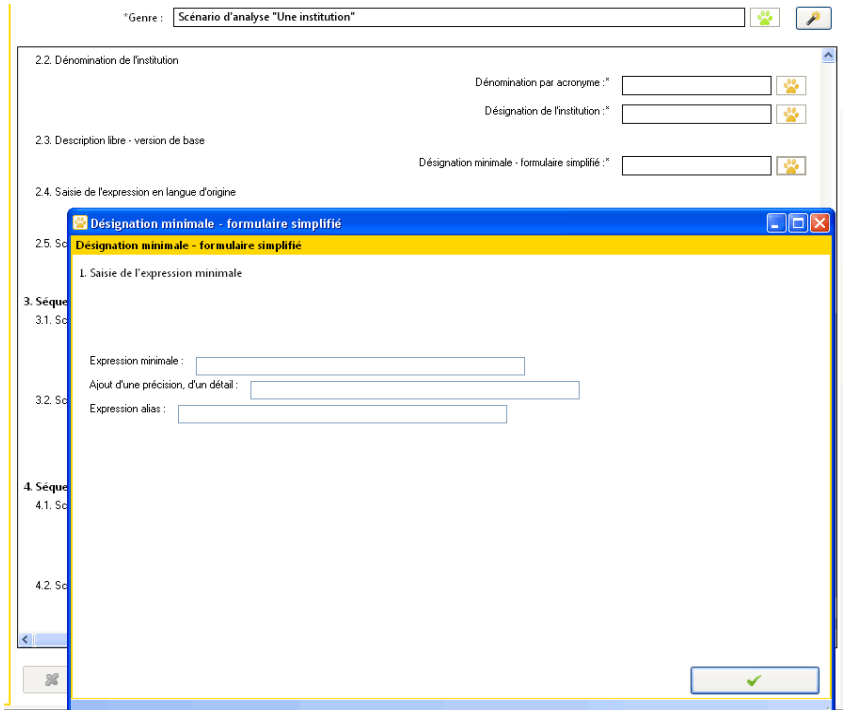

Figure 9.4. *Formulaire simplifié guidant l'analyste dans la réalisation de l'activité* [Désignation minimale]

Enfin, l'analyste est également invité à utiliser une liste de catégories linguistiques pour identifier l'appartenance linguistique de l'expression minimale (ou de sa variante). C'est une option particulièrement utile pour distinguer entre les expressions minimales faisant partie de la catégorie [Nom propre] et celles faisant partie de la catégorie [Nom commun]. En même temps, l'usage systématique de la liste de catégories linguistiques par les analystes de corpus audiovisuels peut faciliter les exploitations terminologiques ultérieures qui se basent sur un traitement (semi-)automatique des banques de données linguistiques générées par les activités de la tâche « indexation libre ».

Insistons sur le fait que l'expression <Epoque Achéménide> (figure 9.3) est la *seule* information *obligatoire* que l'analyste doit fournir pour satisfaire aux exigences formelles d'utilisation du formulaire associé à l'activité [Désignation

minimale]. Cela veut dire aussi, bien sûr, que le formulaire « standard » associé à cette activité de description peut être modifié et adapté à la « politique » spécifique de description propre à une archive audiovisuelle.

Ainsi, la figure 9.4 nous montre l'extrait d'un formulaire d'analyse du contenu utilisé pour identifier et décrire tout type d'institutions (scientifiques, culturelles, politiques, etc.) dont peut être question dans un texte audiovisuel. La partie 2.3 de ce formulaire s'intitule [Description libre – version de base]. Elle est réduite à son expression la plus simple possible : une seule activité de description, à savoir celle, nécessaire, appelée [Désignation minimale] à laquelle est associée un formulaire de saisie d'informations dans lequel on ne trouve que les trois champs suivants : [Expression minimale], [Ajout d'une précision], d'un détail descriptif » et [Expression alias].

L'utilisation systématique de cette version de description libre à laquelle est associé un formulaire de saisie d'information très simplifié permet, bien sûr, un gain de temps important dans l'indexation d'un corpus audiovisuel. Mais ce gain de temps se paye en retour avec une perte de précision de la description. Le système des ressources métalinguistiques ASA permet à l'analyste de faire le choix entre différentes stratégies, différentes politiques de description et d'indexation – mais la responsabilité et les conséquences de celui-ci doivent être assumées par l'analyste.

9.4. L'activité de description [Désignation contextualisée]

La deuxième activité de la procédure de la description libre représentée par la figure 9.1, est appelée [Désignation contextualisée]. Elle invite l'analyste, comme son intitulé l'exprime déjà, à compléter, s'il le juge utile, les informations minimales.

Via cette activité, l'analyste peut notamment faire en sorte que le destinataire – le public – de son travail réussit à évaluer, sans devoir obligatoirement regarder le passage audiovisuel indexé, si celui-ci est réellement pertinent pour lui.

Le formulaire standard que l'analyste peut utiliser pour réaliser cette activité, est, dans sa structure, très proche de celui qu'utilise l'analyste pour réaliser l'activité [Désignation minimale]. Une différence notable est celle de la distinction entre l'entrée [Expression minimale] (figure 9.3) et l'entrée [Expression minimale contextualisée] qui remplace la première dans le formulaire standard attaché à l'activité [Désignation minimale contextualisée] (figure 9.5).

Le « bon usage » des deux formulaires dépend de la compétence et de l'expérience de l'analyste. C'est obligatoirement à l'analyste de définir (en tenant compte du contexte et des objectifs de son analyse) le niveau de pertinence minimal

correspondant à la saisie d'une expression minimale et d'une expression minimale contextualisée appropriées !

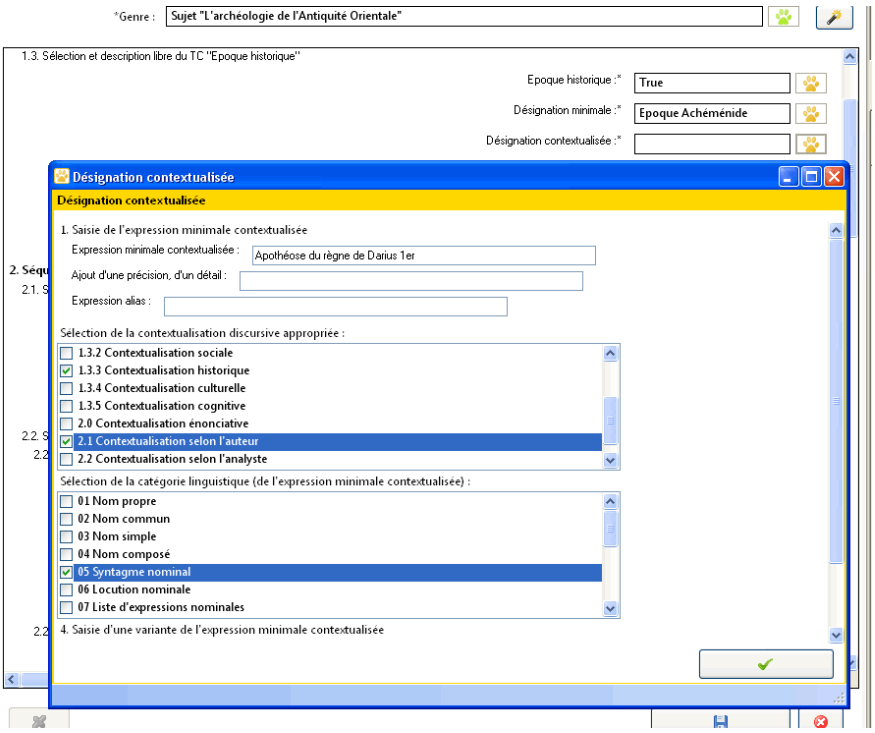

Figure 9.5. *Extrait du formulaire standard guidant l'analyste dans la réalisation de l'activité* [Désignation minimale contextualisée]

S'il le souhaite, l'analyste peut compléter son travail de contextualisation de l'expression minimale, d'une part, par un travail de catégorisation linguistique de l'expression contextualisée et, d'autre part, par l'ajout d'une ou de plusieurs variantes de l'expression contextualisée de référence (voir figure 9.5).

Enfin, comme le montre la figure 9.5, l'analyste est également invitée à spécifier le *type* de contextualisation *qu'il a choisi* pour rendre plus précis, plus parlant l'expression minimale d'un terme conceptuel ou d'une configuration de termes conceptuels. En tenant compte des différentes hypothèses et propositions en analyse du discours, nous distinguons, pour le moment et comme le montre la figure 9.5, entre deux grandes stratégies discursives de contextualisation qui nous rappellent la mise en discours d'une thématique ou d'une topique et son analyse (voir chapitre 7). Il s'agit du *positionnement de l'expression minimale* soit dans un *contexte référentiel* soit dans un *contexte énonciatif*, c'est-à-dire dans le contexte propre à la

production d'un discours thématisant un terme conceptuel ou une constellation de termes conceptuels (voir figure 9.5).

Dans le premier cas, il s'agit typiquement du positionnement d'une expression minimale dans un espace ou dans le temps ou encore dans un environnement social, historique, culturel, voire cognitif (c'est-à-dire de connaissances et de croyances). Dans le deuxième cas, il s'agit du positionnement de l'expression minimale par rapport à l'énonciateur qui produit la contextualisation. L'énonciateur peut être soit l'auteur du discours audiovisuel analysé, soit l'*analyste lui-même*, soit encore une tierce personne comme, par exemple, un expert ou toute autre source d'informations utilisée par l'analyste pour expliciter davantage une expression minimale.

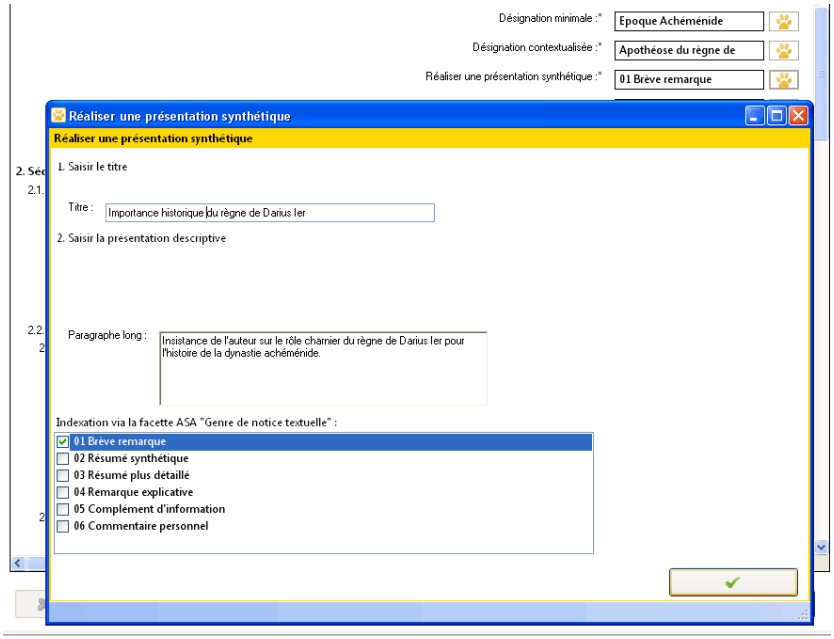

Figure 9.5. *Extrait du formulaire standard guidant l'analyste dans la réalisation de l'activité* [Réaliser une présentation synthétique]

Cependant, il s'agit *ici de catégoriser la contextualisation d'une expression minimale libre*, donc d'une *expression métalinguistique* faisant partie du *métadiscours de l'analyste* ! Il faut très soigneusement distinguer entre, d'une part, la tâche d'analyse de la mise en discours d'un thème ou d'une topique dans un texte audiovisuel qui forme l'objet de l'expertise de l'analyste et, d'autre part, l'effort métalinguistique de rendre plus parlant, plus compréhensible, « plus riche » une description « minimale » du contenu d'un texte audiovisuel !

La classification de la contextualisation de l'expression minimale générée par l'analyste est une stratégie pour résoudre le deuxième point mais pas le premier ! Que nous puissions utiliser les mêmes catégories pour classer la contextualisation discursive d'un thème ou, plutôt, de l'expression linguistique d'un thème, montre seulement qu'il y a une *isomorphie* entre discours et métadiscours (entre le discours qui sert d'objet à une analyse et le discours produit sur le discours servant d'objet à l'analyse) – une des différences les plus importantes entre discours-objet et métadiscours étant que le deuxième s'efforce d'utiliser un langage (un vocabulaire, de modèles, etc.) aussi explicite que possible fondé sur une vision, un cadre théorique de compréhension de l'activité discursive humaine.

9.5. Les activités [Réalisation d'une présentation synthétique] et [Désignation du référent en langue d'origine]

La troisième activité de la procédure de description libre, l'activité *[Réalisation d'une présentation synthétique]*, est une activité qui permet à l'analyste de rédiger une annotation textuelle au sujet de son objet audiovisuel. Le formulaire standard associé à cette activité est relativement simple.

Comme le montre la figure 9.6, il est composé d'un champ [Paragraphe] dans lequel l'analyste peut saisir son commentaire et d'un champ [Titre] que l'analyste peut utiliser s'il souhaite faire précéder son commentaire par un intitulé.

Enfin, l'analyste est invité à classer sa présentation en l'associant à un ou plusieurs genres de notices textuelles tels que <Brève remarque>, <Complément>, <Commentaire personnel>, etc.

La quatrième activité *[Désignation du référent en langue d'origine]* prévoit la possibilité de renseigner, si utile, un terme conceptuel (ou une configuration de termes conceptuels) en la langue d'origine si celle-ci est différente de celle utilisée par l'analyste dans son travail d'analyse. C'est une fonction qui est particulièrement utile si l'analyste souhaite documenter en langue d'origine le nom (d'une ville, d'une institution, d'une œuvre, d'une pratique spécifique, etc.) qu'il emploie comme expression minimale mais sous forme d'une expression faisant partie de sa langue de travail. Ainsi, par exemple, s'il utilise, dans une de ses descriptions, le toponyme <Vienne> comme expression minimale, il pourra toujours indiquer qu'il s'agit ici d'une traduction en langue française de l'expression originale <Wien> faisant partie de la langue allemande.

Comme le montre la figure 9.7, le formulaire standard associé à cette activité spécifique invite l'analyste à choisir la langue d'origine dans un menu déroulant et à saisir, dans cette langue, l'expression et/ou une expression équivalente, voire encore une expression alias.

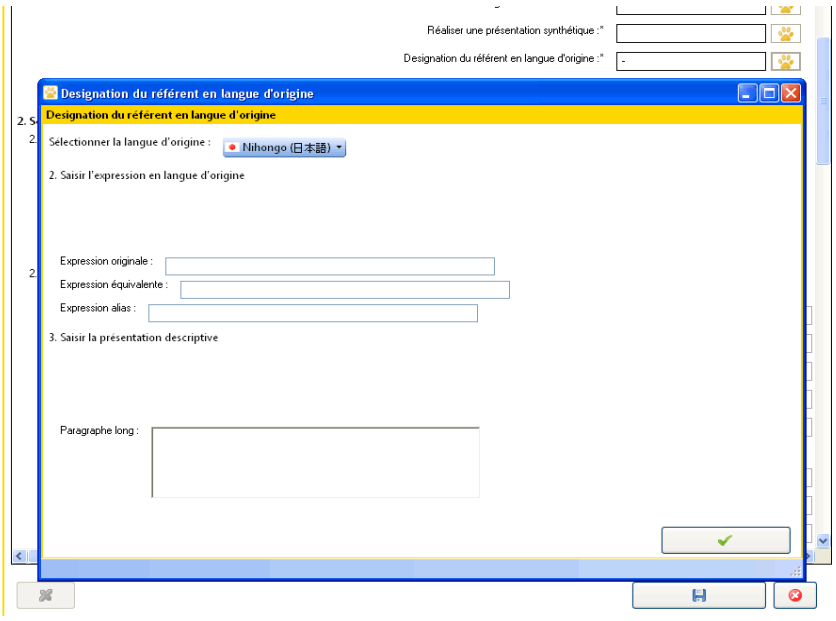

Figure 9.7. *Extrait du formulaire standard guidant l'analyste dans la réalisation de l'activité* [Désignation du référent en langue d'origine]

Si nécessaire, l'analyste peut encore produire une note explicative au sujet de l'expression ou des expressions du terme conceptuel en langue d'origine.

9.6. L'activité de description [Désignation du référent par mot clé]

Enfin, la cinquième activité qui fait partie de notre exemple d'une procédure de description libre (figure 9.1) est l'activité *[Désignation du référent par mots-clés]*. Celle-ci invite l'analyste à produire, s'il le juge utile, un ou une liste d'expressions qui, selon lui, représente ou exemplifie au mieux le terme conceptuel tel qu'il est traité dans l'objet audiovisuel soumis à l'analyse.

Si cette activité est associée à celle intitulée [Désignation minimale], elle a comme seule fonction d'enrichir l'expression minimale – contextualisée ou non – d'un terme conceptuel ou d'une configuration de termes conceptuels. La figure 9.8 montre la version la plus simple de ce procédé qui consiste en l'utilisation d'un formulaire pourvu d'un champ réservé à la saisie d'un ou de plusieurs mots clés. En comparaison à la pratique « standard » d'associer une liste de mots-clés librement choisis par l'utilisateur à un texte audiovisuel (pratique telle qu'elle est couramment

pratiquée, par exemple, sur la plupart des grandes plateformes de partage et de diffusion de vidéos telle que You Tube, Daily Motion, Vimeo, etc.), la seule – mais néanmoins notable – différence consiste dans le fait que chaque liste de mots-clés générés par l'utilisateur est obligatoirement associée soit à un terme conceptuel soit à une configuration de termes conceptuels définissant la structure d'une topique.

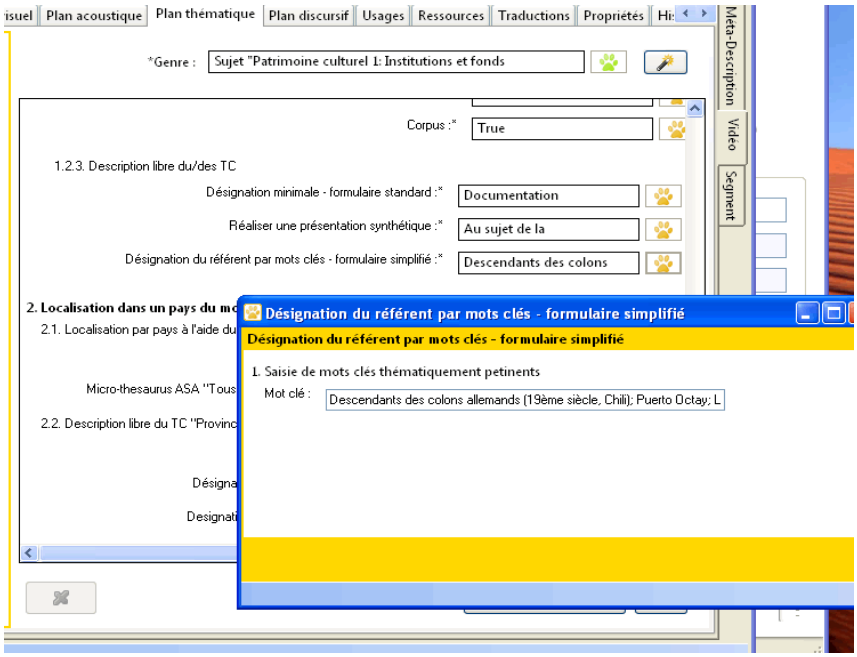

Figure 9.8. *Formulaire minimal guidant l'analyste dans la réalisation de l'activité* [Désignation du référent par mots clés]

Cependant, une description plus explicite par mots-clés a intérêt d'expliciter davantage son procédé. Ainsi, on pourrait attendre d'une version plus sophistiquée de cette activité qu'elle nous indique, par exemple :

– le *type de liste* de mots-clés générés par l'analyste (une simple énumération « associative », une terminologie, un champ lexical, un champ notionnel, etc.) ;
– la *portée empirique* des mots-clés générés par l'analyste (portée globale ou plutôt spécialisée sur tel ou tel aspect de l'objet de connaissance thématisé dans le texte audiovisuel soumis à l'analyse) ;
– l'*origine des mots-clé*s générés par l'analyste (mots-clés provenant du texte audiovisuel lui-même, mots-clés dont l'auteur est l'analyste, mots-clés provenant d'une source autre, etc.) ;

– enfin, éventuellement le rapport qui existe entre la liste de mots-clés générée par l'analyste et l'expression minimale (contextualisée ou non) réalisée lors de l'activité [Désignation minimale] (les mots-clés générés par l'analyste entretiennent-ils avec l'expression minimale un rapport d'*exemplification*, de *développement descriptif*, etc.).

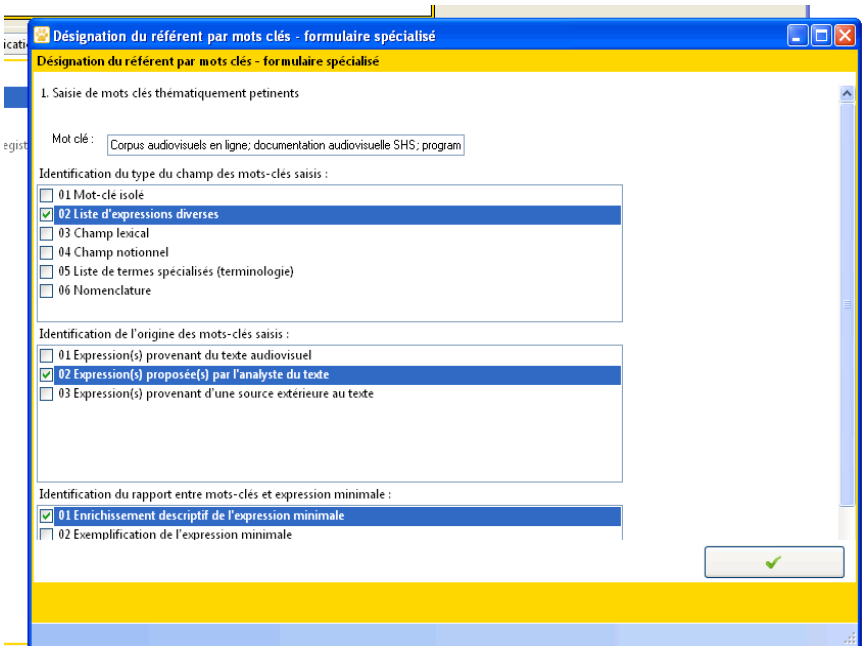

Figure 9.9. *Formulaire expérimental plus élaboré guidant l'analyste dans la réalisation de l'activité* [Désignation du référent par mots clés]

Pour l'instant, nous nous contentons de la version simple de cette procédure (voir figure 9.8). Dans la figure 9.9, on voit cependant une version expérimentale plus élaborée du formulaire que peut utiliser l'analyste afin de préciser, par exemple, le type de la liste de mots-clés générés, l'origine des mots clés utilisés ou encore le rapport que peut exister entre l'expression minimale obligatoirement saisie d'abord par l'analyste et le ou les mots-clés saisis ensuite.

9.7. Variantes pragmatiques et onomasiologiques de l'activité [Désignation minimale]

Comme nous le savons déjà, l'activité [Désignation minimale] est la seule activité obligatoire dans le cadre de la procédure dite de description libre. Mais il s'agit ici d'un type qui connaît toute une diversité d'activités plus spécialisées.

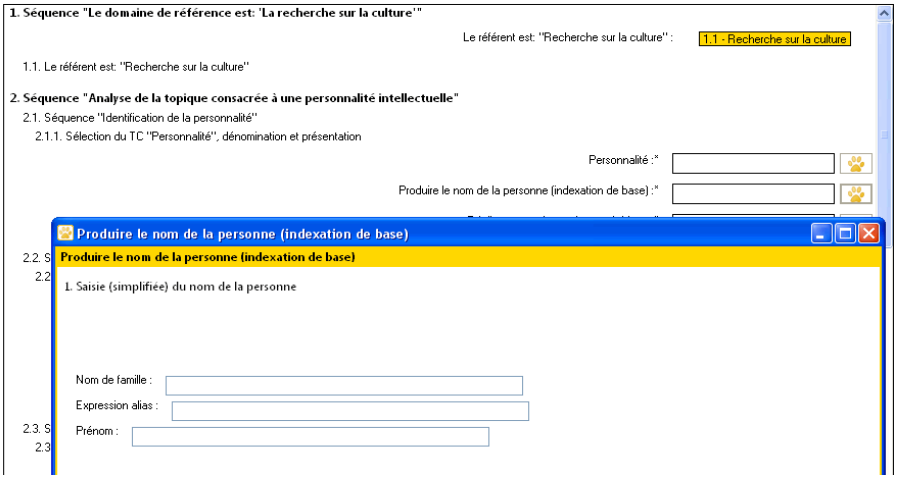

Figure 9.10. *Formulaire simplifié pour produire nom et prénom d'une personne*

Nous avons déjà rencontré deux variantes spécialisées, la première appelée [Désignation minimale standard], la seconde appelée [Désignation minimale simplifiée]. La première variante est réservée à des activités de description détaillées, la seconde à des descriptions rapides et plus superficielles. Bien sûr, cette opposition entre les deux variantes peut être encore différenciée et mieux élaborée sous forme de toute une *liste de variantes* de l'activité de base [Désignation minimale] afin de pouvoir rendre compte de styles et d'options de description les plus divers. En tout cas, il s'agit ici de *variantes dites pragmatiques* de l'activité de base [Désignation minimale].

Mais l'activité [Désignation minimale] comporte encore d'autres types de variantes dont notamment les *variantes onomasiologiques.* Il s'agit des noms (propres) de personnes, de structures collectives, d'œuvres, de marques, de produits, de lieux, de périodes, etc. Parmi les différentes catégories de noms propres identifiées par la recherche onomasiologique (voir à ce propos [EIC 95]), nous avons retenus surtout la catégorie des *anthroponymes* (noms de personne, patronymes, matronymes, pseudonymes, agnoms, surnoms, etc.), des *ethnonymes*

(noms de peuple, d'ethnie, de communautés, etc.), des *ergonymes* (noms de produits, de réalisations, etc.), des *toponymes* et des *chrononymes*.

La figure 9.10 nous montre l'exemple d'une interface de travail d'une des *variantes onomasiologiques* de l'activité [Désignation minimale]. Il s'agit ici de l'interface permettant la saisie du nom et du prénom d'une personne (précisons qu'il s'agit ici d'une interface simplifiée pour dénommer une personne ; si une personne doit être identifiée par son matronyme, patronyme, agnom ou autre surnom, il faut utiliser un autre formulaire de travail).

Enfin, la version standard de l'activité [Désignation minimale], intègre la possibilité de classer linguistiquement une expression minimale saisie. Comme le montre la figure 9.3, la première catégorie linguistique qui s'offre à l'analyste pour classer une expression minimale, est bien celle du *nom propre*. Or, rien n'empêche de développer davantage cette catégorie linguistique en la remplaçant par les différentes catégories de noms propres identifiées et utilisées dans la recherche onomasiologique. Ainsi, le formulaire de travail représenté dans la figure 9.10 pourra se substituer aux différents formulaires utilisés pour saisir les expressions minimales relevant de telle ou telle sous-catégorie spécifique de la catégorie linguistique du *nom propre*.

Chapitre 10

La procédure de la description contrôlée d'un corpus audiovisuel

10.1 Introduction

A côté de la procédure de la *description libre** qui offre à l'analyste la possibilité de produire ses propres métadonnées relatives à un texte ou à un corpus audiovisuel, nous utilisons un deuxième type de procédure de description et d'indexation qui est entièrement ou partiellement guidé par l'usage d'un *thesaurus**. Ce type de procédure est appelée *procédure de description contrôlée*.

Les avantages de cette procédure sont doubles. Un premier avantage est celui d'un gain de temps considérable dans l'analyse même d'un texte audiovisuel. C'est le thesaurus qui, en quelque sorte, impose à l'analyste le choix des expressions les plus appropriées pour décrire le contenu d'un texte ou d'un corpus audiovisuel. Comme nous le verrons encore, une procédure concrète de description contrôlée ne sollicite jamais l'ensemble du thesaurus mais seulement une de ses multiples *facettes* qui couvre le *champ des valeurs référentielles* d'un terme conceptuel ou d'une configuration de termes conceptuels à renseigner par l'analyste. Le champ des valeurs référentielles d'une facette est constitué d'une (voire de plusieurs) listes (hiérarchiques ou non) d'*expressions normalisées* (de descripteurs). Mais pratiquement aucune de ces listes comportent un nombre élevé d'expressions

normalisées n'exigeant donc pas, de la part de l'analyste, un effort énorme pour être maitrisé convenablement

Un deuxième avantage réside dans le fait que les expressions censées représenter le contenu d'un texte ou d'un corpus audiovisuel possèdent toutes la même *forme* (orthographique, linéaire) *normalisée* étant donné qu'elle est déterminée *a priori* et non pas générée librement par l'analyste. C'est un avantage important si on considère notamment que les expressions saisies par l'analyste, pour décrire le contenu d'un texte audiovisuel, constituent, entre autres, les *expressions des interfaces de recherche* et de *localisation* de contenus audiovisuels dans une bibliothèque ou une archive numérique donnée.

Mais, l'usage d'un thesaurus possède aussi ses limites évidentes dont les deux les plus importantes nous semblent être d'abord celle de l'exhaustivité empirique et ensuite celle d'imposer à l'analyste une terminologie qu'il ne souhaite pas obligatoirement utiliser. Nous y reviendrons davantage au chapitre 15 consacré à une discussion plus systématique de l'organisation et du rôle du thesaurus dans notre approche de description du contenu d'un texte ou d'un corpus de textes audiovisuels.

Dans la section 10.2 nous présentera rapidement la structure canonique de la procédure de la description contrôlée.

La procédure de description contrôlée n'est pas obligatoirement limitée à l'usage d'un seul thesaurus. Au contraire, elle peut s'appuyer sur deux ou plusieurs thésaurus, chacun exprimant un « point de vue » particulier. La section 10.3 montre l'exemple concret d'une procédure de description contrôlée qui utilise deux thesaurus différents.

Dans la section 10.4 nous considérerons rapidement les principales fonctions que remplit le thesaurus comme outil d'indexation dans le cadre d'une description contrôlée. Nous parlerons plus particulièrement de trois fonctions essentielles que sont la fonction de l'*identification* d'un terme conceptuel (ou d'une configuration de termes conceptuels) par sa valeur appropriée, la fonction de la *classification* et, enfin, de la fonction de l'*évaluation* ou de l'*appréciation* d'un terme conceptuel instancié.

Enfin, nous discuterons encore, dans la section 10.5, l'approche combinée qui se base à la fois sur la procédure de description libre et celle de description contrôlée.

10.2 Anatomie de la procédure de la *description contrôlée*

La figure 10.1 nous montre une version très simple mais récurrente de la procédure de la description contrôlée (au moins dans le cadre de notre recherche

consacrée à la description de corpus audiovisuels documentant de patrimoines scientifiques et, plus généralement, culturels). Il s'agit de l'extrait d'un modèle d'analyse de corpus audiovisuels consacrés aux langues naturelles.

On y voit le terme conceptuel [Langue naturelle] qui fait partie du méta-lexique des *termes conceptuels** ASA accompagné de deux *activités d'analyse** spécifiques et simples qui sont :

– l'activité [(Sélection dans le) Micro-thesaurus ASA « Langues naturelles par ordre alphabétique] ;
– et l'activité [Réalisation d'une présentation synthétique].

Seule l'activité de la sélection d'une ou de plusieurs langues dans la liste des langues composant le micro-thesaurus ASA (figure 10.1) est obligatoire. Elle possède ainsi le même statut que celle appelée [Désignation minimale] dans la procédure de la *description libre**.

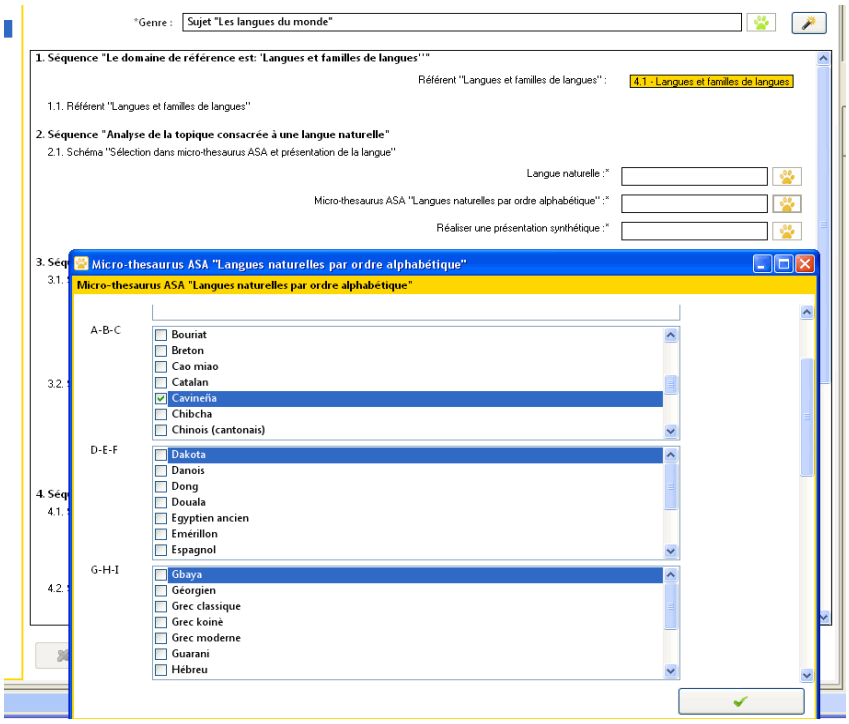

Figure 10.1. *La procédure de la description contrôlée – version canonique*

La deuxième activité composant la tâche de la description contrôlée (figure 10.1), l'activité *[Réalisation d'une présentation synthétique]*, est facultative. Elle permet à l'analyste, comme nous l'avons déjà vu dans le chapitre précédent, d'expliciter davantage les spécificités d'un sujet développé dans le texte audiovisuel soumis à l'analyse sous forme, par exemple, de commentaires personnels, de compléments d'information, etc.

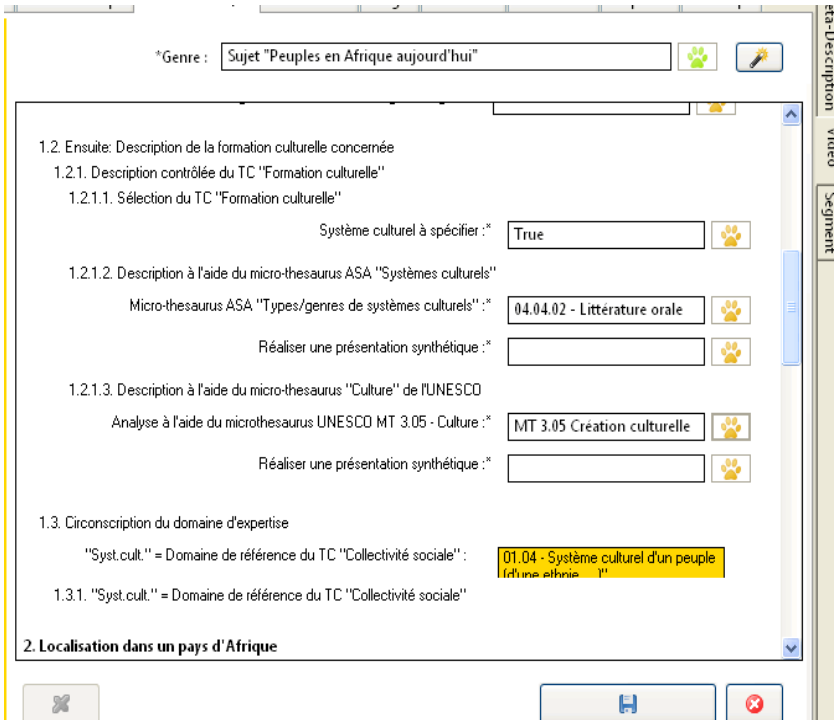

Figure 10.2. *La procédure de la description contrôlée se réalisant à l'aide de deux micro-thesaurus différents*

Prenons un exemple concret (voir figure 10.1) : s'il est question, dans un entretien avec un spécialiste de langues amérindiennes, des langues *Cavineña* et/ou *Reyesano*[47], l'analyste coche simplement, dans la liste des langues composant le

[47] Nous nous référons ici à un entretien avec le linguiste Antoine Guillaume, spécialiste des deux langues citées qui font partie de la famille des langues *tacananes* parlées en Bolivie (entretien réalisé par Elisabeth de Pablo dans le cadre du programme « Archives Audiovisuelles de la Recherche » en 2006 au Laboratoire *Dynamique du Langage (DDL)* à Lyon)

micro-thesaurus *ASA* des langues par ordre alphabétique, l'une ou l'autre ou les deux langues en question.

Nous appelons *micro-thesaurus* la liste (hiérarchique ou non) des expressions normalisées (des descripteurs) qui désignent les *valeurs* que peut prendre un terme conceptuel ou une configuration de termes conceptuels. Autrement dit, le *micro-thesaurus* est l'ensemble des expressions qui interprètent une *facette* tandis qu'une facette représente un *axe* ou une *dimension du contenu* (du *signifié*, au sens structuraliste du terme) d'un *terme conceptuel* ou d'une *configuration de termes conceptuels*.

Le contenu ou le signifié d'un terme conceptuel peut être interprété, bien entendu, par toute une diversité de facettes. A son tour, une expression (un « descripteur ») peut faire partie de plusieurs micro-thesaurus, peut interpréter différentes facettes d'un terme conceptuel ou d'une configuration de termes conceptuels. Nous y reviendrons plus en détail dans le chapitre 15.

En tout cas, un thesaurus, lui, est essentiellement composé d'un ensemble de *facettes* et d'un ensemble d'*expressions normalisées* pouvant faire partie d'une ou de plusieurs facettes.

10.3 Travailler avec plusieurs micro-thesaurus

Dans la même procédure de description contrôlée d'un texte audiovisuel, l'analyste peut recourir, à la limite, à deux ou plusieurs micro-thesaurus différents.

La figure 10.2 nous montre comme exemple la description et l'indexation d'une formation culturelle spécifique à l'aide d'une part du micro-thesaurus ASA « Systèmes Culturels » et d'autre part du micro-thesaurus de l'UNESCO MT 3.05 « Culture »[48].

L'intégration de deux ou plusieurs micro-thesaurus dans une même procédure de description montre qu'il est parfaitement possible de concevoir et de réaliser des *analyses contrôlées* et *normalisées* qui tiennent compte de *différents points de vue*, de différentes « logiques sociales ou institutionnelles » (si on accepte l'hypothèse qu'un thesaurus représente, *grosso modo*, une vision, un point de vue propre à « son » acteur qui peut être une institution ou une organisation sociale). En ce qui concerne notre exemple (figure 10.2), la description contrôlée d'une formation culturelle s'appuie à la fois sur le point de vue du projet ASA-SHS et sur le point de vue de l'UNESCO. Mais rien n'empêche, si le contexte et les objectifs d'une analyse l'exigent, de faire varier ces points de vue. La seule exigence est celle que le ou les micro-thesaurus utilisés développe, chacun, sous forme d'une liste

48 Voir http://databases.unesco.org/thesfr/

hiérarchique ou non les *expressions normalisées* qui constituent les valeurs d'une *facette* d'un terme ou d'une configuration de termes conceptuels composant l'ontologie du domaine *univers du discours** ASA (voir la partie IV de ce livre).

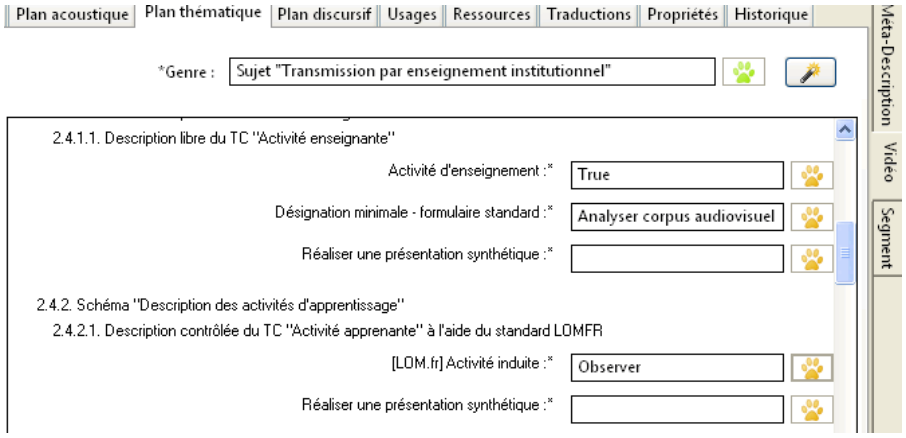

Figure 10.3. *La procédure de la description contrôlée se réalisant à l'aide d'un thesaurus LOMFR*

Signalons ici un cas particulier qui est celui de l'utilisation locale d'un standard (ou, le plus souvent, d'une partie d'un standard) lors de la description contrôlée d'un terme ou d'une configuration de termes conceptuels. La figure 10.3 montre l'extrait d'un formulaire de description servant à analyser des enregistrements de manifestations pédagogiques (cours, ateliers de formation, etc.). Etant donné l'objet particulier de l'analyse, il est normal qu'on essaie de respecter, dans la mesure du possible, les exigences d'un standard tel que LOM (*Learning Objet Metadata*) ou, pour la France, LOMFR (correspondant à la norme française NF Z76-040, publiée en 2006 par l'AFNOR)[49]. Dans notre exemple, il s'agit de spécifier la ou les activités induites du côté de l'apprenant dans le cadre d'un cours portant sur l'analyse de corpus audiovisuels. La spécification de la ou des activités se fait via un petit micro-thesaurus qui intègre le vocabulaire proposé à cet effet par LOMFR. Il est tout à fait possible de compléter ce vocabulaire normalisé si un type d'activités n'est pas retenu dans le vocabulaire fourni par LOMFR.

Sans vouloir développer ici davantage la possibilité de travailler avec un, deux ou plusieurs micro-thesaurus dans le cadre d'une procédure de description contrôlée, précisons, que cette possibilité offerte à l'analyste, tient compte du fait que tout axe ou toute dimension du contenu (du signifié) d'un terme conceptuel est forcément de

[49] Pour plus d'informations, voir : http://www.lom-fr.fr/

nature *indexicale*. En d'autres termes et d'une manière générale, nous devons tenir compte du fait que le *sens spécifique* d'un terme conceptuel (d'une configuration de termes conceptuels), sa – pour parler ainsi – *facette*, se forme (et se stabilise) en relation :

– à la *spécificité de l'objet* auquel réfère le terme (la configuration des termes conceptuels) en question ;
– à la *compétence* (ou à la *culture*, à la *connaissance*) et aux *attentes* des acteurs qui se servent du terme conceptuel (de la configuration des termes conceptuels) ;
– et, enfin, aux *contextes* dans lesquels le terme ou la configuration des termes conceptuels sont utilisés, employés.

Ce triple rapport caractérise intrinsèquement le sens spécifique d'un terme conceptuel ou d'une configuration de termes conceptuels. Il est largement responsable des variations dites contextuelles du sens d'un terme (d'une configuration de termes conceptuels), c'est-à-dire du fait que le même terme conceptuel peut être compris, interprété et surtout utilisé de façon parfois seulement légèrement différente si le contexte d'usage change ou si les acteurs qui s'en servent changent ou encore si l'objet auquel il réfère subit des changements graduels (dus, par exemple, à une évolution historique, sociale, etc.). Comme on le sait, le fait que le sens d'un terme conceptuel soit toujours indexé, ce fait n'a pas toujours bien été compris notamment dans les recherches « classiques » sur les terminologies et les langages de spécialité (les « Fachsprachen », en allemand).

10.4 Sélectionner, classer et hiérarchiser à l'aide d'un micro-thesaurus

Comme déjà dit ci-dessus, nous considérons qu'une facette représente un axe ou une dimension du contenu (du signifié) d'un terme conceptuel. Les expressions normalisées composant le micro-thesaurus d'une facette forment les valeurs (référentielles) que peut prendre un terme conceptuel (ou une configuration de termes conceptuels) dans un contexte d'analyse donné – contexte déterminé entre autre d'une part par les objectifs de l'analyse et d'autre part par la spécificité de l'*univers du discours** d'une archive.

En considérant maintenant de plus près les usages qu'on peut faire d'un micro-thesaurus dans le cadre de la procédure de la description contrôlée, on peut se rendre compte que les dimensions du contenu ne se distinguent pas seulement – ce qui est évident – du point de vue de l'objet ou du domaine *référentiel* auquel elles correspondent (en référence à leur contexte d'emploi) mais aussi du point de vue *fonctionnel*.

En ce qui concerne l'objet ou le domaine référentiel, nous distinguons, bien sûr, toute une diversité de facettes qui servent à interpréter les termes conceptuels du méta-lexique ASA pour lesquels elles ont été définies. Ainsi :

– le micro-thesaurus « Langues du monde » correspond au terme conceptuel [Langue naturelle] ;
– le micro-thesaurus « Pays du monde (début 21ème siècle) » correspond à la configuration conceptuelle partiellement instanciée {[Pays] + {[Siècle : <21ème siècle>] + [Moment :<Début>]}} ;
– le micro-thesaurus « Auteurs de la littérature française » correspond à la configuration conceptuelle partiellement instanciée {[Auteur] + [Littérature par origine : <Littérature française>] + [Siècle : <20ème siècle>]} ;
– etc.

Tous ces micro-thesaurus servent, avant tout, à identifier une valeur spécifique à un terme conceptuel : [Langue] -> [Langue : <Quechua>] ; [Pays] -> [Pays : <France>] ;, etc.

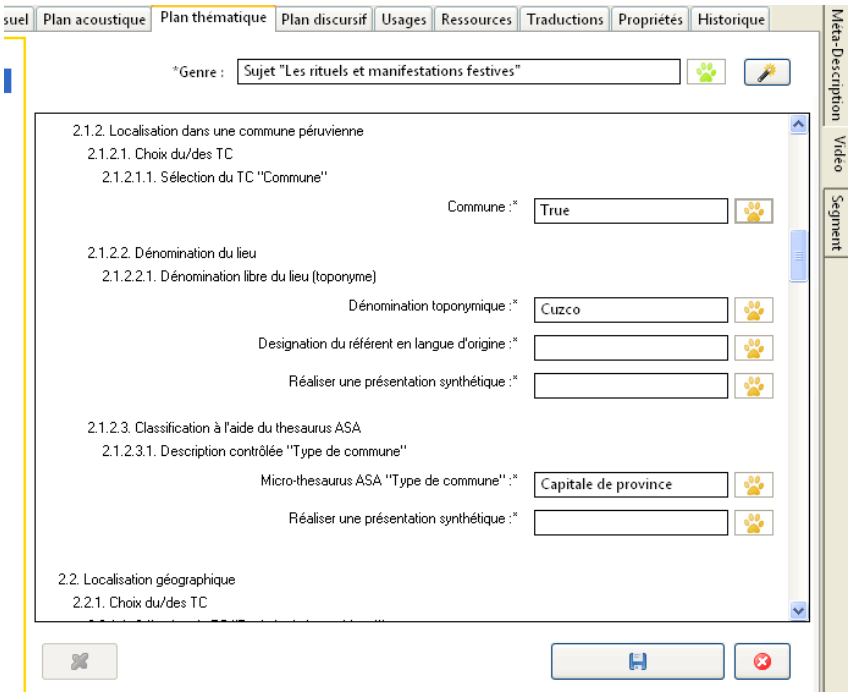

Figure 10.4. *La classification d'une expression saisie à l'aide d'un micro-thesaurus*

Mais outre sa fonction identificatrice évidente, un thesaurus peut également servir à d'autres fonctions très importantes dans la description d'un texte ou d'un corpus de textes audiovisuels.

Mentionnons ici le fait qu'il puisse aussi remplir la fonction centrale, à toute activité humaine, de la classification. Comme on le sait, la classification au sens d'une pratique sociale est indispensable non seulement pour la maîtrise (toujours relative) de l'environnement pertinent à l'homme mais aussi pour se distinguer de l'autre et pour se donner une identité (nous renvoyons à ce propos à Lévi-Strauss et à son ouvrage magistral *Pensée Sauvage* [LEV 62]. Ainsi, en parlant de la fonction de classification qu'occupe le thesaurus dans la procédure de description contrôlée, nous ne pensons pas seulement aux classifications dites scientifiques (par exemple en biologie ou en linguistique) qui peuvent être utilisées, exploitées dans le cadre d'une analyse d'un corpus audiovisuel mais aussi aux diverses ethno- socio-taxinomies, aux *folksonomies*, etc. Après tout – l'analyse d'un corpus audiovisuel peut parfaitement bien être conçue et réalisée du point de vue des connaissances et croyances propres à un groupe, une communauté, voire propre à une époque donnée.

Enfin, citons encore une troisième fonction que l'usage du thesaurus peut revêtir – fonction qui nous paraît essentielle pour l'analyse d'un texte ou d'un corpus de textes audiovisuels. Il s'agit de la *fonction évaluative* ou *appréciative* qui fournit à l'analyste la possibilité de hiérarchiser l'objet analysé ou telle ou telle partie de cet objet. L'appréciation ne peut pas seulement être de nature cognitive permettant de hiérarchiser l'objet analysé, par exemple, selon le point de vue de son informativité ou de sa nouveauté épistémique pour un public donné. Elle peut être aussi de nature plutôt morale (ou moralisante), esthétique, émotionnelle, pratique et utilitaire. L'utilisation des micro-thesaurus appropriés permet à l'analyste de documenter assez facilement ses préférences et/ou ses réserves quant à la qualité (intellectuelle, morale, esthétique, utilitaire, etc.) de l'objet qu'il décrit.

10.5 Une approche combinée de description contrôlée et libre

Dans le cadre du projet de recherche ASA-SHS, nous avons recours également à une procédure d'analyse qui fait appel aussi bien à la *description libre** qu'à la description dite *contrôlée*. La figure 10.5 montre comment ces deux types de procédures peuvent être combinés.

L'exemple concret est tiré des archives ALIA (« Atelier Littéraire d'Ici et d'Ailleurs ») consacré au patrimoine littéraire. Le modèle d'analyse en question est dédié à l'identification et la présentation, dans un corpus audiovisuel donné mais ouvert (c'est-à-dire s'enrichissant au fur et à mesure), d'un ou plusieurs auteurs de la littérature française du XXe siècle. Ce modèle est pourvu d'un thesaurus ASA d'auteurs de la littérature française mais, bien évidemment, il est intrinsèquement impossible, d'en avoir une version complète.

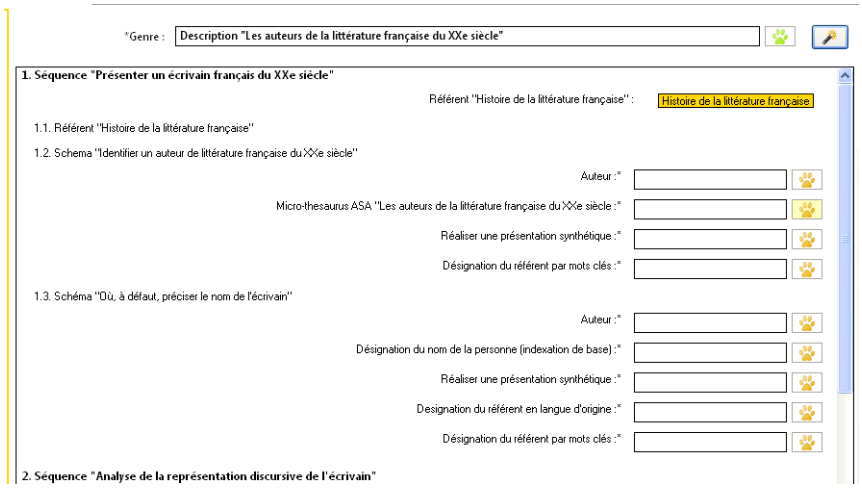

Figure 10.5. *Une approche combinée « description libre et contrôlée »*

En considérant de plus près cette combinaison *Description contrôlée + Description libre*, nous pouvons y déceler deux fonctions particulières :

– une première est celle de compléter une description contrôlée par une description libre – ce cas nous est montré dans la figure 10.5 ;
– une deuxième fonction est celle de préciser, de développer une description contrôlée par une description libre.

Cette deuxième fonction nous est représentée dans la figure 10.6 qui nous montre l'exemple de la description du sujet relatif à une discipline en sciences humaines et sociales – description contrôlée – complétée éventuellement par l'identification et la description minimale (contextualisée ou non) d'une branche de discipline spécifique.

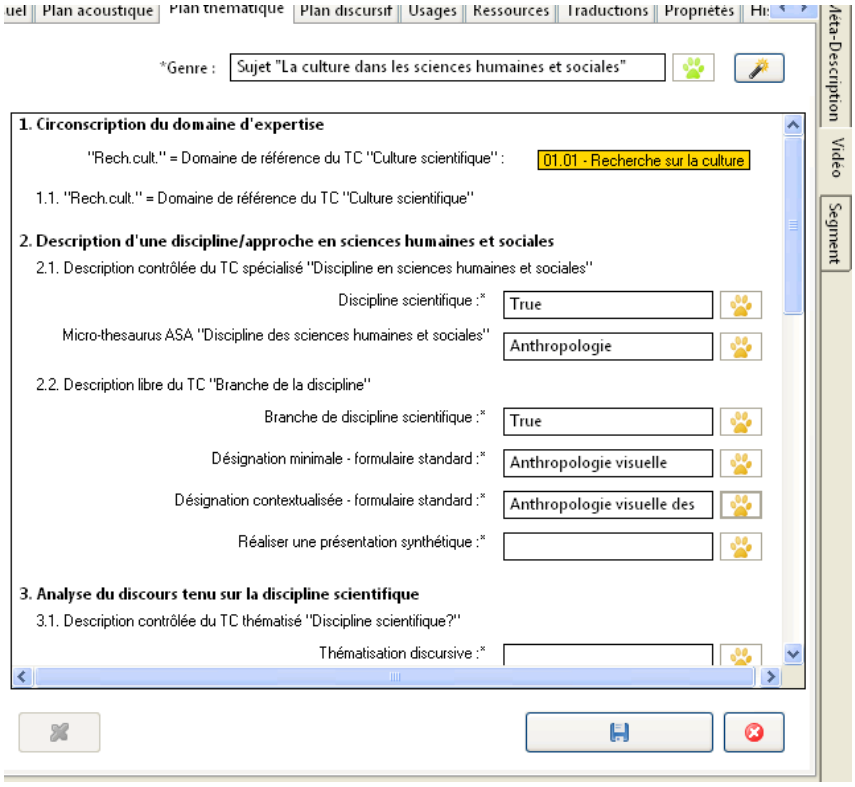

Figure 10.6. *La description libre développe, enrichit la description contrôlée : l'exemple de l'identification (libre) d'une sous-discipline à partir d'une discipline sélectionnée dans le micro-thesaurus ASA « Disciplines scientifiques »*

Concrètement parlant, si sont développées, dans un passage vidéo soumis à l'analyse, des informations relatives à l'anthropologie visuelle, l'analyste peut sélectionner dans la liste ASA la discipline <Anthropologie> et présenter brièvement de quoi est question dans le passage vidéo en question.

Il peut se contenter avec une telle description mais il peut également utiliser le deuxième schéma de la séquence « Analyse de la topique consacrée aux disciplines et approches en sciences humaines et sociales », à savoir le schéma *Description libre du TC [Branche de la discipline]*, et saisir le terme <Anthropologie visuelle> en tant qu'expression minimale via le formulaire standard associé à l'activité [Désignation minimale].

Il peut également contextualiser le terme « anthropologie visuelle » via le formulaire associée à l'activité [Désignation contextualisée] en précisant, par

exemple, un contexte soit référentiel, soit énonciatif particulier dans lequel ce terme est utilisé dans le passage vidéo en question. En tout cas, la procédure *Description libre* remplit bel et bien ici une fonction de précision, d'approfondissement d'une information générale fournie par la procédure *Description contrôlée* (voir figure 10.7).

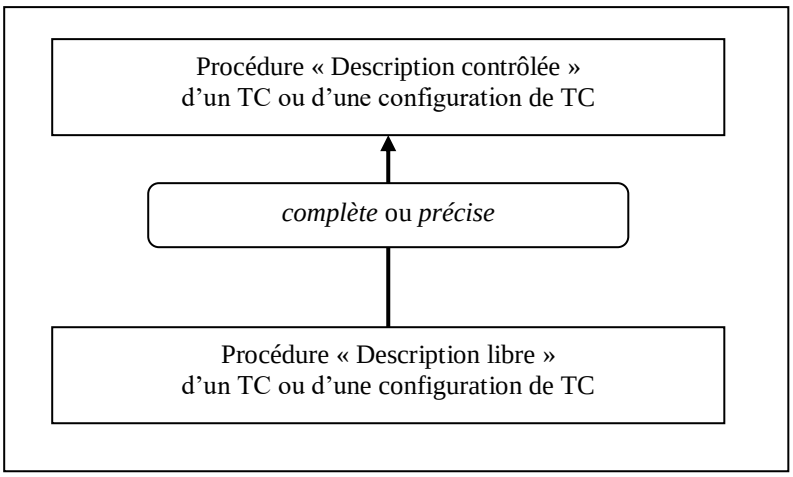

Figure 10.7. *Anatomie des rapports entre description contrôlée et description libre*

Quatrième partie

Le système des ressources métalinguistiques ASA

Chapitre 11

Une vue d'ensemble des ressources métalinguistiques ASA

11.1. Introduction

La possibilité de définir, de développer et d'adapter un *modèle de description** aux exigences d'analystes et de communautés d'analystes les plus différentes, nous est fournie par un ensemble intégré de ressources métalinguistiques qui constituent le *métalangage de description de l'univers du discours ASA* – métalangage à l'aide duquel les analystes peuvent traiter leurs objets et corpus audiovisuels afin de les transformer en des ressources intellectuelles *sui generis* pour des publics et des contextes d'utilisation spécifiques.

Dans ce chapitre, nous allons présenter d'une manière globale, le système des ressources métalinguistiques ASA ; les chapitres 12 à 15 seront consacrés à une discussion plus détaillée des différentes parties de ce système métalinguistique.

La section 11.2 est consacrée à une présentation synthétique du système des ressources métalinguistiques ASA. Nous montrerons également les principaux rapports qui existent entre ces ressources métalinguistiques et l'interface de travail (du Studio ASA) telle qu'elle se présente à l'analyste.

La section 11.3 introduit rapidement le méta-lexique des termes conceptuels ASA composé de deux vocabulaires conceptuels complémentaires, l'un consacré à la représentation des *objets d'analyse** et l'autre consacré aux *activités d'analyse**.

Dans la section 11.4 nous présenterons d'une manière très synthétique un autre type de ressources métalinguistiques qui est celui du thesaurus ASA. Comme nous l'avons déjà vu dans les chapitres précédents, le thesaurus ASA joue un rôle essentiel dans la cadre d'une procédure de description dite contrôlée.

Les sections 11.5 et 11.6 sont réservées à la présentation des *briques** configurationnelles qui composent les modèles de description. Comme également déjà expliqué dans les chapitres précédents, nous distinguons entre deux classes de telles briques : les séquences de description et les schémas de description.

Dans la section 11.7, nous aborderons la question des ressources métalinguistiques extérieures de celles composant le système métalinguistique ASA.

Enfin, dans la section 11.8, nous présenterons encore rapidement l'environnement de travail *Atelier de Modélisation ASA* à l'aide duquel nous définissons et gérons d'une part le métalangage générique de l'*univers du discours** ASA et d'autre part tous les *métalangages de domaine** propres à l'univers du discours d'une archive audiovisuelle particulière.

11.2. Présentation globale du système des ressources métalinguistiques ASA

La figure 11.1 montre les différentes parties qui composent le système des ressources métalinguistiques ASA. Nous y distinguons d'abord la partie *Bibliothèque de modèles de description*. Une bibliothèque de modèles de description est composée au minimum d'une, mais en règle générale de plusieurs *modèles de description** spécifiques qui représentent la *vision* de l'*univers du discours* d'une archive audiovisuelle donnée*. La *vision* de l'univers du discours d'une archive peut évoluer en fonction des intérêts, des objectifs ou tout simplement en fonction de son objet textuel. Ainsi une bibliothèque de modèles peut subir des changements. Mais les conséquences éventuelles d'un tel changement sur les analyses déjà réalisées doivent être évaluées avec précaution.

Une bibliothèque de modèles de description est composée de modèles spécialisés dans l'analyse d'une partie ou d'un aspect spécifique de l'objet textuel tel qu'il est appréhendé dans une perspective *sémiotique* (voir chapitre 1.6). Ainsi, distinguons-nous, comme déjà expliqué dans la première partie de ce livre, entre des modèles qui nous servent à réaliser :

– la description d'une analyse elle-même (c'est-à-dire la tâche de la *méta-description**) ;
– la *description paratextuelle** (c'est-à-dire de l'identité formelle du texte audiovisuel dans une perspective tout à fait comparable à celle que nous dresse le standard Dublin Core) ;

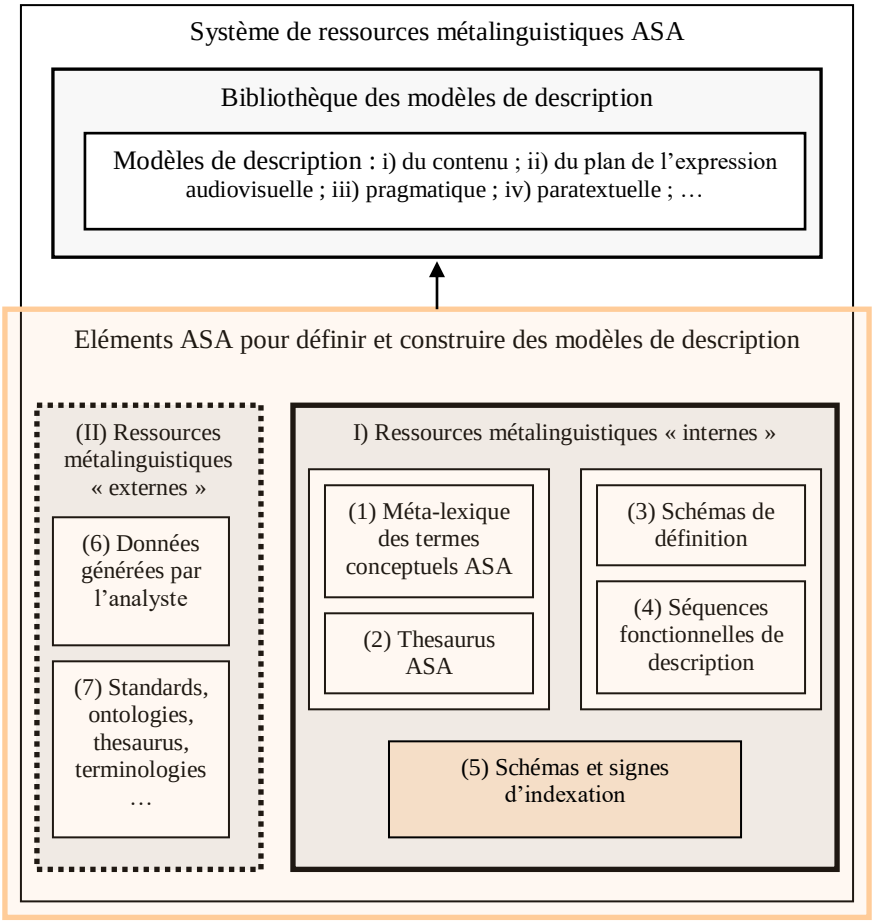

Figure 11.1. *Vue d'ensemble des ressources métalinguistiques ASA*

– la *description audiovisuelle** *stricto sensu* (c'est-à-dire des plans visuels, acoustiques, et/ou audiovisuels du texte audiovisuel) ;
– la *description thématique** (c'est-à-dire des sujets développés dans le texte audiovisuel) ;

– la *description pragmatique** (c'est-à-dire l'ouverture de l'identité, de la spécificité du texte audiovisuel à la culture et aux attentes d'un public ainsi qu'aux exigences propres au contexte dans lequel il doit être utilisé) ;
– l'adaptation-traduction à proprement parler que nous rangeons également sous l'étiquette de *description pragmatique** [STO 08, STO 11a] et qui vise de surmonter l'obstacle de la langue dans la réception et l'appropriation de l'objet audiovisuel.

Ces différents types de modèles de description organisent, comme nous le savons déjà, l'interface de travail de l'Atelier de Description du Studio ASA développé par notre équipe de recherche ESCoM dans le cadre du projet de recherche ASA-SHS[50].

La moitié inférieure de la figure 11.1 montre les différentes parties du système des ressources métalinguistiques ASA qui permettent la construction d'un modèle de description spécifique ou d'une bibliothèque de tels modèles. Nous y distinguons deux ensembles complémentaires :

1. l'ensemble des ressources métalinguistiques qui font partie du système ASA,
2. et l'ensemble des ressources qui lui sont externes mais qui sont *mis en relation avec ce dernier*.

L'ensemble des *ressources métalinguistiques propres au système ASA* connaît trois classes de ressources plus spécifiques et fonctionnellement différentes :

– une classe de ressources lexicales constituées d'une part d'un méta-lexique hiérarchique de *termes conceptuels génériques* et d'autre part d'un vocabulaire contrôlé qui est le *thesaurus du système ASA* ;
– une classe de ressources *structurales* ou *configurationnelles* qui sélectionnent et positionnent les termes génériques et/ou du thesaurus les uns par rapport aux autres selon les spécificités d'un domaine d'expertise donné et selon les besoins de l'analyse ;
– une classe de ressources – appelée *schémas d'indexation* – qui permettent de réaliser une description, de l'exécuter.

La figure 11.1 identifie également un ensemble de ressources métalinguistiques qui sont externes au système ASA. On y trouve d'abord les *données générées par les analystes eux-mêmes* qui se servent des modèles de description ASA pour traiter (décrire, indexer, annoter, etc.) leurs corpus audiovisuels. Ces données générées par les analystes constituent une base d'expressions sémio-linguistiques (verbales mais aussi iconiques, acoustiques, etc.) qui peut servir de référentiel à des nouvelles analyses du même texte audiovisuel.

[50] Pour plus d'informations, voir le site du projet ASA-SHS : http://www.asa-shs.fr/ ainsi que le carnet de recherche consacré au même projet : http://asashs.hypotheses.org/

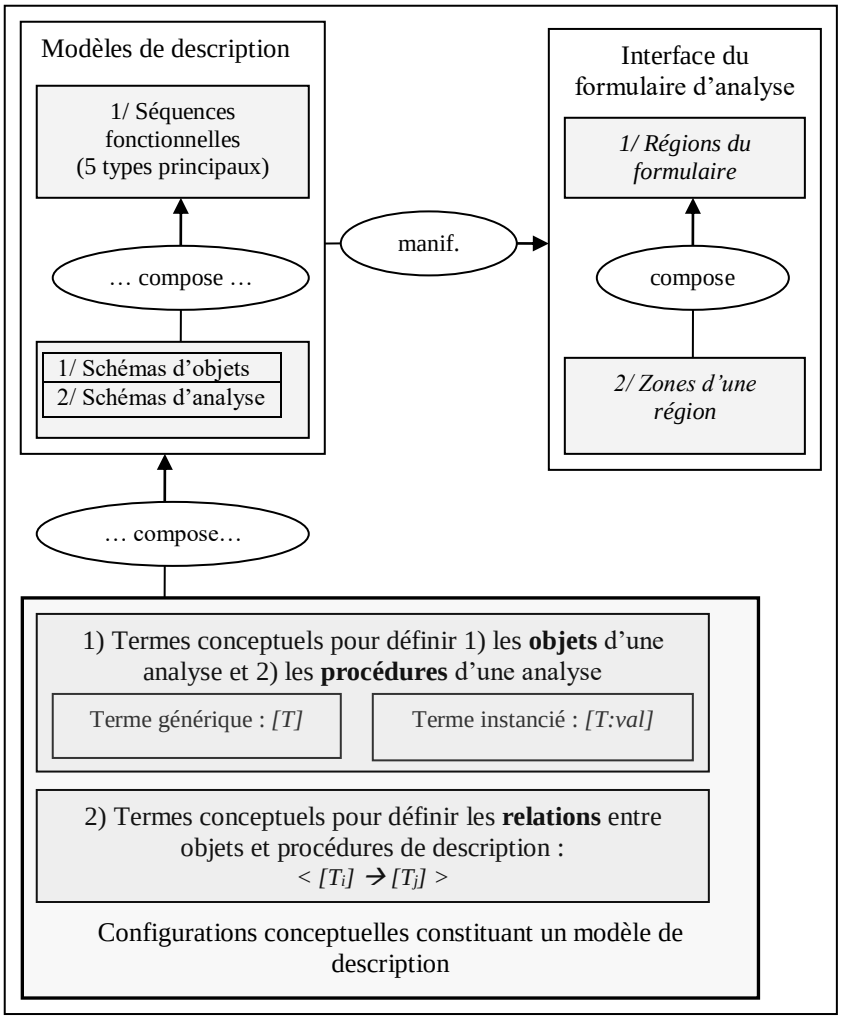

Figure 11.2. *Vue synthétique des rapports entre d'une part interface de travail et modèle de description et d'autre part modèle de description et constituants métalinguistiques d'un modèle de description*

Une deuxième catégorie de ressources métalinguistiques externes au système ASA est composée de *standards*, *ontologies*, *thesaurus* et autres *terminologies* avec lesquels ont été ou peuvent être créés des correspondances, des *ponts*. Ces correspondances ou ponts servent à rendre – dans la mesure du possible – *interopérables* les résultats des analyses concrets réalisés à l'aide du Studio ASA

avec ceux réalisés en référence à d'autres ressources métalinguistiques (ontologie, thesaurus, etc.).

Considérons maintenant de plus près la figure 11.2. Celle-ci nous montre une vue d'ensemble des principaux rapports qu'existent entre les différents éléments composant le système métalinguistique ASA. On y voit notamment :

– le rapport entre l'*interface de travail de l'analyste* (ici il s'agit de l'Atelier de Description du Studio ASA) et les *modèles* qui *sous-tendent cette interface*, qui la définissent (dans notre cas, il s'agit des modèles discutés dans ce livre, à savoir, ceux réservés à la description des sujets développés dans un texte ou un corpus de textes audiovisuels). Un *modèle de description** est visualisé sous forme de *régions* et de *zones* spécialisées qui forment l'organisation physique et formelle de l'interface (pour une discussion approfondie de l'analyse et de la *conception d'interfaces numériques*, voir [STO 05]) ;

– le rapport entre un modèle de description (ici : du contenu d'un texte ou d'un corpus de textes audiovisuels) et ses constituants que sont d'abord les *séquences* fonctionnellement spécialisées* (description du domaine de connaissances, description de la mise en discours, etc.) et ensuite les *schémas** de définition (c'est-à-dire d'une part les *schémas définissant l'objet d'analyse* et d'autre part les schémas définissant la *procédure d'analyse*). Les schémas de description configurent, définissent les séquences de description d'un modèle, un schéma pouvant investir soit une seule soit plusieurs séquences, soit en principe toutes les séquences (c'est le cas, par exemple, des schémas définissant une procédure de description libre ou contrôlée) ;

– le rapport, enfin entre les parties composant un modèle de description, c'est-à-dire les *schémas* et *séquences*, et leurs constituants que sont les *termes conceptuels** génériques ou déjà référencés (dans le *thesaurus**) ainsi que les relations permettant de positionner les termes conceptuels pertinents sous forme justement de schémas de description ou de parties de schémas de description. En ce qui concerne les termes conceptuels, nous distinguons entre deux grandes classes : celle qui représente le *discours-objet* ASA*, c'est-à-dire les *objets de l'univers du discours ASA pouvant être soumis à une analyse* (voir chapitres 12 et 13) et celle qui représente les *activités d'analyse* de l'objet du discours (voir chapitre 14).

11.3. Le méta-lexique des termes conceptuels ASA

Le « cœur » des ressources métalinguistiques du système métalinguistique ASA est constitué par un lexique hiérarchiquement organisé de *concepts* ou, plutôt de *termes conceptuels*, c'est-à-dire d'*expressions linguistiques* de concepts. Ensemble, ces termes conceptuels constituent le vocabulaire de l'*univers du discours ASA*. Autrement dit, les termes conceptuels sont les « mots » ou les « expressions » introduites et définies par le *cogniticien* (« concept designer », ingénieur de connaissances) et utilisées par l'analyste pour pouvoir parler et communiquer sur un domaine de connaissances documenté par des corpus de textes (audiovisuels, etc.). *Métalinguistique*, employé dans ce contexte, veut dire *langage* ou *langue* construite en vue de *traiter* (décrire, indexer, etc.) un objet textuel donné, c'est-à-dire dans notre cas, en vue d'*analyser* des corpus audiovisuels.

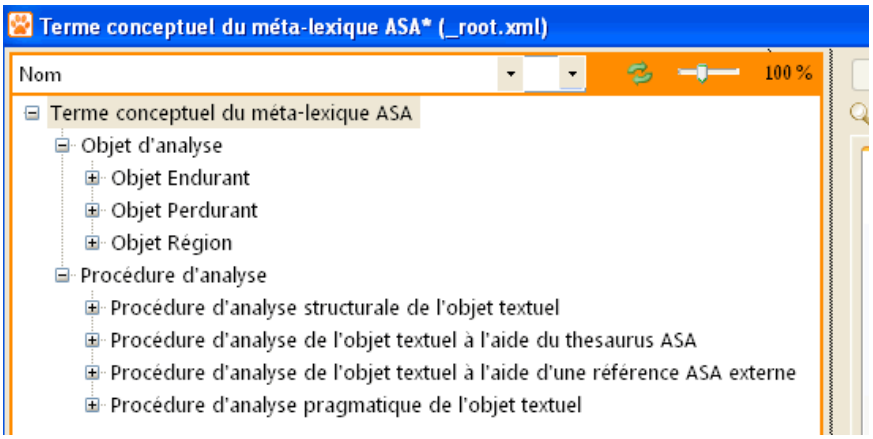

Figure 11.3. *Les deux parties composant le méta-lexique des termes conceptuels ASA – le vocabulaire conceptuel relatif à l'objet d'analyse et le vocabulaire conceptuel relatif à la procédure d'analyse*

Souvenons-nous : tout *modèle de description** ASA se réduit, *in fine* à deux types de configurations conceptuelles qui sont d'une part les *schémas des objets d'analyse* et d'autre part les *schémas des activités d'analyse* ou *de description*. Par conséquent, le méta-lexique des termes conceptuels ASA se constitue de deux ensembles complémentaires de termes conceptuels (voir figure 11.3) :

1. le premier ensemble du méta-lexique : le vocabulaire des termes conceptuels dont le terme racine est *[Objet d'analyse]* ;
2. le deuxième ensemble du méta-lexique : le vocabulaire des termes conceptuels dont le terme racine est *[Procédure d'analyse]*.

Dans les chapitres précédents, nous avons distingué entre plusieurs grandes catégories d'objets pour analyser un texte audiovisuel ou un corpus de textes audiovisuels :

– les objets qui constituent le domaine de référence (c'est-à-dire le domaine dont « parle » le texte audiovisuel soumis à l'analyse) ;
– les objets qui contextualisent les premiers dans le temps et l'espace ;
– les objets proprement discursifs (qui permettent donc à comprendre *comment* le texte soumis à l'analyse *parle* de son objet) ;
– les objets visuels et acoustiques (permettant de comprendre l'expression, la mise en scène audiovisuelle d'un objet dans le texte soumis à l'analyse) ;
– enfin, les objets dits réflexifs qui servent à l'explicitation du contenu et des objectifs d'une analyse elle-même.

Ces différentes catégories d'objets d'analyse interviennent plus particulièrement, souvenons-nous, dans la définition et le développement des modèles de description du contenu de corpus audiovisuels. Mais le Studio ASA (c'est-à-dire son Atelier de Description) est pourvu encore d'autres familles de modèles servant à la description paratextuelle d'un texte ou d'un corpus de textes audiovisuels, à la traduction et à l'adaptation de ce dernier à un public donné, etc. Ces modèles font appel à des catégories d'objets d'analyse que nous n'utilisons pas pour la description du contenu. Il s'agit, par exemple, des catégories d'objets d'analyse qui servent à l'identification d'un texte audiovisuel ou d'un corpus de textes audiovisuels, à l'explicitation de leur spécificité cognitive ou intellectuelle, ou encore à l'explicitation de leurs usages.

Toutes ces catégories d'objets d'analyse (utilisées ou pas utilisées dans la construction des modèles destinés plus spécifiquement à l'analyse du contenu d'un texte ou d'un corpus de textes audiovisuels) forment le domaine de référence du vocabulaire de la première partie du méta-lexique des termes conceptuels ASA, c'est-à-dire du vocabulaire dont le terme conceptuel de base est *[Objet d'analyse]* du *discours-objet** ASA.

La figure 11.3 montre que ce vocabulaire des termes conceptuels se décompose en trois branches dont les termes conceptuels de base s'intitulent : [Objet Endurant], [Objet Perdurant] et [Objet Région]. Ensemble, ils forment la base canonique du vocabulaire du méta-lexique ASA relatif aux domaines de connaissances thématisés (voire thématisables) dans l'univers du discours ASA. Nous en proposerons des explications et une présentation détaillée dans les chapitres 12 et 13.

L'interface visualisant le méta-lexique des termes conceptuels ASA dans la figure 11.3 est celle d'un outil de conception et de développement de ressources métalinguistiques s'intitulant OntoEditeur[51] (voir plus loin, section 11.8).

Dans les chapitres précédents, nous avons discuté *in extenso* toute une série d'activités d'analyse ou de description spécifiques qui composent les procédures de description libre et/ou contrôlée d'un objet ou d'un ensemble d'objets d'analyse. Ces activités ainsi que toutes celles dont nous avons besoin pour exécuter les autres tâches d'analyse ou de traduction-adaptation, forment le domaine de référence du vocabulaire de la deuxième partie du méta-lexique des termes conceptuels ASA, c'est-à-dire du vocabulaire dont le terme conceptuel de base est *[Procédure d'analyse]*.

Comme le montre la figure 11.3, le vocabulaire des termes conceptuels désignant les procédures d'analyse, se décompose à son tour en quatre branches plus spécialisées dont les termes de base s'intitulent comme suit :

– [Procédure d'analyse structurale de l'objet textuel] ;
– [Procédure d'analyse de l'objet textuel à l'aide du thesaurus ASA] ;
– [Procédure d'analyse de l'objet textuel à l'aide d'une référence ASA externe] ;
– [Procédure d'analyse pragmatique de l'objet textuel].

Ensemble, ces quatre branches forment la base canonique des activités d'analyse de l'univers du discours ASA. Nous les discuterons plus en détail dans le chapitre 14.

Les deux vocabulaires [Objet d'analyse] et [Procédure d'analyse] forment donc ensemble le méta-lexique des termes conceptuels de l'*univers du discours ASA*. *Univers de discours ASA* veut bien dire qu'il forme un espace de connaissances qui possède sa structure, ses règles, sa « grammaire ». Cet univers n'est pas le seul, bien évidemment – il en existe autant qu'on veut. Un but important est de rendre l'univers de discours ASA suffisamment général et ouvert pour qu'il puisse accueillir les univers de discours d'autres archives que celles examinées ici.

11.4. Le thesaurus ASA

Le thesaurus ASA est un *vocabulaire contrôlé* de termes et d'expressions linguistiques normalisées qui forment les valeurs (instances ou référents) prédéfinies des termes conceptuels génériques du méta-lexique ASA. Comme nous le savons

[51] OntoEditeur est un éditeur xml qui a été développé par Francis Lemaitre de l'ESCoM dans le cadre des projets de recherche et de développement SAPHIR et ASA-SHS, financés par l'Agence Nationale de la Recherche (ANR). Il constitue, actuellement, la partie logicielle du Studio de Modélisation ASA.

déjà, le thesaurus ASA se compose d'un ensemble de *facettes*, et de listes (hiérarchiques) d'*expressions normalisées* (de *descripteurs*) dont chacune fait partie d'une ou de plusieurs facettes. Une facette, elle, forme un *axe sémantique* ou encore une *dimension du contenu* (du signifié au sens structuraliste du terme) d'un terme conceptuel.

Une expression normalisée d'une facette représente donc une *valeur possible*, une instance ou un référent du terme conceptuel qui possède comme signifié la même facette. Par exemple, une des facettes du terme conceptuel [Pays] est *Tous les pays du monde du début du XXIe siècle*. Cette facette est composée de la liste de tous les pays reconnus par la communauté internationale (et non pas seulement par l'ONU). Ainsi le micro-thesaurus ASA (c'est-à-dire *facette + liste (hiérarchique) des expressions normalisées*) *Tous les pays du monde du XXIe siècle* contient, outre les 194 pays officiellement reconnus par l'ONU, également des territoires tels que l'île de Taïwan, les îles Cook, l'Abkhazie, la Palestine, etc. (figure 11.4).

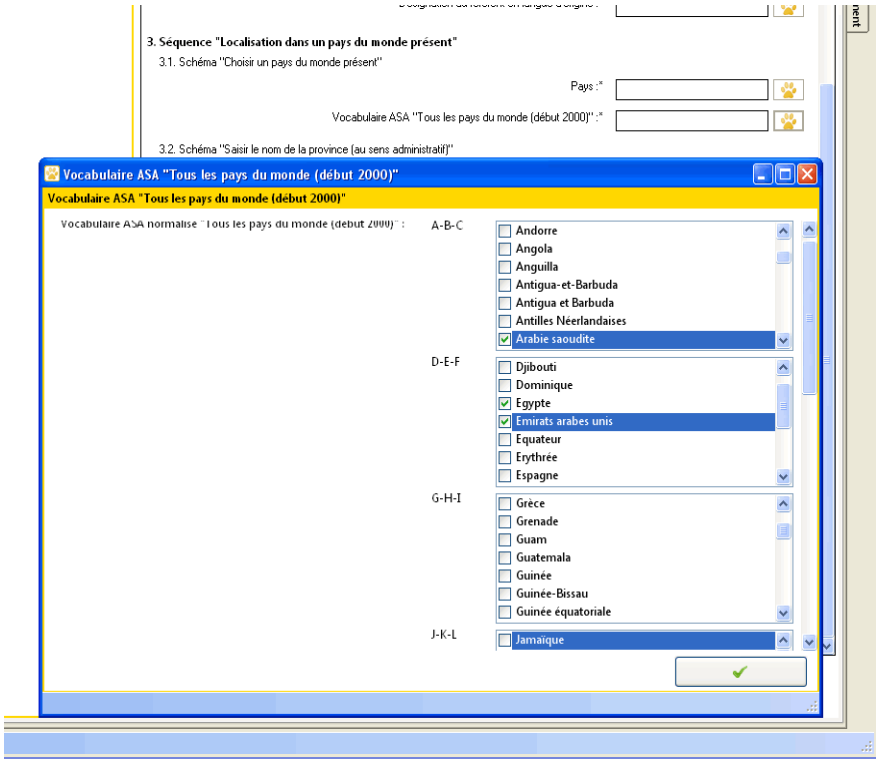

Figure 11.4. *Les expressions normalisées désignant les pays du monde et composant la facette « Tous les pays du monde (début 2000) »*

Si l'inclusion des pays non officiellement reconnus par l'ONU pose un problème à telle ou telle communauté d'analyste, celle-ci peut être évidemment remplacée par une liste des 194 pays reconnus par l'ONU réunis, cette fois-ci, sous la facette Les *194 pays reconnus officiellement par l'ONU*.

Pour une description concrète, on peut utiliser ainsi l'une ou l'autre des deux facettes ou encore la réunion des deux (la réunion des deux facettes veut dire, dans ce cas concret, un micro-thesaurus qui est formellement identique à celui s'organisant sur la base de la facette *Tous les pays du monde du XXIe siècle* mais où les territoires pas reconnus officiellement par l'ONU sont marqués comme tels).

Bien sûr, une expression normalisée représentant une valeur possible d'un terme conceptuel du méta-lexique ASA peut, elle, faire partie de différentes facettes. Cela signifie qu'une expression normalisée peut avoir des sens différents selon son utilisation dans tel ou tel micro-thesaurus et pour telle ou telle analyse concrète.

Ainsi la figure 11.5 nous montre que l'expression normalisée *Argentine* peut être employée, dans le cadre du thesaurus ASA, pour trois facettes (légèrement) différentes : la facette *Pays contemporains d'Amérique* ; la facette *Pays de culture littéraire* ; la facette *Tous les pays du monde (2000)*. On peut, bien sûr, s'imaginer toute une diversité d'autres facettes dans lesquelles l'expression normalisée *Argentine* peut être employée. Les trois facettes identifiées ci-dessus ainsi que dans la figure 11.5 sont pertinentes pour l'analyse de corpus des ateliers d'expérimentation du projet ASA-SHS. Ce pour cette raison qu'elles ont été créées.

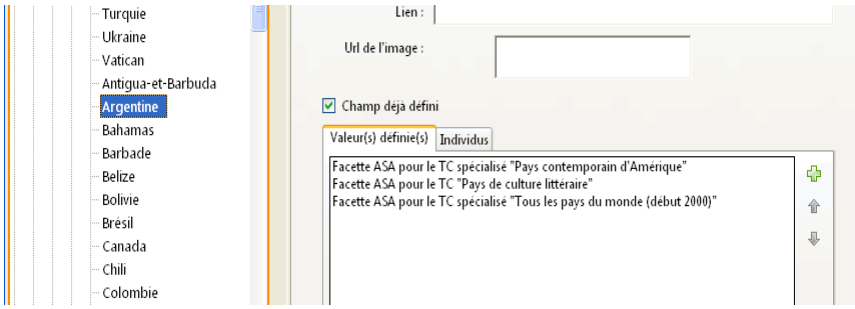

Figure 11.5. *Emploi de l'expression normalisée Argentine dans trois facettes différentes faisant partie du thesaurus ASA*

La figure 11.6 nous présente d'une manière synthétique et globale les principales parties du thesaurus de l'univers du discours ASA. Nous distinguons entre :

1/ le *thesaurus commun* du domaine d'analyse ASA. Le thesaurus commun est mis à la disposition à chaque utilisateur (analyste), à chaque groupe d'utilisateurs

(d'analystes) mais ne peut être modifié (enrichi) que par le comité (l'instance) responsable de la gestion du thesaurus ASA ;

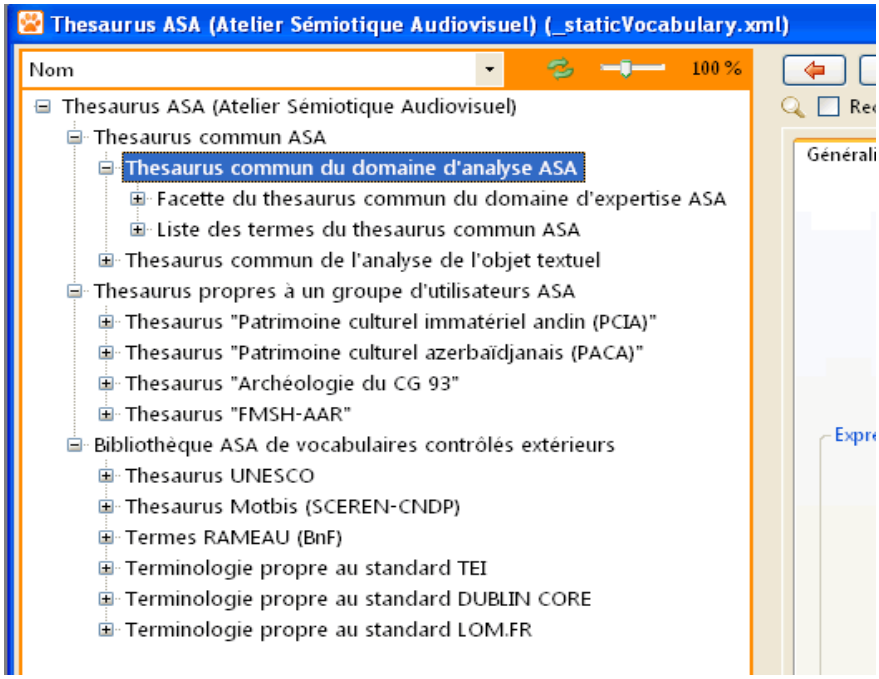

Figure 11.6. *Organisation générale du thesaurus ASA*

2/ les *thesaurus propres* à un groupe ou une communauté d'utilisateurs des ressources ASA. Ce sont des thesaurus représentant les points de vue spécifiques relatives à l'univers du discours d'une archive. Ainsi, par exemple, les archives audiovisuelles PCIA (*Patrimoine culturel immatériel andin*) possèdent leurs propres facettes et expressions pour analyser les textes audiovisuels faisant partie de leurs fonds. En même temps, on peut copier des facettes et listes d'expressions normalisées provenant du thesaurus commun dans cette partie qui es propre à une communauté d'analystes afin de pouvoir modifier librement soit les facettes soit les listes d'expressions normalisées ;

3/ les vocabulaires contrôlés extérieurs au système ASA provenant notamment d'autres thesaurus, langages ou standards qu'on souhaite utiliser « directement » (seuls ou ensemble avec des ressources terminologiques propres au système ASA) lors de l'analyse d'un texte ou d'un corpus de textes audiovisuels. Nous en avons vu un exemple au chapitre 10 (section 10.3) qui s'appuie sur l'usage d'un micro-thesaurus ASA commun et d'un micro-thesaurus de l'UNESCO pour décrire des

textes audiovisuels sur les formations culturelles d'un peuple ou d'une région géopolitique.

11.5. Les schémas de définition

Un schéma de définition est une configuration qui positionne deux ou plusieurs éléments métalinguistiques du système ASA les uns par rapport aux autres. En plus, une configuration de type « schéma » est considéré comme une configuration élémentaire (contrairement à une configuration de type « séquence » qui est elle-même une composition d'un ou de plusieurs schémas).

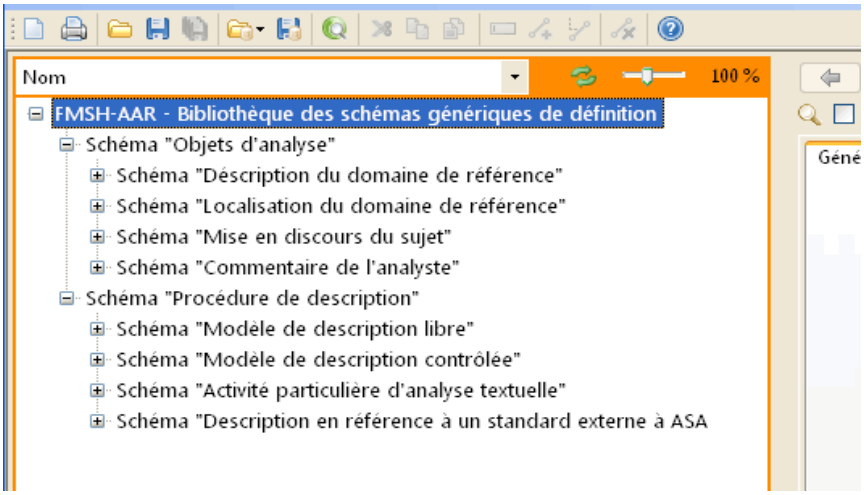

Figure 11.7. *Bibliothèque de schémas générique de définition définissant l'univers du discours des archives FMSH-AAR*

La figure 11.7 nous montre une bibliothèque de schémas génériques de définition que nous utilisons pour traiter des corpus audiovisuels qui font partie du fonds du Programme AAR (Archives Audiovisuelles de la Recherche). Comme on peut le constater, nous y distinguons, conformément à notre approche théorique, entre, d'une part, la classe des schémas définissant les objets d'analyse et, d'autre part, la classe des schémas définissant les procédures de description utilisées pour décrire un objet particulier.

Les schémas génériques de définition, comme déjà dit, forment des *mini-structures*, des *structures locales* dont se composent les *configurations thématiques** (topiques, de mise en discours, etc.) qui définissent les modèles de description

accessibles via l'interface de travail du Studio ASA (section 11.2). Les figures 11.8 et 11.9 montrent à quoi ressemblent ces mini-structures ou ces structures locales.

La figure 11.8 focalise sur le schéma générique de l'objet d'analyse *Description d'un courant de pensée, d'une théorie …* – schéma générique qui fait partie d'une famille de schémas utilisée pour définir des sujets relatifs au domaine de la culture scientifique (lato sensu). Ce n'est qu'au niveau de la bibliothèque des séquences que se décide lequel ou lesquels des schémas faisant partie de cette famille sont réellement pertinent(s) pour être sélectionnés dans une séquence particulière spécialisée dans l'analyse d'un sujet relevant de la culture scientifique : dans les cas les plus simples, sélection uniquement d'*un seul schéma* ou, au contraire, dans les cas plus complexes, sélection de *plusieurs* schémas.

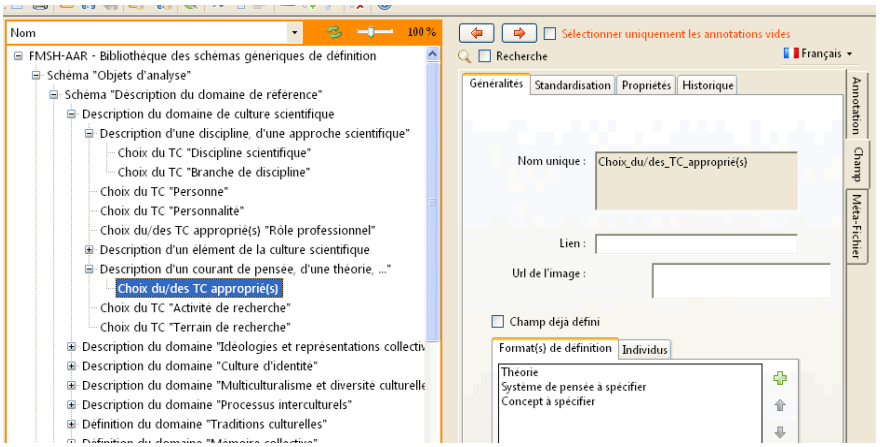

Figure 11.8. *Définition du schéma générique de l'objet d'analyse* Description d'un courant de pensée, d'une théorie, etc.

Mais ce que nous montre surtout la figure 11.8, c'est le fait que le schéma *Description d'un courant de pensée, d'une théorie…* contient une variante intitulée *Choix du/des TC approprié(s)* – variante qui se définit par la sélection des trois termes conceptuels : [Théorie], [Système de pensée à spécifier], [Concept à spécifier]. Notre schéma (ou, pour être précis, la variante de notre schéma) se définit donc sous forme d'une petite structure composée de trois termes conceptuels (faisant partie du méta-lexique ASA) qui se positionnent les uns par rapport aux autres sous forme d'une *relation* appelée *« ou non-exclusive »* ou encore *« disjonction inclusive »* (c'est-à-dire *vel*, en latin) : soit l'un, soit l'autre, soit le troisième, soit les deux premiers, etc.

Même si nous n'avons pas eu le temps ni les moyens d'implémenter au niveau informatique les différentes relations logiques, notons que la structure qui définit un schéma est faite :

– d'une part par la *sélection* d'un ou de plusieurs termes conceptuels,
– et d'autre part par une *relation logique* définissant le rapport précis entre les termes conceptuels sélectionnés.

En principe et dans la très grande majorité des cas que nous avons rencontré jusqu'à maintenant, il s'agit soit d'une relation de *simple affirmation* de la présence d'un terme conceptuel dans un schéma soit de la relation appelée *« ou non-exclusive »* ou encore *« disjonction inclusive »* (*vel*, en latin).

Ajoutons que toutes les autres relations dont on fait état dans la littérature spécialisée – relations *casuelles*, relations *attributives*, relations *partitives*, relations *de localisation*, relations *rhétoriques*, etc. – n'interviennent plutôt qu'au niveau des *séquences* lors qu'il s'agit de sélectionner et de positionner non pas des termes conceptuels (des « concepts, si l'on veut) mais des *schémas génériques de description* composés d'un ou de plusieurs termes conceptuels.

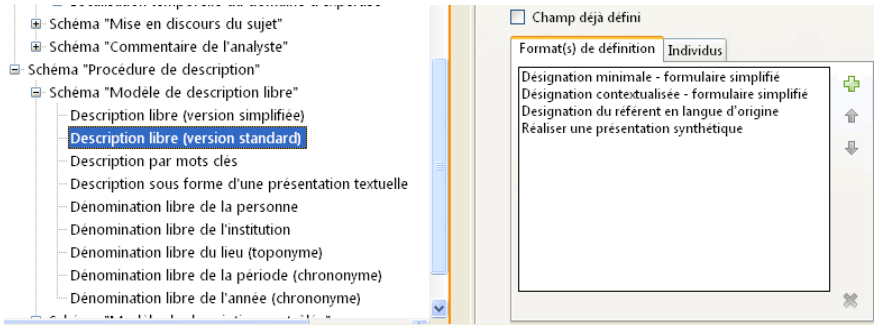

Figure 11.9. *Définition du schéma générique de la procédure de description* Description libre (version standard)

La figure 11.9 montre un schéma générique qui fait partie de la deuxième classe des schémas identifiée ci-dessus, à savoir de celle de la classe des schémas réservés à la définition des procédures de description. Dans notre cas, il s'agit de la procédure de description libre, version standard.

Contrairement à la version simplifiée qui n'a recours qu'à une seule activité de description (c'est-à-dire l'activité [Désignation minimale – formulaire simplifié]), le schéma définissant la version standard de la procédure de description libre a recours, elle, comme le montre la figure 11.9, à quatre activités d'analyse qui font toutes, bien sûr, partie du vocabulaire conceptuel du méta-lexique ASA réservé à l'analyse du texte audiovisuel. Il s'agit des activités [Désignation minimale – formulaire

simplifié], [Désignation contextualisée – formulaire simplifié], [Désignation du référent en langue d'origine] et [Réaliser une présentation synthétique].

D'une manière identique au schéma représenté par la figure 11.8, aussi celui-ci est donc défini d'une part par une sélection d'un ensemble de termes conceptuels pertinents et d'autre part par la relation appelée *« ou non-exclusive »* ou encore *« disjonction inclusive »*. Cela veut dire, que l'analyste, utilisant ce schéma peut performer soit l'une soit l'autre soit plusieurs des activités identifiées par le schéma en question pour renseigner son objet d'analyse. Cependant, il existe ici une *condition supplémentaire* qui doit être respectée par l'analyste. Cette condition supplémentaire veut que si ce schéma est utilisé par l'analyste, alors l'activité de description appelée [Désignation minimale – formulaire simplifié] devient une activité à la fois *obligatoire* et *présupposée* par toutes les autres activités : en utilisant ce schéma, l'analyste doit réaliser l'activité [Désignation minimale – formulaire simplifié] et, cela, *avant* toutes les autres activités.

Cette condition que nous n'avons pas pu, non plus, implémenter au niveau informatique, ajoute une *contrainte pragmatique spécifique* à l'usage de ce schéma qui n'est pas contenue dans la relation logique régissant les rapports entre les quatre termes conceptuels sélectionnés.

11.6. Les séquences de description

Une séquence possède une fonction spécifique et propre à elle dans un *modèle de description** d'un objet audiovisuel. En nous limitant aux modèles de description thématique (c'est-à-dire du contenu audiovisuel), nous distinguons, comme déjà expliqué (voir chapitre 5), les principaux types fonctionnels suivants de séquences de description :

– séquences d'identification et description du domaine thématisé dans un objet audiovisuel ;
– séquences de localisation temporelle ou spatiale du domaine thématisé ;
– séquences de description de la mise en discours de l'objet thématisé (et éventuellement contextualisé) ;
– séquences de description de l'expression visuelle ou audiovisuelle de l'objet thématisé ;
– séquences réservées au commentaire de l'analyste à propos soit de la description de l'objet thématisé soit de l'objet thématisé lui-même (et/ou de ses conditions de mise en discours et d'expression audiovisuelle).

Toutes les *bibliothèques des séquences** se définissent en référence à ces cinq types fonctionnels. La figure 11.10 nous montre la bibliothèque des séquences de description à l'aide desquelles l'analyste procède à la description et indexation des corpus audiovisuels composant le fonds du Programme AAR (Archives

Audiovisuelles de la Recherche). On y distingue cinq grandes familles de séquences qui sont les suivantes :

1) Première famille de séquences de description regroupant les séquences réservées à l'analyse du domaine d'expertise propre à l'univers du discours d'une archive. En ce qui concerne les archives AAR, il s'agit de domaines tels que ceux de la *globalisation/mondialisation*, des *mouvements sociaux*, de la *diversité culturelle*, etc. Tous ces domaines (et bien d'autres encore) sont abordés dans la production audiovisuelle des archives AAR, possèdent un intérêt évident pour le monde de la recherche en sciences humaines et sociales et contribuent à définir la spécificité, la valeur ajoutée de ces archives.

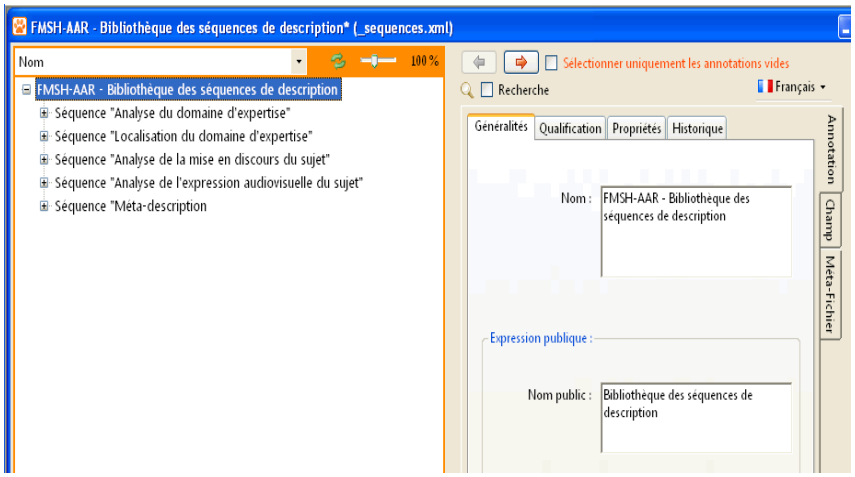

Figure 11.10. *Bibliothèque des séquences définissant l'univers du discours des archives FMSH-AAR*

2) Deuxième famille de séquences de description réunissant les séquences qui sont réservées à la localisation (géographique, géopolitique, chronologique, historique, etc.) des domaines d'expertise. Précisons que cette famille de séquences n'est pas propre (contrairement à la première famille de séquences !) à l'univers du discours des archives AAR. On la retrouve pratiquement dans chaque bibliothèque de séquences servant à définir les modèles de description de l'univers du discours d'une archive particulière.

3) Troisième famille de séquences de description regroupant les séquences servant à l'analyse de la mise en discours d'un domaine d'expertise tel que celui des mouvements sociaux ou de la globalisation. Aussi cette famille de séquences se retrouve, en principe, dans chaque bibliothèque de séquences définissant les modèles de description du fonds audiovisuel d'une archive.

4) Quatrième famille de séquences de description regroupant toutes les séquences qui servent à l'analyse de l'expression audiovisuelle et/ou verbale d'un domaine d'expertise thématisé dans un texte audiovisuel. De nouveau, ces séquences ne sont pas propres à la bibliothèque des modèles de description des archives AAR mais peuvent être réutilisées telles quelles ou moyennant quelques modifications locales pour la définition de n'importe quelle bibliothèque de modèle de description.

5) Cinquième famille de séquences de description regroupant les séquences servant à expliciter le point de vue de l'analyste soit sur le domaine thématisé soit sur son analyse, sur sa description du domaine thématisé. Cette famille de séquences, de nouveau, peut être réutilisée pour la définition de toute bibliothèque de modèles de description.

Grosso modo, toutes les bibliothèques de séquences d'analyse se construisent donc selon le modèle de référence représenté par la figure 11.10.

En ce qui concerne plus particulièrement la catégorie des modèles de description thématique (c'est-à-dire du contenu audiovisuel), c'est notamment la première famille de séquences, c'est-à-dire celle qui sert à la description des objets ou domaines de connaissance, qui différencie les bibliothèques des séquences les unes les autres, les quatre autres familles n'étant pas réellement propres à une archive, à l'univers du discours d'une archive spécifique. Ainsi, malgré la complexité apparente du processus de l'élaboration d'une bibliothèque de modèles de description thématique (c'est-à-dire du contenu audiovisuel), celle-ci ne concerne que la première famille des séquences servant à l'analyse des objets et domaines de connaissance thématisé dans un corpus ou un fonds audiovisuel.

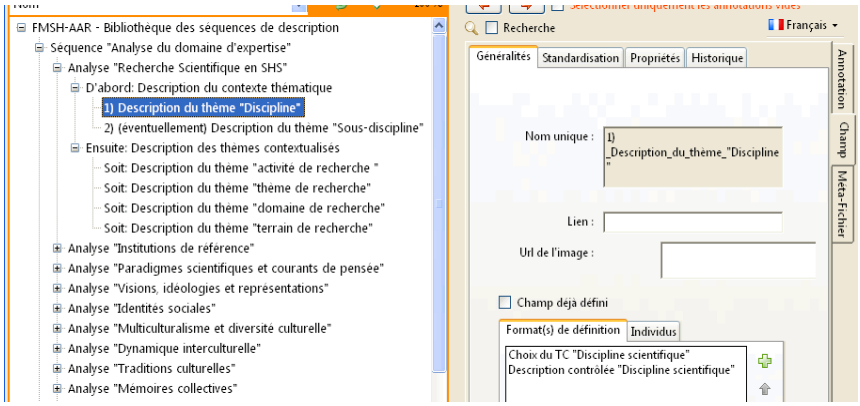

Figure 11.11. *Organisation fonctionnelle d'une séquence et rapports avec schémas composant une séquence*

Voyons encore l'*organisation interne* d'une séquence et ses *rapports* avec les schémas de description. L'exemple représenté dans la figure 11.11, nous montre la séquence *Analyse « Recherche Scientifique en SHS »*. Cette séquence se déploie en une structure syntagmatique qui est composée, comme nous pouvons le voir, en deux grandes sous-séquences dont la deuxième présuppose la première. Autrement dit, la première sous-séquence doit être obligatoirement renseignée avant la seconde. En regardant de plus près encore le rapport entre les deux sous-séquences, nous constatons que la première sous-séquence joue le rôle de la délimitation du contexte pertinent (dans notre cas, le contexte est donné par la ou les disciplines scientifiques dont est question dans un texte ou un corpus de textes audiovisuels), tandis que la deuxième sous-séquence a en charge de détailler ce contexte (dans notre cas, il s'agit d'expliciter des aspects spécifiques d'une recherche disciplinaire ou interdisciplinaire : thème, domaine, terrain, etc.).

Le dispositif *« d'abord le contexte pertinent, ensuite différents facettes ou aspects du contexte »* est un dispositif syntagmatique très récurrent dans la construction de séquences plus complexes se déployant, comme celle représentée par la figure 11.11, en deux ou plusieurs sous-séquences. Mais il existe, bien entendu, toute une diversité de tels dispositifs fonctionnels et syntagmatiques tels que le *déploiement énumératif* (certainement le plus simple), le *déploiement causal*, le *déploiement chronologique*, le *déploiement à proprement parler narratif*, etc. Nous retrouvons ici la problématique de la cohérence syntagmatique du texte – problématique que nous avons traitée d'une manière plus extensive dans un autre livre consacré à l'analyse et la conception de « nouveaux produits d'information » ([STO 99, STO 92]).

Comme le montre la figure 11.11, chacune des deux sous-séquences peut, à son tour, se diviser – selon le même impératif de cohérence syntagmatique – en des séquences encore plus spécialisées. Dans notre exemple, la première sous-séquence est elle-même composée de deux séquences plus spécialisées dont la première doit être renseignée obligatoirement, tandis que le renseignement de la seconde est facultatif. Simplement dit, l'analyste est ici invité à préciser de quelle(s) discipline(s) scientifique(s) il est question dans le texte audiovisuel qu'il décrit. Si pertinent et si l'analyste le souhaite, il peut spécifier encore la ou les disciplines. C'est une option particulièrement utile pour bien expliciter des disciplines (telles que la sociologie ou l'anthropologie) très prolifiques en sous-disciplines et approches dont on ne peut pas réellement savoir si leur existence relève de la décision du chercheur individuel, de tel ou tel groupe de recherche ou si elles peuvent être considérées comme des faits institutionnels, voire des références scientifiques communément partagées. Les séquences spécialisées qui composent la deuxième sous-séquence suivent une logique de déploiement énumératif : rien n'empêche d'ajouter d'autres séquences spécialisées pertinentes pour expliquer le contexte de la recherche scientifique et rien n'empêche de changer la place de telle ou telle séquence spécialisée dans l'ordre syntagmatique choisi.

En tout cas, chaque unité de la séquence qui sert « directement » à la description d'un texte audiovisuel est, comme le montre également la figure 11.11, définit par au moins deux types de schémas de description. L'exemple illustré par la figure 11.11 est l'unité de la séquence *Description du thème Discipline*. Celle-ci est définie, comme on peut la voir, par le schéma *Description du TC (= terme conceptuel) Discipline* ainsi que par le schéma *Description contrôlée Discipline scientifique*.

En règle générale, toutes les unités de séquences servant « directement » à la description d'un texte audiovisuel se construisent sur ce couple de schémas de description que nous venons de discuter :

– le premier schéma sert à définir la structure topique (dans notre cas – le plus simple – il s'agit simplement de confirmer le fait que l'analyse porte bien sur le terme conceptuel [Discipline scientifique] ; pour des cas plus complexes, voir les chapitres 5 et 8),

– le deuxième schéma sert à décrire la structure topique préalablement définie (comme nous le savons déjà, il s'agit soit du schéma définissant une procédure de description contrôlée soit du schéma définissant une procédure de description libre ; dans quelques cas, on peut également trouver ici un schéma quelque peu plus élaboré intégrant les deux procédures).

11.7. Les ressources externes au système ASA

Par « ressources externes » au système métalinguistique ASA, nous entendons les ressources qui ne sont pas des constituants à proprement de ce système ou bien qui ne le sont pas au même titre que, par exemple, le méta-lexique des termes conceptuels, le thesaurus ASA ou les schémas d'indexation. Cependant, « externe au système ASA » ne veut pas dire, ici, « sans relation » (existante ou possible) avec le dit système.

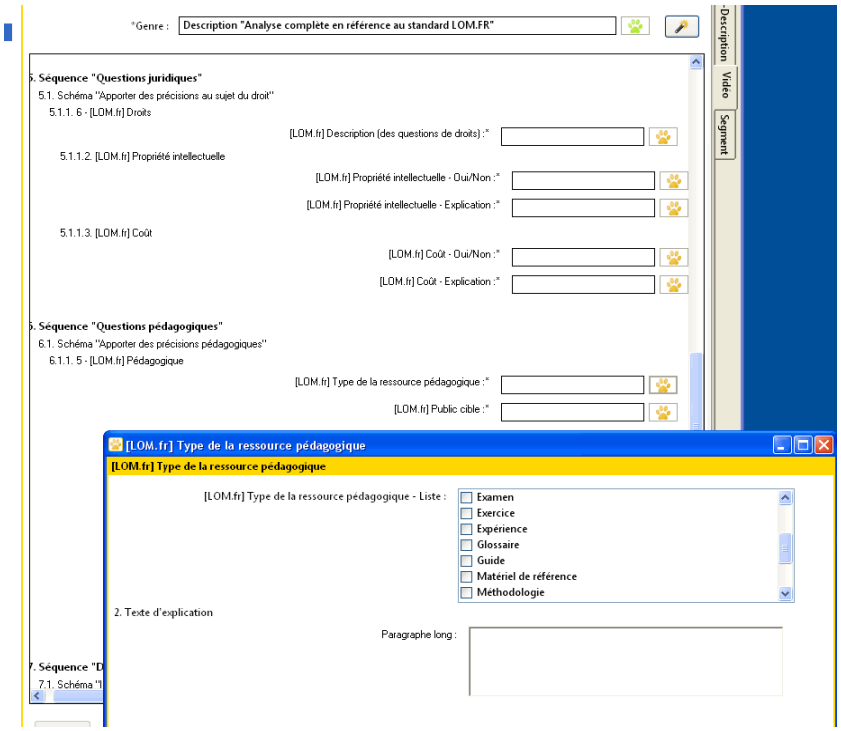

Figure 11.12. *Le standard LOMFR intégré dans l'interface de travail du Studio ASA*

Nous distinguons entre deux classes de ressources externes au système ASA : la classe des ressources formée des données générées par l'analyste, et la classe des ressources formée de la diversité des métadonnées déjà existantes. Ces deux classes sont soit contrôlées par le système ASA (c'est le cas des données générées par l'analyste), soit – dans la mesure du possible – mis en relation avec les éléments du système ASA afin d'augmenter le pouvoir d'expressivité du système ASA lui-même et de contribuer à ce que l'on appelle l'interopérabilité des systèmes de description-indexation utilisés pour générer des métadonnées relatives à une ressource numérique.

Les standards, terminologies, ontologies, etc. sont des ressources métalinguistiques externes au système ASA mais dont, potentiellement, il faudrait se servir dans la mesure où ils constituent les « langages » utilisés par diverses communautés (institutions, etc..) pour traiter (décrire et/ou indexer) des corpus numériques audiovisuels documentant des domaines de connaissance comparables à ceux que l'on souhaite investiguer à l'aide de l'atelier sémiotique de la description audiovisuelle.

Il existe des types très variés de ressources externes. Nous pensons, bien sûr, aux thesaurus et terminologies monolingues ou multilingues, aux ontologies conceptuellement souvent très proches des terminologies, aux standards utilisés pour archivées, diffusées et partagées des ressources numériques ou encore aux réseaux sémantiques et conceptuels comparables aux *modèles de description** ASA.

C'est un véritable « paysage multilingue » qui se dessine ainsi devant nous transposant l'image de la tour de Babel du niveau des langues naturelles au niveau du métalangage. Il est impossible de demander à l'analyste de faire ce travail de « traduction » de son analyse (réalisée à l'aide de modèles de description ASA) vers d'autres métalangages. Ces « traductions » devraient être plutôt intégrées dans le système lui-même.

Il existe aux moins deux options pour rendre (au moins partiellement) interopérable les données générées par un analyste à l'aide du métalangage de description ASA :

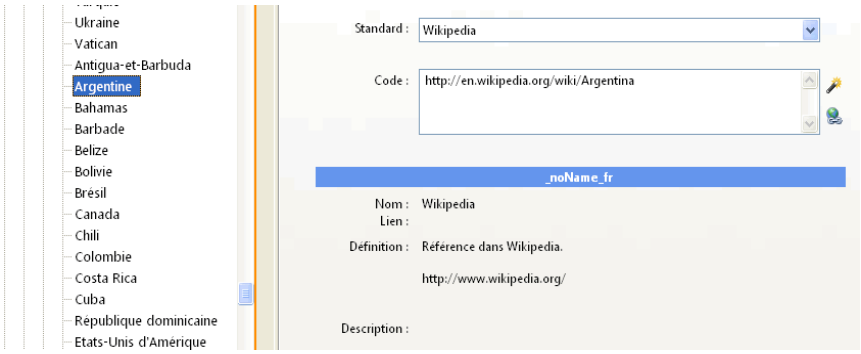

Figure 11.13. *Création d'un pont entre l'expression normalisée ASA « Argentine » et l'article Wikipedia (en anglais) pour pouvoir récupérer entre autre des données de localisation géographique*

1/ développement de *modèles de description** ASA qui intègrent telles quelles des références externes au système métalinguistique ASA. La figure 11.12 nous fournit un exemple sous forme d'un extrait d'un modèle de description ASA élaboré et réalisé conformément à la norme française LOMFR ;

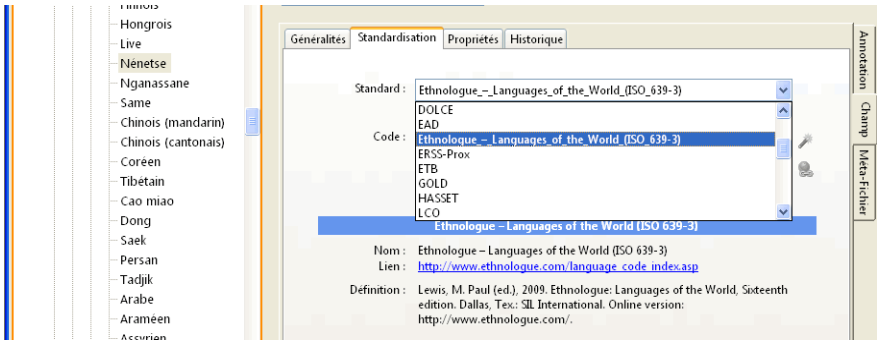

Figure 11.14. *Création d'un pont entre l'expression normalisée ASA « Nénètse » (langue samoyèdes) et son correspondant dans Ethnologue « Languages of the World »*

2/ mise en relation des termes conceptuels, des expressions normalisées faisant partie du thesaurus ASA, des schémas, séquences ou modèles de description avec leurs équivalents métalinguistiques dans un thesaurus, une ontologie, un standard, etc. externe au système métalinguistique ASA. Les figures 11.13 et 11.14 nous montrent deux exemples concrets : le référencement dans Wikipedia de l'expression normalisée <Argentine> qui fait partie du thesaurus commun ASA (figure 11.13) ; le référencement de l'expression normalisée <Nénètse> dans le glossaire des langues du monde de l'Ethnologue (figure 11.14) afin de récupérer via ce pont les données linguistiques sur cette langue samoyèdes et de pouvoir communiquer avec tous les systèmes métalinguistiques utilisant ce glossaire.

11.8 L'Atelier de Modélisation ASA

L'Atelier de Modélisation ASA fait partie du Studio ASA (voir chapitre 2 ainsi que [STO 11a]). Il sert d'abord au développement et à la gestion de toutes les ressources métalinguistiques pour l'analyse de l'*univers du discours** ASA que nous venons de présenter brièvement dans ce chapitre. Cette fonction – capitale – est réservée à l'administrateur, au responsable, voire au groupe de responsables des ressources métalinguistiques communes et ouvertes à toutes les archives audiovisuelles.

Mais l'Atelier de Modélisation ASA sert également à la spécification, au développement et à la gestion des *bibliothèques de modèles de description** pour analyser l'*univers du discours** d'une archive audiovisuelle précise comme, par exemple, celui des archives ARC, ALIA et ADA (c'est-à-dire celui des trois chantiers principaux d'expérimentation du projet ASA-SHS) dont chacune dispose de sa propre bibliothèque.

L'outil principal de l'Atelier de Modélisation ASA est, pour le moment, un outil qui s'intitule OntoEditeur. « Pour le moment » veut bien dire que cet outil pourra être remplacé dans l'avenir par d'autres outils (plus élaborés) mais sans que l'*organisation conceptuelle* du système métalinguistique en soit affectée.

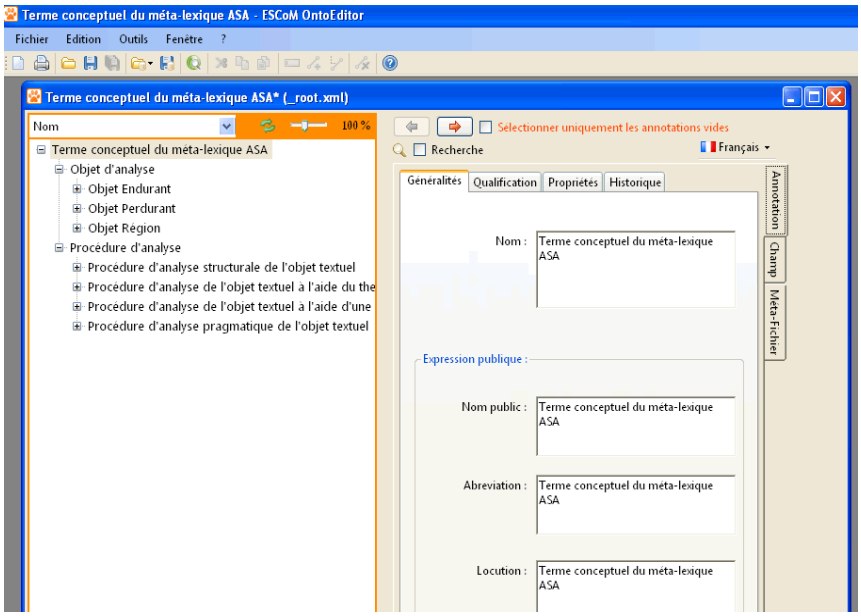

Figure 11.15. *Interface générale de l'outil OntoEditeur*

OntoEditeur est un éditeur de fichiers xml utilisés pour le développement, la gestion et l'enrichissement d'une part des ressources métalinguistiques ASA et d'autre part des ontologies de domaine qui sont propres aux différentes archives audiovisuelles faisant partie de l'*univers du discours** ASA.

Dans cette section, nous proposons une présentation très brève de l'interface de travail OntoEditeur et de l'organisation des ressources métalinguistiques ASA sous forme d'un ensemble de fichiers xml.

La figure 11.15 montre un extrait de l'interface de travail de l'outil OntoEditeur. La partie gauche est réservée à l'affichage du fichier xml concerné pour travailler tel ou tel aspect du système métalinguistique. Celui-ci est composé d'un ensemble de fichiers que nous présenterons ci-après.

Le travail de développement et/ou de gestion d'une ressource métalinguistique telle que les deux méta-lexiques, le thesaurus ou encore les briques configurationnelles pour construire une bibliothèque de modèles de description, se déroule sur trois niveaux :

1) Le premier niveau (appelé *Annotation* ; voir figure 11.15) est réservé à la saisie de la dénomination d'un élément métalinguistique (terme conceptuel ; valeur d'un terme conceptuel ; intitulé d'un schéma ou d'une séquence ; intitulé du modèle de description ; etc.) ainsi qu'à sa qualification (c'est-à-dire à sa définition verbale et sa description, son exemplification, etc.). Une distinction très utile ici est celle qui est faite entre *différents noms* (i.e. entre différents registres sociolinguistiques de dénomination) qu'on peut attribuer à un élément métalinguistique selon le contexte d'usage de cet élément : *usage interne* et *technique* de l'élément en question, *usage public* (par exemple sur un site portail), *usage* sous forme d'une *locution explicative* ou encore *usage* sous forme d'une *abréviation* (d'un acronyme, d'une icône identificatrice, etc.).

2) Le deuxième niveau (appelé *Champ (de définition)* ; voir figure 11.15) est réservé, comme son nom l'indique déjà, à la définition opérationnelle d'un élément métalinguistique : définition d'un schéma de définition à l'aide d'une sélection de termes conceptuels ; définition d'un terme conceptuel à l'aide d'un schéma d'indexation *stricto sensu* (voir chapitre 16) ; définition d'un modèle de description à l'aide d'une sélection de séquences de description ; etc. On trouve ici également l'opération de la mise en relation entre un item métalinguistique du système ASA et une référence externe au dit système : mise en relation d'un terme conceptuel du méta-lexique désignant les objets d'analyse de l'univers du discours ASA avec une expression (ou une liste d'expressions) d'un thesaurus ou d'une ontologie externe ; mise en relation d'un schéma ou d'une séquence de description ASA avec un élément faisant partie d'un standard ou d'une norme.

3) Le troisième niveau (appelé *Méta-Fichier* ; voir figure 11.15) est enfin réservé à un ensemble d'activité de gestion des ressources métalinguistiques : mise en relation des fichiers composant l'ontologie du domaine d'une archive audiovisuelle particulière ; suivi des principales propriétés de chaque élément métalinguistique (dont notamment l'identifiant unique définissant un élément donné) et de l'historique des activités relatives à un élément métalinguistique donné.

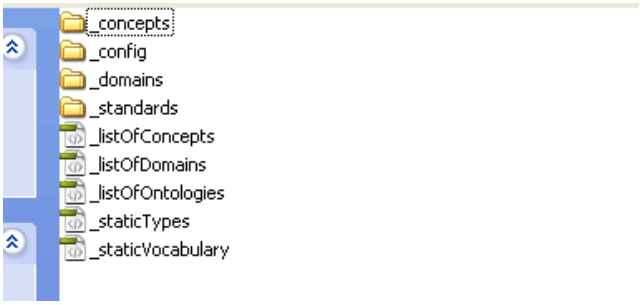

Figure 11.16. *Organisation générale du système des ressources métalinguistiques ASA*

Voyons encore rapidement l'organisation générale du système des fichiers xml contenant les ressources métalinguistiques ASA. La figure 11.16 nous montre que ce système de fichiers contient, entre autres, quatre grands dossiers dont chacun possède une fonction spécifique.

On y trouve notamment le dossier « _concepts » où sont placés tous les fichiers qui contiennent les ressources métalinguistiques communes à l'univers du discours* ASA (voir figure 11.17) : les deux méta-lexiques identifiant les *objets d'analyse** de l'univers du discours ASA et les *activités d'analyse** ; la bibliothèque des schémas d'indexation à proprement parler (voir chapitre 16) ainsi que les types de données avec lesquels nous travaillons (données textuelles ; données numériques ; données de localisation physique, etc.).

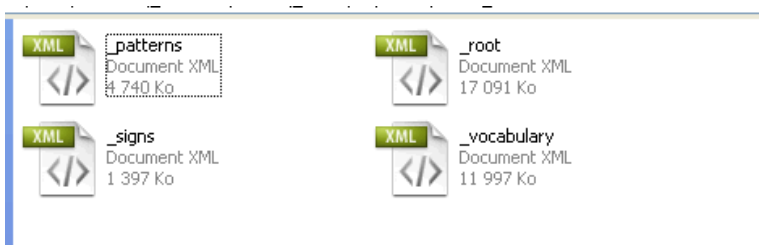

Figure 11.17. *Organisation canonique du domaine ASA*

Un autre dossier qui fait partie du système des fichiers (voir figure 11.16) contenant les ressources métalinguistiques ASA s'intitule « _domaines ». Comme le montre la figure 11.18, ce dossier réunit les ontologies spécifiques à un domaine et qui sont développées à l'aide des ressources métalinguistiques communes à l'univers du discours ASA. On y trouve, actuellement, les ontologies du domaine ARC (consacré à la diversité culturelle et à la communication interculturelle), du domaine ALIA (consacré au patrimoine littéraire), du domaine ADA (consacré à la recherche

en archéologie), du domaine PCIA (consacré au domaine de l'héritage culturel immatériel du monde andin) et du domaine PCA (consacré au domaine de l'héritage culturel de l'Azerbaïdjan). Cette liste est, bien évidemment, totalement ouverte.

Figure 11.18. *Les univers du discours propres à chaque archive*

La figure 11.19 nous montre le dossier à l'aide duquel nous gérons l'ontologie d'un domaine spécifique (tel que celui des archives ALIA ou ADA). Ce dossier se compose d'une manière canonique de trois fichiers :

1) le fichier contenant les *schémas de définition* des objets d'analyse* et des *activités d'analyse* propres à l'*univers du discours** de l'archive concernée ;
2) le fichier contenant les *séquences de description** considérées pertinentes pour traiter la thématisation des objets de connaissance ainsi que, le cas échéant, leur mise en discours, leur expression audiovisuelle ou encore la position de l'analyste par rapport à son objet analysé ;
3) ainsi que le fichier contenant la *bibliothèque des modèles de description** à l'aide de laquelle l'univers du discours de l'archive concernée est décrit, explicité, adapté, exploité, etc.

Le cas échéant, d'autres fichiers peuvent s'ajouter aux trois fichiers mais ceux derniers sont *obligatoirement* présents dans tous les dossiers dédiés à la gestion de l'ontologie d'un domaine spécifique, voir d'une version spécifique de l'ontologie du domaine).

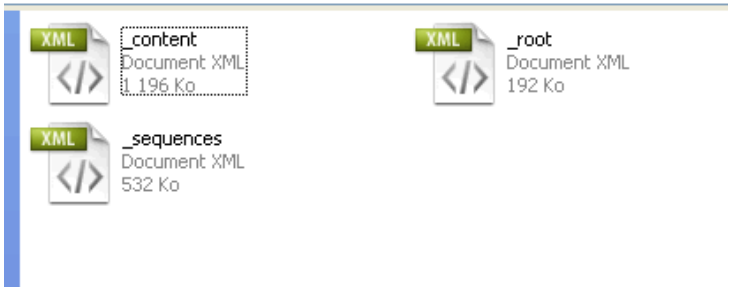

Figure 11.19. *Organisation canonique d'un domaine*

Enfin, le système de fichiers contenant les ressources métalinguistiques ASA inclut encore d'autres dossiers spécialisés ainsi qu'un petit ensemble de fichiers xml. Signalons ici seulement le fichier « static vocabulary » qui contient le thesaurus ASA (voir figure 11.16) ainsi que le fichier « listOfOntologies » qui identifie les thesaurus, terminologies, ontologies et autres standards ou normes extérieurs avec lesquels il est possible de créer des ponts d'interopérabilité.

Chapitre 12

Le méta-lexique représentant l'univers du discours ASA

12.1. Introduction

Rappelons que le méta-lexique des termes conceptuels constitue en quelque sorte le « cœur » des *ressources métalinguistiques** ASA. C'est lui qui fournit le *vocabulaire* nécessaire pour la définition et l'élaboration des modèles de description dont se sert l'analyste – via l'interface de travail du Studio ASA (et, plus particulièrement, de l'Atelier de Description) – pour analyser les textes audiovisuels d'une archive.

Ce méta-lexique est composé, comme nous l'avons déjà dit dans le chapitre précédent, de deux parties parfaitement complémentaires :

– la première partie est réservée au vocabulaire conceptuel qui couvre le *discours-objet* ASA, c'est-à-dire les différents types d'*objets d'analyse** qui font partie de l'*univers du discours** ASA et qui sont susceptibles d'être décrits et indexés par l'analyste ;
– la deuxième partie est réservée au vocabulaire conceptuel qui couvre les *activités d'analyse** mises à la disposition de l'analyste pour décrire les dits objets.

Dans ce chapitre, nous allons présenter la partie du méta-lexique des termes conceptuels qui est réservée à l'identification et à la désignation des types d'objets d'analyse de l'univers du discours ASA.

La section 12.2 est consacrée à quelques explications relatives aux rapports entre « terme conceptuel » et « thème ».

Dans la section 12.3, nous discuterons encore une fois sur la question de la définition de la structure topique qui, comme déjà expliqué dans le chapitre 5, constitue une partie centrale de la structure thématique que l'analyste se propose de décrire et d'expliciter lors de son travail d'analyse du contenu d'un texte ou d'un corpus de textes audiovisuels.

Dans la section 12.4, nous reviendrons sur la notion de l'univers du discours d'une archive audiovisuelle qui, pratiquement parlant, est traitée à l'aide d'une bibliothèque de modèles de description propre à l'archive en question.

Les sections 12.5 et 12.6 sont consacrées à une discussion approfondie des principes d'organisation du méta-lexique des termes conceptuels identifiant et représentant les objets d'analyse de l'univers du discours ASA.

Enfin, la section 12.7 décrit encore rapidement les différentes étapes d'élaboration du méta-lexique en question ainsi que de celui, présenté dans le chapitre 14 est consacré à l'identification et à la représentation des activités d'analyse de l'univers du discours ASA.

12.2. « Terme conceptuel » et « thème » – quelques explications

Avant de présenter la partie du vocabulaire conceptuel représentant l'univers de discours ASA, précisons encore ce que nous entendons par terme conceptuel.

Le terme conceptuel *exprime* un *concept* ou, plutôt, une *notion*, un *thème*. Un thème est un lieu de connaissances qui permet à un acteur (à un *agent*) de reconnaître et de classer des situations, objets ou événements, d'interagir avec eux et de les utiliser d'une manière appropriée selon ses intérêts, besoins ou désirs. Dans ce sens, le grand phénoménologue et sociologue Alfred Schütz définit le thème comme un *schéma typique* ou encore comme un *schéma de typification* (en allemand, *Typisierungsschema* [SCH 03]). Par exemple, un grand nombre de villages historiques possède une topographie typiquement concentrique avec, au milieu, une place souvent limitée par l'église, la mairie, parfois l'école, des lieux de rencontres et d'échanges économiques, etc. Cet agencement à la fois spatial et social constitue une structure typique qui organise notre compréhension culturelle d'un village, d'une petite commune rurale. Il fournit en quelque sorte une *définition implicite*, une

définition acquise par expérience nous permettant de classer tel ou tel agglomération sous le terme *village* ou encore sous le terme *village historique européen*.

Mais comme on le sait aussi, cette représentation peut devenir un obstacle à nos agissements si nous nous trouvons dans une agglomération possédant une organisation spatiale (et sociale) différente, voire inclassable et donc incompréhensible, « chaotique », etc. par rapport au schéma – au lieu de connaissances – que nous utilisons *instinctivement*, *par routine* pour reconnaître et classer des agglomérations et d'interagir avec ces dernières.

Ce petit exemple montre qu'un thème compris comme un lieu de connaissances ou encore au sens phénoménologique comme un *schéma typique*, possède toujours une fonction *indexicale*. Cela veut dire qu'il est toujours dépendant d'un contexte historique, social et culturel. Il peut être plus ou moins commun à un acteur social, plus ou moins controversé, plus ou moins spécialisé, plus ou moins formalisé, etc. Nous renvoyons ici de nouveau aux excellentes explications de Schütz [SCH 03] au sujet de la *structuration thématique du monde social* (explications reprises par Habermas [HAB 81] dans sa théorie du faire communicatif). L'explicitation (toujours révisable) d'un thème est, dans ce sens, affaire d'une *sémiotique** des cultures, voire d'une *anthropologie sémiotique* au sens de C. Geertz [GEE 86].

Compris ainsi, un *thème* ressemble beaucoup à un *modèle de description**. En effet, l'expression (française) « village » pour désigner notre compréhension intuitive d'agglomérations spatiales de type [Village] est plutôt une abréviation pour l'expression linguistique plus appropriée « Village historique en Europe » (l'expression abrégée « village » est en effet commode mais dangereuse : elle insinue une sorte de prétention à l'universalité de notre *définition implicite et culturellement indexée* du terme « village » et constitue donc, comme tant d'autres expressions linguistiques que nous utilisons tous les jours, le foyer potentiel pour une attitude classifiable comme culturo-centriste.

Le terme conceptuel [Village] dans son acceptation implicite de « village historique européen » est organisé – pour parler ainsi – par un jeu d'interactions entre différents termes conceptuels qui désignent le *contexte historique* et *géographique* pertinent pour notre compréhension de l'objet « village », de son architecture, de sa structure topographique, de sa la taille sociodémographique, etc. Selon les préférences individuelles ou selon des visions stéréotypées dominantes, certains de ces éléments peuvent devenir plus importants que d'autres, le schéma lui-même peut s'adapter et intégrer de nouveaux éléments permettant de tenir compte des évolutions des villages historiques qui se trouvent, par exemple, près des grandes métropoles.

Ce que nous retenons de ce petit exemple et, en nous référant de nouveau à Schütz ainsi qu'aux travaux sémantiques lexicales de Greimas [GRE 66], c'est le fait qu'un thème au sens d'un lieu de connaissances (et de reconnaissance) doit être

appréhendé plutôt par le recours à une *configuration thématique* exprimée par une sélection et un regroupement de termes conceptuels que par le recours à un seul terme conceptuel pris individuellement (voir à ce propos nos remarques dans le chapitre 6, section 6.3). Autrement dit, un seul terme conceptuel n'acquiert un sens qu'en relation avec d'autres termes conceptuels avec lesquels il exprime un thème, une notion.

Ainsi, il ne faut pas réduire le métalangage de description au seul vocabulaire taxinomiquement organisé de termes conceptuels. Il repose, au contraire, et comme nous l'avons déjà montré à l'aide de toute une série d'exemples concrets, sur le concept fondamental de la *configuration* (concept central également dans la théorie *sémiotique** de Greimas [GRE 79]), c'est-à-dire sur la *sélection* d'un ensemble de *termes conceptuels* et leur *positionnement* à l'aide de relations spécifiques telles que les relations logiques, les relations attributives, les relations de localisation, les relations rhétoriques, et ainsi de suite.

Cependant, pour qu'il y ait sélection de termes conceptuels, il faut déjà un vocabulaire bien défini de tels termes conceptuels sélectionnables.

12.3. La structure définitionnelle d'une topique

Le vocabulaire des termes conceptuels du méta-lexique ASA dont le terme de base est [Objet d'analyse] (voir chapitre 11, figure 11.3) sert d'abord (mais pas exclusivement !) à représenter les *domaines référentiels de connaissance* thématisés ou thématisables dans les corpus composant les archives qui forment les chantiers expérimentaux du projet ASA-SHS[52] : diversité culturelle (domaine référentiel du portail Archives Rencontre des Cultures), patrimoine littéraire (domaine référentiel du portail Atelier Littéraire d'Ici et de Loin) et archéologie (domaine référentiel du portail Atelier des Arkéonautes).

Ces domaines référentiels sont abordés à l'aide de *modèles de description thématique** (et, plus spécifiquement, *topique*[53]), c'est-à-dire à l'aide de configurations ou structures de situations, de pratiques, d'acteurs, d'œuvres, d'environnements et milieux, etc. – structures ou configurations conceptuelles que l'analyste, le cas échéant, adapte à son travail d'analyse et renseigne avec des informations provenant du texte ou du corpus de textes audiovisuels soumis à l'analyse.

Mais la notion de l'*univers du discours** d'un texte audiovisuel (d'un corpus ou d'une archive audiovisuelle) n'est pas limité à la seule prise en compte des domaines

---

[52] carnet de recherche du projet : http://asashs.hypotheses.org/

[53] au sujet du rapport entre « configuration thématique » et « configuration topique » stricto sensu, voir nos explications au chapitre 5.3 ainsi que la figure 5.1

référentiels de connaissance. Il recouvre également, comme nous l'avons déjà fait remarquer à plusieurs reprises, les objets du *texte* et du *discours*, c'est-à-dire les spécificités et contraintes des instruments ou outils de la *médiatisation d'une connaissance*, de son *expression*, de sa *communication*, de son *appropriation* et aussi de sa *conservation* et *transmission*. Comme déjà expliqué (voir chapitre 1), le texte audiovisuel traite un domaine de connaissances selon un certain point de vue et en référence à un cadre culturel. Il en thématise et exprime certains aspects qu'il hiérarchise et développe « à sa façon » et en réponse à un contexte donné de médiatisation de connaissances. Les *termes conceptuels** permettant de représenter les objets « texte » et « discours » constituent donc des éléments indispensables du vocabulaire des termes conceptuels dont la racine est le terme [Objet d'analyse].

Souvenons-nous de notre exemple développé dans le chapitre 3 relatif à la *description – indexation* de textes audiovisuels traitant de formations culturelles (cultures techniques, vestimentaires, immatérielles, etc.) de civilisations sur le continent américain à une époque historique donnée. La définition (la plus simple) de la structure thématique pour pouvoir traiter systématiquement ce type de contenu prévoit :

– d'une part une définition de la partie purement référentielle de la structure topique en question,
– et d'autre part une définition des parties de la mise en discours de la topique, de l'expression (audiovisuelle) de la topique ou encore de l'explicitation du « regard » de l'analyste sur la topique telle qu'elle est traitée dans un texte donné.

Chaque partie de la structure topique est définie par un ou un ensemble de termes conceptuels entre lesquels s'établissent des relations spécifiques. Ainsi, la *partie purement référentielle* de la structure topique *Formation culturelle d'une civilisation sur le sol américain* est composée par la sélection suivante de termes conceptuels :

– [Formation culturelle],
– [Civilisation],
– [Amérique],
– [Période].

Ces termes conceptuels se positionnent les uns par rapport aux autres à l'aide des *relations* suivantes :

– *{réfère à} :* {[Civilisation] réfère à [Formation culturelle]} ;
– *{est géographiquement localisé à} :* {{[Civilisation] réfère à [Formation culturelle]} est localisé géographiquement à [Région géographique : <Amérique>]} ;
– *{est chronologiquement localisé dans} :* {{[Civilisation] réfère à [Formation culturelle]} est chronologiquement localisé dans [Période]}.

Les relations conceptuelles forment une autre partie du *métalangage de description** ASA où elles constituent en effet la *bibliothèque de schémas* et de *séquences* (voir chapitre 16) à l'aide desquels les termes conceptuels sont positionnés les uns par rapport aux autres.

Les formulaires interactifs de travail représentés dans le chapitre 3 par les figures 3.1 à 3.6 intègrent ces structures composant ensemble la configuration définitionnelle sous-tendant la description – indexation de la topique *Formation culturelle d'une civilisation sur le sol américain*. L'analyste d'un texte audiovisuel, lui, sélectionne un, voire plusieurs termes conceptuels dans son interface de travail et en rend compte en l'indexant librement, le décrivant, l'annotant, etc. tout en respectant certaines *règles d'usage* dont, notamment celle qui veut que certains termes conceptuels *présupposent* d'autres termes conceptuels (par exemple, dans notre cas, le terme conceptuel [Civilisation] est présupposé par le terme [Formation culturelle] et le terme conceptuel instancié [Région géographique : <Amérique>] est présupposé par le terme [Période] ; pour des explications plus détaillées, voir chapitre 3).

12.4. L'univers du discours ASA

Les structures ou configurations thématiques forment les principales collections des bibliothèques ARC, ALIA ou ADA des modèles de description du contenu audiovisuel. Ainsi, la figure 12.1 montre un extrait des collections composant la *bibliothèque des modèles de description** ALIA. Comme on peut le voir, le domaine référentiel de connaissances ALIA est abordé sous forme d'un système hiérarchique de *topoï* appelés *sujets*. A la racine de la bibliothèque des modèles de description ALIA, on trouve quatre grandes classes de sujets :

1. *Sujet « Littérature française »* – classe des sujets qui traitent différents aspects de la littérature française (histoire, auteurs, œuvres, courants littéraires, etc.) ;
2. *Sujet « Littérature du monde »* – classe des sujets qui traitent des littératures nationales, des littératures par langues, des œuvres et des auteurs, et ainsi de suite ;
3. *Sujet « Vie littéraire »* – classe des sujets qui traitent des pratiques littéraires, des usages de la littérature, de l'édition, de la diffusion et des critiques littéraires ;
4. *Sujet « Recherche littéraire »* – classe des sujets qui traitent, enfin, la recherche à proprement parler consacrée à la littérature, au texte littéraire, à l'histoire littéraire, à la réception de la littérature, etc.

Ces quatre classes de sujets forment en même temps les quatre principales collections des formulaires interactifs dédiés à la description du contenu audiovisuel des textes faisant partie des archives ALIA. L'analyste choisira parmi ces collections

le formulaire approprié pour décrire et indexer son texte ou son corpus de textes audiovisuels. Précisons en passant que ni l'organisation des collections (et sous-collections) de formulaires d'analyse de sujets ni le nombre de ces derniers ne sont figés – ils peuvent évoluer selon les besoins de l'analyste ou, plus généralement, des missions et politiques d'analyse de telle ou telle archive. Cependant, le fait de faire évoluer une bibliothèque de formulaires destinés à l'analyse de sujets concrets dans un corpus de textes audiovisuels est une tâche ardue et compliquée qui peut avoir des conséquences importantes sur les corpus de textes audiovisuels déjà analysés et publiés.

La figure 12.1 nous montre que le « grand » sujet *Littérature française* est lui-même décomposé en cinq sujets plus spécialisés parmi lesquels on trouve le sujet *Littérature française par type*. Ce sujet est à son tour composé des quatre sujets suivants :

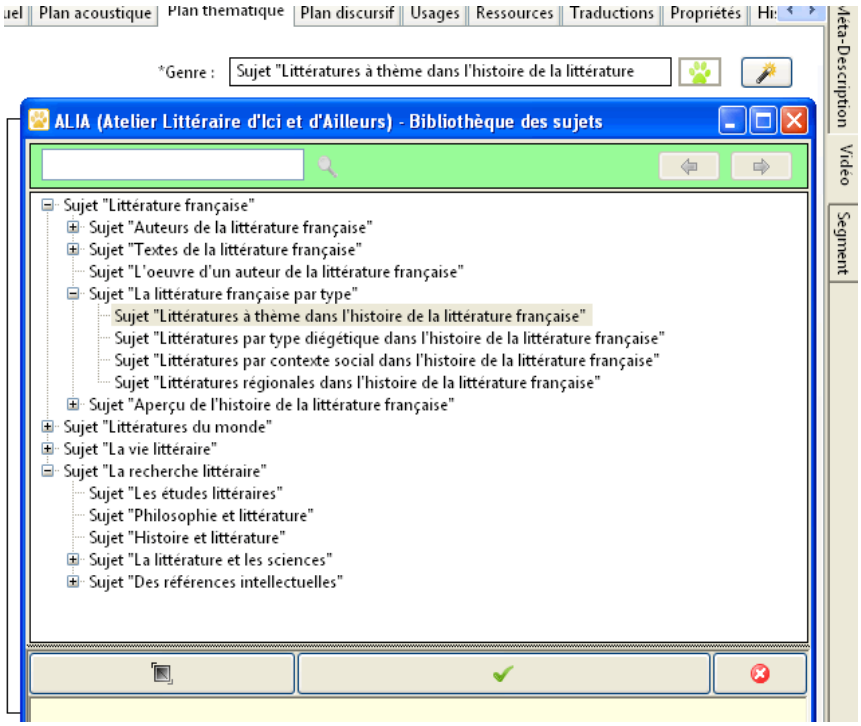

Figure 12.1. *Vue d'ensemble de la bibliothèque des modèles de description des corpus audiovisuels faisant partie des archives ALIA*

– Sujet *Littérature à thème dans l'histoire de la littérature française* – formulaire permettant de décrire des sujets relatifs, par exemple, à la

littérature du voyage, à la littérature géographique, à la littérature fantastique, etc. ;
– Sujet *Littérature par type diégétique dans l'histoire de la littérature française* – formulaire permettant d'aborder des sujets relatifs à, par exemple, la littérature romanesque, la poésie, le théâtre, etc. ;
– Sujet *Littérature par contexte social dans l'histoire de la littérature française* – formulaire permettant d'aborder des sujets relatifs, par exemple, à la littérature populaire, la littérature de la cour, la littérature pour les jeunes lecteurs, etc. ;
– Sujet *Littératures régionales dans l'histoire de la littérature française* – formulaire permettant de décrire, par exemple, des sujets relatifs à la littérature bretonne, alsacienne, picarde, occitane, corse.

Décomposer ainsi le domaine ALIA en un ensemble de *topoï*, de *configurations* ou *structures topiques** typiques, est un choix motivé par une « politique » qui sous-tend la ou les missions d'une archive : mission relative à la couverture empirique du domaine, mission relative à la publication et diffusion de son patrimoine audiovisuel, mission relative à la préservation à long terme et à la transmission de son patrimoine.

Autrement dit, les *topoï* que nous avons retenus dans le cadre de notre projet de recherche ASA-SHS et qui a donné naissance aux archives ALIA n'ont, bien évidemment rien de nécessaire ou d'exclusif. D'autres archives ou bibliothèques numériques qui s'occupent du domaine littéraire peuvent le concevoir d'une manière différente. Toujours est-il que toute entreprise d'analyse doit obligatoirement s'appuyer sur des structures ou configurations thématiques (et, plus spécifiquement, topiques) et donc assumer des choix intellectuels et « politiques ». Une fois adopté le choix pour un ou un ensemble de *topoï* représentatifs d'un domaine référentiel de connaissance, chaque topos doit être explicité et décrit d'une manière générique sous forme d'une *configuration définitionnelle*, c'est-à-dire sous forme d'une structure qui définit l'organisation interne du topos. Nous en avons vu toute une série d'exemples concrets dans le chapitre 5 et nous y reviendrons encore ci-après (voir section 12.4).

La qualification d'un topos sous forme d'une structure ou configuration définitionnelle explicite ne peut se faire qu'à l'aide d'un *métalangage*. D'où l'importance capitale, pour tout métalangage, du méta-lexique ASA en général et du vocabulaire des termes conceptuels dont la racine est [Objet d'analyse], en particulier.

Ensemble avec la *bibliothèque des schémas* et des *séquences* (voir chapitre 16) représentant la partie relationnelle (c'est-à-dire la *bibliothèque des relations* entre termes conceptuels) et le *thesaurus* (voir le chapitre 15), le *méta-lexique* des termes conceptuels, rappelons-le, forme l'un des constituants les plus essentiels de tout métalangage de description de corpus textuels – audiovisuels ou autres.

Revenons à la question du domaine référentiel des connaissances thématisé ou simplement thématisable dans un corpus audiovisuel. Tous les topoï identifiés et qualifiés sous forme de configurations définitionnelles constituent, ensemble, la vision particulière que possède une archive (ou une bibliothèque) de son domaine de connaissances.

On comprend bien, qu'il subsiste ici une petite ambiguïté dans l'emploi du terme *domaine référentiel de connaissances* :

> – d'une manière intuitive et « pré-analytique », ce terme signifie la *réalité donnée* dont traite ou dont est question dans une archive ou une bibliothèque ;
> – d'une manière explicite et analytique, ce terme signifie, en revanche, la *représentation*, la « vision » que possède une archive ou une bibliothèque d'une réalité donnée sous forme d'un ensemble de structures topiques.

Nous ne sommes concernés ici que par la deuxième acceptation du terme « domaine référentiel de connaissances ». Or, même si deux archives traitant du même domaine référentiel (au *premier sens*, c'est-à-dire au sens intuitif et pré-analytique) en possèdent deux visions différentes se manifestant sous forme de deux systèmes de structures topiques différentes, chacune des deux archives peut s'appuyer sur le même métalangage de description – et donc sur le même méta-lexique des termes conceptuels – pour élaborer ses structures topiques et sa bibliothèque de modèles de description. Autrement dit, une archive consacrée aux connaissances littéraires mais n'utilisant pas les mêmes modèles de description que les archives ALIA, peut néanmoins s'appuyer sur le métalangage ASA pour élaborer sa vision du domaine référentiel en question sous forme d'un ou d'un système de structures topiques.

On comprend bien ici l'intérêt d'un tel métalangage qui, entre autre, permet non seulement la prise en compte d'une certaine *relativité*, d'une certaine diversité de visions d'une même « réalité donnée » mais qui peut servir également de *ressource commune* à des entreprises de conception et de modélisation concurrentes. Enfin, il – ce métalangage – permet également d'assurer (jusqu'à un certain niveau) – la traductibilité et l'interopérabilité des métadonnées relatives au contenu (aux sujets) des textes audiovisuels même si elles sont réalisées à l'aide de formulaires de travail qui font partie de bibliothèques de formulaires concurrents.

Par ailleurs, même si un domaine référentiel de connaissances (au *deuxième sens*, voir *supra*) est propre à une archive (ou une bibliothèque) audiovisuelle, telle ou telle structure topique ne l'est pas obligatoirement. Ainsi, si deux archives s'occupant du « même » domaine référentiel de connaissances (au *premier sens*) en possèdent une vision différente (c'est-à-dire sous forme d'une bibliothèque divergente de modèles de description du contenu), une structure topique particulière

peut trouver emploi telle quelle ou moyennant quelques modifications locales dans les deux bibliothèques. D'une manière encore plus générale, une structure topique telle que celle représentée par la figure 3.2 (voir chapitre 3) et définissant le fait (très général) qu'une *civilisation (quelconque) se réfère à une formation culturelle (quelconque)*, peut être pertinente *pour toute une diversité d'archives* et de bibliothèques même si ces dernières n'ont en commun que très peu d'intérêts et de domaines de connaissances (au *premier sens*).

Néanmoins, il ne faut pas perdre de vue non plus que le méta-lexique des termes conceptuels ASA des objets d'analyse représente, *in fine*, une certaine vision (théorique) du *monde de vie des acteurs sociaux et de sa médiatisation sous forme d'un discours et d'un texte* que l'analyste peut utiliser pour renseigner son domaine d'expertise tout en l'adaptant aux spécificités de ce dernier. Ainsi, comme toute ontologie, le vocabulaire des termes conceptuels du méta-lexique ASA représentant les objets d'analyse dans l'univers du discours ASA, est triplement « limité » :

– le vocabulaire exprime une certaine vision des *objets d'analyse* qui proviennent notamment (comme nous le verrons ci-après) du *monde social* et de sa *médiatisation* ;
– le vocabulaire exprime cette vision sur un certain niveau de généralité ou de « granularité » ;
– la vision exprimée par le vocabulaire est intrinsèquement partielle.

Or, les modifications du vocabulaire des termes conceptuels des objets d'analyse de l'*univers du discours** ASA, se référeront à ces trois types de limitation et deviendront ainsi contrôlables. C'est un enjeu important de pouvoir conjuguer la « rigidité » de l'organisation du méta-lexique des termes conceptuels génériques avec sa nécessaire adaptation aux attentes et besoins spécifiques des utilisateurs (des « concept designers » et des analystes) et aux évolutions historiques du monde.

12.5. L'organisation générale du vocabulaire relatif aux objets d'analyse dans l'univers du discours ASA

Considérons maintenant la figue 12.2 qui nous montre la base canonique et les catégories supérieures du vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA. Ces termes conceptuels dont le terme de base est [Objet d'analyse], sont identifiés, définis et classés pour répondre aux besoins d'analyse de corpus audiovisuels variés parmi lesquels on trouve ceux qui documentent les domaines de l'histoire et de la littérature, de l'archéologie et de la diversité culturelle.

Comme nous l'avons déjà fait remarquer, la portée empirique d'un terme conceptuel *pris isolément* de ce vocabulaire dépasse de loin la portée empirique des trois domaines cités ce qui veut dire qu'il peut être utilisé pour la définition de modèles de description de ressources audiovisuelles qui n'ont rien à voir avec lesdits domaines. Une *distinction* est donc à faire entre les *deux plans* suivants :

– le plan du *modèle de description** d'un texte ou d'un corpus de textes audiovisuels ;
– et, le plan des *termes conceptuels** qui composent le modèle de description.

Pris isolément, un terme conceptuel n'est évidemment pas spécifique à un domaine d'analyse choisi ; considéré en relation avec un ou plusieurs autres termes avec lesquels il forme une *configuration conceptuelle** [STO 87, STO 93], il devient spécifique et propre à un domaine d'analyse (tel que celui du corpus audiovisuel documentant, par exemple, les grands courants de la littérature française, ou les fouilles archéologiques autour du monde).

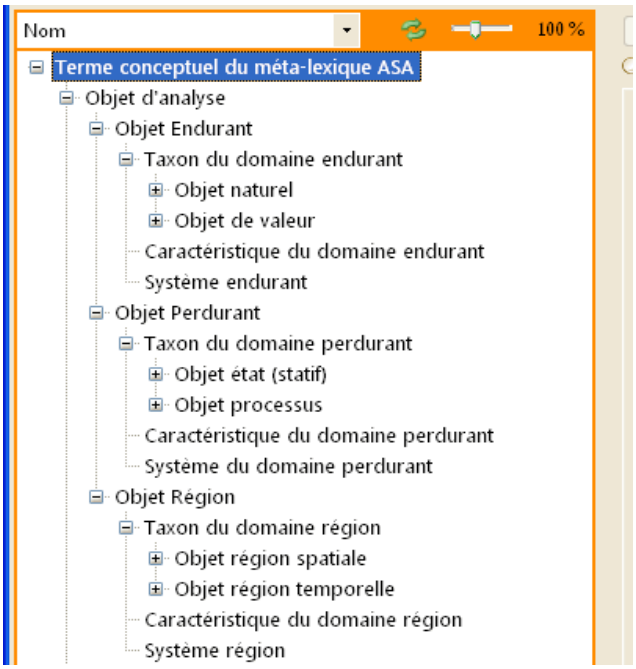

Figure 12.2. *Base canonique du vocabulaire des termes conceptuels représentant les objets d'analyse de l'univers du discours ASA*

La base canonique du vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA est composée de *trois* termes conceptuels (voir figure 12.2) qui n'ont pas une valeur réellement descriptive mais plutôt *organisationnelle*. En raison de leur très grande généralité, il n'est guère envisagé qu'ils forment des concepts ou des notions à renseigner et à indexer lors d'un travail concret d'analyse d'un texte ou d'un corpus audiovisuel (seule exception : un formulaire d'analyse du contenu qui offre à l'analyste le choix de sélectionner librement dans tout le vocabulaire des termes conceptuels des objets d'analyse ASA, ceux dont il/elle a besoin pour définir sa structure topique, etc.). Ces trois termes conceptuels sont, par contre, essentiels pour la structure taxinomique elle-même du méta-lexique.

Tout en restant « fidèle » à notre cadre théorique et pratique de description de corpus audiovisuels, le choix des termes conceptuels composant la base canonique du vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA s'inspirent de certaines ontologies formelles (en anglais, *top level ontologies*) dont plus particulièrement de l'ontologie DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*)[54]. Même si le vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA ne reprend pas « telle quelle » les catégories de base de l'ontologie DOLCE certains termes conceptuels de base réfèrent, comme le montre la figure 12.2, directement aux travaux de N. Guarino et de son équipe.

Ainsi avons-nous repris (voir figure 12.2) la distinction entre termes conceptuels faisant partie de la branche [Objet Endurant] et termes conceptuels faisant partie de la branche [Objet Perdurant]. Cette distinction qui est amplement décrite, discutée et formalisée dans la littérature spécialisée (voir par exemple, [WON 03]), trouve son « écho » au niveau descriptif dans la distinction qu'on fait en *sémiotique** structurale entre :

– d'une part les *objets* et les *entités* ainsi que leurs assemblages ou réunions en des structures plus complexes ;
– et d'autre part les *procès* et *pratiques* ou encore les *situations* et *états* (au sens d'état d'action) qui perdurent dans le temps.

Comme nous le verrons plus loin, cette distinction aussi catégorique qu'elle semble être à un haut niveau d'abstraction, devient parfois difficile à manier lorsqu'il s'agit de catégoriser des termes conceptuels possédant une valeur descriptive réelle ainsi qu'un impact concret sur le travail d'analyse.

Le troisième terme conceptuel de base – le terme conceptuel [Objet Région] – est également inspiré de l'ontologie DOLCE. Sous le terme conceptuel [Objet Région], nous classons en effet tous les termes qui conceptualisent des étendues et lieux de

[54] L'ontologie DOLCE est développée par le Laboratory for Applied Ontology du CNR (Consiglio Nazionale delle Ricerche) à Trento dont le directeur est Nicola Guarino ; pour plus d'informations, voir le site du laboratoire : http://www.loa-cnr.it/index.html.

l'espace physique, des espaces imaginaires ou abstraites ainsi que des périodes et moments temporels.

L'organisation générale du vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA, repose ainsi sur trois branches principales. « Sous » ces branches, on trouve des catégories conceptuelles supérieures, c'est-à-dire des catégories qui organisent d'autres catégories conceptuelles plus spécialisées, plus riches d'un point de vue intensionnel et moins large, plus circonscrit du point de vue extensionnel. La branche [Objet Endurant] connaît ainsi deux sous-branches (figure 12.2) :

– la sous-branche [Objet naturel] qui réfère aux entités physiques (matérielles, biologiques, etc.) ;
– la sous-branche [Objet de valeur] qui réfère d'une part aux entités physiques possédant un *statut fonctionnel particulier* (c'est-à-dire jouant un rôle particulier dans la vie d'un agent humain ou anthropomorphe) et d'autre part aux *entités du sens*, c'est-à-dire aux entités (à support non-spécifié !) qui font partie de la *culture*, de l'*horizon du sens* (selon Schütz [SCH 03]) d'un tel agent.

La distinction entre [Objet naturel] et [Objet de valeur] retrace également, d'une manière ou d'une autre, la distinction entre « Nature » et « Culture » dans la théorie sémantique de Greimas [GRE 70] et dans l'anthropologie structurale de Lévi-Strauss [LEV 58] – distinction capitale formulée et définie selon les contraintes et les particularités de tel ou tel langage social (c'est-à-dire propre à un *acteur social*) ou encore tel ou tel jeu de langage [WIT 03]. Elle fait également écho à la distinction opérée entre *caractéristiques intrinsèques* d'un objet et *caractéristiques d'un objet en fonction d'un observateur* (d'un sujet) dans l'ontologie sociale développée par Searle [SEA 95].

Ceci dit, les deux termes conceptuels [Objet naturel] et [Objet de valeur] n'ont guère non plus une valeur descriptive – leur intérêt heuristique réside plutôt dans leur capacité de pouvoir classer et organiser des catégories conceptuelles inférieures, c'est-à-dire des catégories conceptuelles plus spécialisées.

La deuxième branche du vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA possède comme taxon le terme conceptuel [Objet Perdurant]. D'une manière générale, les termes de cette catégorie servent à la description d'actions et de pratiques sociales, de situations ou d'états. Ils servent également à la description de processus causaux pour lesquels il n'y a pas vraiment d'agent intentionnel identifiable mais qui peuvent se dérouler aussi bien dans le monde naturel que dans le monde social et historique. La figure 12.2 montre l'organisation générale cette branche sous forme d'une taxinomie de termes conceptuels autour des deux catégories de base [Objet état (statif)] et [Objet processus].

Le taxon [Objet Région] inaugure la troisième branche représentée dans la figure 12.2. Elle réunit tous les termes conceptuels qui réfèrent aux lieux et étendues physiques (naturels ou sociaux), aux moments et périodes temporels ou encore à des régions abstraites (imaginaires, mathématiques, etc.). Comme le montre la figure 12.2, la taxinomie des termes conceptuels de cette branche se développe autour des deux catégories de base suivantes [Objet région spatiale] et [Objet région temporelle].

Dans le prochain chapitre (chapitre 13), nous considérerons plus en détail certaines parties du méta-lexique des objets d'analyse.

12.5. Questions relatives à l'organisation du vocabulaire des termes conceptuels des objets d'analyse ASA

Comme tout travail visant la mise en place d'un métalangage de description, celui-ci est dicté par son objet (c'est-à-dire les corpus composant les chantiers d'expérimentation du projet ASA-SHS) et ses objectifs (analyse du contenu de textes audiovisuels relevant des chantiers cités et utilisation de ces analyses pour une meilleure diffusion dudit contenu).

Ainsi, le vocabulaire des termes conceptuels des objets d'analyse de l'*univers du discours** ASA n'est, bien sûr, *pas achevé* : certaines de ses parties semblent être solidement développées, d'autres moins. C'est le cas, par exemple, des domaines de connaissance qui relèvent de la psychologie, du droit et de l'économie – domaines qui sont peu représentés dans nos corpus de travail même s'ils font évidemment partie de l'univers du discours ASA, c'est-à-dire de l'univers du discours qui s'intéresse au *monde de vie des acteurs sociaux*.

Enfin, la granularité (c'est-à-dire le niveau de précision descriptive) est bien calée par rapport à l'objet et aux objectifs cités de notre projet. Mais rien ne dit qu'elle soit pertinente (telle quelle) pour d'autres projets. Par exemple, un projet d'analyse qui ne vise qu'un « sous-domaine » en archéologie (c'est-à-dire archéologie de telle ou telle époque, telle ou telle approche de l'archéologie, etc.) aura vraisemblablement besoin de termes conceptuels plus spécialisés que ceux dont disposent l'actuel vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA.

Ou prenons encore les archives AAR qui disposent de corpus audiovisuels traitant des grands processus du monde moderne (industrialisation, développement économique, migration, mondialisation, etc.), des situations socio-économiques (emploi, niveau de vie, richesses et misères du monde, etc.), de situations conflictuelles (révoltes, guerres, etc.), de situations d'oppression et de négation de

l'autre (exclusions, génocides, etc.), de représentations et d'idéologies (nationalisme, communautarisme, racisme, etc.), etc. Afin de rendre compte de tels *sujets**, le vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA a du être adapté encore très récemment[55]. Etant donné la nouveauté mais aussi l'importance de ce chantier dans le cadre de la préservation et de l'exploitation de patrimoines scientifiques, il fera l'objet d'une publication ultérieure.

Se pose donc ici la question de l'évolution du méta-lexique des termes conceptuels ASA afin pour qu'il puisse répondre aux besoins et exigences propres à des projets d'analyse concrets. Pour cela, au moins deux aspects doivent être distingués :

– l'*enrichissement des parties taxinomiques* spécifiques du méta-lexique des termes conceptuels (ce processus présuppose la distinction entre modules taxinomiques communs et non-modifiables et modules taxinomiques ajoutés aux premiers) ;
– la *diversification* des *ponts* entre les termes conceptuels du méta-lexique ASA et des ressources métalinguistiques externes au système métalinguistique ASA – ressources métalinguistiques telles que thesaurus, langages d'indexation, terminologies, ontologies ou encore normes et standards (voir nos remarques à ce sujet au chapitre 11).

Revenons aux trois termes de la base canonique du vocabulaire conceptuel des objets d'analyse de l'univers du discours ASA. Ceux-ci ne constituent pas le résultat d'un pur choix apriorique au sens d'un choix pour telle ou telle ontologie supérieure. Comme nous l'avons déjà dit, nous les avons choisis, *in fine*, en référence à l'ontologie supérieure DOLCE. Mais ce choix s'est imposé tardivement dans notre travail d'organisation du méta-lexique. Il a été précédé d'un travail de classification des termes conceptuels retenus pour produire des modèles d'analyse de corpus audiovisuels ainsi que de différentes tentatives de définition d'une *base canonique* et de catégories supérieures de classification.

En somme, l'organisation du vocabulaire en question telle qu'elle est présentée par la figure 12.2, est le résultat d'une double approche :

– d'une part, d'une *approche « lexicale » ou « terminologique »* de reconstruction sémantique de séries de termes conceptuels identifiés lors de la phase préalable de l'analyse des corpus audiovisuels, de recherches comparatives menées sur des ressources métalinguistiques déjà existantes (telles que thesaurus, terminologies et ontologies) ou encore sur des

[55] C'est-à-dire fin octobre 2011

ressources lexicales que nous fournit des programmes tels que le projet *WordNet* de l'université de Princeton[56] ;
– d'autre part, d'une *« enquête » de concepts* (ou catégories conceptuelles) à la fois suffisamment généraux, explicitement définis, « philosophiquement solides » et enfin servant déjà de fondements à d'autres entreprises d'élaboration de métalangage de description.

Le vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA est une *ontologie descriptive* ou, pour reprendre l'expression de B. Bachimont [BAC 05][57], une ontologie *au sens épistémologique* (contrairement à une *ontologie formelle* ou *catégoriale*) qui s'appuie sur l'analyse concrète de corpus de textes audiovisuels documentant un domaine de connaissance (voir à ce propos nos remarques au chapitre 1 et aussi [STO 89, STO 96, STO 98]).

Les concepts de base et l'organisation canonique de ce vocabulaire reflète cependant notre volonté de le rapprocher aux ontologies formelles (catégoriales) existantes[58] et d'en faire, *in fine*, une ontologie formelle dans ce sens précis qu'*elle n'est pas dépendant d'un domaine empirique bien circonscrit* de connaissance.

Une ontologie *formelle* doit être distinguée, comme le remarque très justement Bruno Bachimont [BAC 05] d'une *ontologie formalisée* (au sens mathématique ou logique du terme). Le méta-lexique ASA constitue ainsi une ontologie (ou, plutôt une partie d'une ontologie) *formelle mais non-formalisée* pouvant être utilisée à des tâches descriptives concrètes (sur un certain niveau de granularité) de corpus audiovisuels documentant des domaines empiriques divers de connaissance qui relèvent du monde de vie (sociale, historique, culturelle, naturelle). Comme déjà expliqué ci-auparavant, les corpus de textes audiovisuels soumis à l'analyse documentent trois domaines d'expertise spécifiques : diversité culturelle, littérature et archéologie. Mais les termes conceptuels composant le vocabulaire en question se prêtent plus ou moins facilement à la définition et à l'élaboration de modèles d'analyse de corpus audiovisuels qui n'ont aucun rapport direct avec les trois

---

[56] http://wordnet.princeton.edu/

[57] Voir aussi la présentation enligne de la conférence de B. Bachimont: http://www.spim.jussieu.fr/doc/ontologies/Bachimont-SticSante-08122005.pdf

[58] Outre DOLCE, nous nous sommes référés également aux ontologies suivantes : BFO (*Basic Formal Ontology* ; pour plus d'informations : http://www.ifomis.org/bfo/); SUMO (*Suggested Upper Merged Ontology* ; pour plus d'informations : http://www.ontologyportal.org/); OCHRE (Online Cultural Heritage Research Environment) Core Ontology (http://ochre.lib.uchicago.edu/index_files/Page845.htm) plus spécialisée sur les question relatives aux patrimoines culturels ; …) ; le *Conceptual Référence Model* (CRM) du CIDOC (*Comité International pour la Documentation de l'ICOM* ; pour plus d'informations, voir : http://www.cidoc-crm.org/scope.html) ainsi qu'à l'ontologie GOLD (General Ontology for Linguistic Description) consacrée à la recherche en linguistique (http://linguistics-ontology.org/).

domaines cités. Ils se prêtent cependant moins bien à l'analyse d'objets filmiques qui ont une prétention poétique et esthétique comme c'est le cas, notamment, des œuvres audiovisuelles de fiction.

L'ensemble des termes conceptuels formant le méta-lexique qui nous sert à identifier et à désigner les objets d'analyse dans l'univers du discours ASA, ne dépasse guère les 1100 termes conceptuels[59]. Ceux-ci se regroupent en quelques 85 domaines taxinomiques (pour plus d'explications, voir le chapitre suivant, section 13.3). Chaque domaine taxinomique connaît deux catégories de termes fonctionnellement différentes :

– une *première catégorie* de termes conceptuels qui servent d'abord à l'organisation du domaine taxinomique,
– et une *deuxième catégorie* de termes conceptuels qui servent essentiellement à l'identification du type d'objets de connaissance thématisable dans le discours audiovisuel d'un texte ou d'un corpus de textes soumis à l'analyse.

A ces deux premières catégories de termes conceptuels, il faut encore ajouter une *troisième catégorie* dont la fonction principale est de positionner et de hiérarchiser les domaines taxinomiques eux-mêmes.

Cela explique le fait que parmi les 1100 termes conceptuels composant la version actuelle (fin 2011) de ce méta-lexique, ce ne sont qu'à peine la moitié des termes conceptuels (c'est-à-dire entre 500 et 550 termes conceptuels) qui ont réellement une pertinence dans la construction des modèles de description, c'est-à-dire des modèles dont se sert l'analyste pour traiter un corpus audiovisuel. L'autre moitié des termes conceptuels possèdent plutôt une fonction d'organisation, de structuration du méta-lexique ou de tel ou tel domaine taxinomique spécifique qu'une fonction descriptive à proprement parler.

Aujourd'hui, le méta-lexique des termes conceptuels représentant les objets d'analyse de l'univers du discours réunit :

– quelques *30 termes conceptuels* de la branche [Objet région spatiale] pouvant être utilisés pour décrire des lieux, des étendues, des reliefs naturels, des régions et territoires, etc. Pour la description des corpus audiovisuels faisant partie des archives du projet ASA-SHS[60], nous n'utilisons, actuellement, pas plus qu'une quinzaine de termes conceptuels de ce groupe ;

---

[59] C'est-à-dire dans la version actuelle (dernière mise à jour, fin novembre 2011)
[60] Rappelons qu'il s'agit d'abord des trois archives ARC, ALIA et ADA (qui forment les chantiers principaux des activités d'expérimentation pour ce projet)

– quelques *25 termes conceptuels* de la branche [Objet région temporelle] pouvant être utilisés pour décrire des périodes, dates ou encore événements. Pour le moment, seulement quelques 18 termes conceptuels de ce groupe sont réellement utilisés dans les modèles de description ;

– quelques *75 termes conceptuels* de la branche [Perdurant] qui nous permettent de décrire des activités, des processus intentionnels ou causaux, des pratiques sociales ainsi que des états ou situations (naturelles, sociales, etc.). Nous utilisons actuellement de ce groupe environ 40 termes conceptuels ;

– quelques 400 termes conceptuels de la branche [Objet Endurant] – branche qui est de loin la plus importante dans notre méta-lexique. Les termes conceptuels de cette branche nous permettent de décrire des entités non-vivantes et vivantes, des artefacts et produits, des acteurs individuels et collectifs, des expressions de toute sorte, des systèmes de pensée et de croyance, etc. Ici de nouveau, nous n'utilisons aujourd'hui dans nos modèles de description pas plus que la moitié des termes conceptuels disponibles.

12.6. Le processus de l'élaboration du vocabulaire des termes conceptuels des objets d'analyse ASA

Revenons à notre travail d'analyse des trois grands corpus audiovisuels documentant des patrimoines culturels et scientifiques spécifiques (diversité culturelle, littérature et archéologie), l'objectif de ce travail a été d'abord d'identifier des *séries* de *sujets* – de « thématiques » – abordés d'une manière récurrente dans lesdits corpus. Prenons un exemple simple et assez évident. Dans le discours filmé d'archéologues relatant leurs travaux de recherche, parmi les sujets les plus récurrents on trouve la présentation et la description des sites archéologiques sur lesquels ils travaillent, où ils organisent et conduisent des fouilles en vue de découvrir et de sauvegarder des objets d'intérêt archéologique ou encore en vue d'étayer ou, au contraire, d'invalider telle ou telle hypothèse, telle ou telle théorie. Dans le corpus de discours filmés sur la littérature française, des sujets récurrents sont, par exemple, la présentation, la description et la mise en relation (historique, intertextuelle, etc.) de l'œuvre de tel ou tel auteur, la discussion et la présentation d'un courant littéraire dans une époque historique donnée, la discussion d'un thème ou d'un motif littéraire, l'exploration des rapports entre les différents arts et l'évolution historique des arts, la discussion des différentes approches pour traiter l'objet littéraire, et ainsi de suite.

Bien sûr, l'identification de telles séries de sujets (ou *topoï*) ne se fait pas « à l'aveugle ». Elle est toujours orientée par un cadre méthodologique qui guide les analyses et par une certaine demande (sociale, *lato sensu*). Dans notre cas concret, l'identification et la description de séries de sujets ou thèmes développés dans les

corpus audiovisuels choisis ont été réalisés à l'aide de scripts d'analyse dont nous avons déjà discuté *in extenso* structure interne et usage dans un autre ouvrage consacré à l'analyse du document audiovisuel [STO 03].

Par ailleurs, l'analyse des corpus audiovisuels à l'aide de tels scripts de travail a été réalisé en concertation avec des groupes de personnes intéressées (de *parties prenantes*) qui exprimaient une attente particulière par rapport à l'analyse de ces corpus : enseignants de littérature française, formateurs et enseignants en communication interculturelle, jeune chercheurs spécialisés dans la collecte et la préservation de patrimoines culturels immatériels, archéologues professionnels ayant en charge la préservation du patrimoine matériel d'un département français, réseau international de chercheurs s'occupant de l'enrichissement d'une vidéothèque consacrée à la documentation d'une région géopolitique, et ainsi de suite.

Une activité importante avec ces groupes de personnes était, bien sûr, celle qui consistait à identifier et à hiérarchiser les sujets (ou thèmes) les plus pertinents, les plus importants pour une partie prenante donnée. Il s'agissait, en somme, de faire une *analyse de besoins* d'information ou de connaissance auprès de et avec les parties prenantes concernées.

Dans une *deuxième étape*, les scripts décrivant les thèmes identifiés étaient soumis à une *analyse comparative*. La comparaison des scripts réalisés par les analystes permettait d'identifier – en référence au cadre théorique exposé synthétiquement au chapitre 1 – et de décrire les *traits* (thématiques, discursifs, de mis en scène visuelle) les plus récurrents, de les classer en des groupes sémantiquement homogènes pour en faire des types d'éléments, c'est-à-dire des termes conceptuels exprimant des « lieux de savoir », des *topoï* relatifs à tel ou tel type d'objets analysés. On le sait bien : il s'agit ici d'un travail délicat qui repose nécessairement sur une sorte de *principe de révision cognitive permanente* due aux limites intrinsèques et au « subjectivisme » que comporte toute catégorisation et classification.

Une *troisième étape de travail* dans la construction du méta-lexique des termes conceptuels ASA, consistait en le regroupement des termes conceptuels entre eux. Ce travail de regroupement recouvre trois points complémentaires mais bien distincts :

– le regroupement des termes conceptuels pour en dégager des types de termes conceptuels de plus en plus généraux ;
– le regroupement des termes conceptuels formant un *domaine taxinomique* spécifique (voir ci-après, section 13.3) ;
– l'identification des termes conceptuels qui forment la *base* ou encore la *base canonique* du vocabulaire des termes conceptuels des objets d'analyse de l'*univers du discours** ASA.

En ce qui concerne plus particulièrement les termes conceptuels devant former la base canonique du vocabulaire conceptuel, un des soucis principaux était de les évaluer par rapport à et de les mettre en relation avec des catégories conceptuelles déjà existantes. Dans notre cas, il s'agissait notamment de catégories approuvées et formellement définies dans les ontologies dites supérieures (en anglais, *upper-level* ou *top-level ontology*).

Ceci dit, par rapport aux ontologies supérieures, nous nous sommes toujours contentés du rôle de l'utilisateur critique des catégories conceptuelles définies dans ces ontologies sans (pouvoir ni vouloir) prendre partie dans les discussions fort abstraites et formelles qui agitent les spécialistes en la matière. L'important pour nous était – et il l'est toujours – que l'usage des catégories définies dans les ontologies supérieures nous permettait d'imposer une certaine structure aux termes ou groupes de termes conceptuels identifiés et regroupés au préalable sur la base de critères sémantiques et de faire en sorte que le *méta-lexique dans son ensemble reste interopérable* avec des terminologies et autres ontologies qui adoptent les catégories définies par les principales ontologies supérieures.

Chapitre 13
Présentation plus détaillée du vocabulaire conceptuel [Objet d'analyse]

13.1. Introduction

Ce chapitre est consacré à une présentation plus détaillée de certaines parties du vocabulaire des termes conceptuels désignant les *objets d'analyse de l'univers du discours ASA*.

La section 13.2 est consacrée à la présentation générale de la branche [Objet Endurant] qui se divise en deux branches plus spécialisées, la branche [Objet naturel] et [Objet de valeur].

Dans la section 13.3, nous reviendrons encore une fois sur l'organisation du méta-lexique. Nous discuterons plus particulièrement la notion du domaine taxinomique dont nous nous servons pour classer et structurer en petits groupes sémantiquement homogènes les termes conceptuels.

Dans la section 13.4, nous discuterons encore la question épineuse pour toute tentative de mise en place d'un cadre unique de description qui est celle de savoir comment rendre compte, à l'aide du méta-lexique des termes conceptuels, de différentes classifications (scientifiques, « populaires », pratiques, etc.) d'un domaine de connaissance.

Les sections suivantes – c'est-à-dire sections 13.5 à 13.11 – sont enfin consacrées à une présentation plus détaillée des différentes branches composant le méta-lexique des termes conceptuels des objets d'analyse de l'*univers du discours** ASA.

13.2. Les deux branches [Objet naturel] et [Objet de valeur]

La branche du vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA qui trouve sa base dans le terme conceptuel [Objet Endurant] (voir la figure 12.2 dans le chapitre 12.5), réfère à l'ensemble des objets, systèmes ou assemblages d'objets du monde « naturel » ou social. Elle réfère, pour utiliser une distinction opérée par Searle [SEA 95], d'un côté aux objets et faits à *caractéristiques intrinsèques* et de l'autre, aux objets qui possèdent leurs caractéristiques uniquement *en référence à un observateur*, qui sont, comme Greimas l'a déjà signalé depuis longtemps [GRE 76, GRE 79], des *objets de valeur* (relatifs à un *sujet*).

Ainsi, par exemple, le *diamant* possède-t-il des caractéristiques « naturelles » qui lui sont « intrinsèques », qui sont indépendantes de l'homme. Il s'agit de caractéristiques telles que le fait d'être composé de carbone, le fait de posséder une certaine structure cristalline, le fait d'être pourvu de propriétés physiques spécifiques, etc. Cependant, le fait que le diamant est un *objet très recherché dans l'industrie et la joaillerie* (donc, le fait qu'il est un *objet de valeur*), pour lequel on n'hésite pas à débourser des sommes d'argent énormes et pour l'obtention duquel on est prêt à accepter parfois les pires compromissions humanitaires, ce fait-là n'est évidemment pas intrinsèque à ce minéral. Le fait d'être un objet très recherché (parce que *très rare et « donc » socialement distinctif, beau, précieux, utile, indispensable*, etc.) qualifie le diamant comme un *objet physique à statut particulier* en référence à telle industrie ou tel groupe social[61].

[61] Notons, en passant que le terme *objet de valeur* exprime seulement le fait que l'objet joue un *certain rôle* pour le sujet-observateur, possède une *certaine fonction* pour un agent doté d'une capacité intentionnelle. Ce n'est pas, bien sûr, la rareté ou le fait d'être cher qui

Pour rendre compte de cette distinction la branche [Objet Endurant] est donc divisée en deux sous-branches principales : la sous-branche qui débute par le terme conceptuel [Objet naturel] et la sous-branche qui débute par le terme conceptuel [Objet de valeur].

Comme le montre la figure 13.1, la branche des termes conceptuels qui a comme base le terme [Objet naturel] se construit autour de deux *taxa* principaux – le taxon [Objet non-vivant] et le taxon [Objet vivant]. Le premier taxon réunit les termes conceptuels pour la description des topiques se référant au monde des objets et systèmes d'objets inanimés dont notamment (dans notre cas) aux principaux éléments naturels qui font partie de la vie de tous les jours tels que *minéraux*, *métaux*, *sédiments* mais aussi *eau*, *feu*, *air* ou *terre* (*sol*). Le deuxième taxon réunit les termes conceptuels qui nous permettent de décrire des topiques thématisant telle ou telle espèce de *flore* ou de *faune*, tel ou tel *organe physiologique*, tel ou tel *écosystème*, etc.

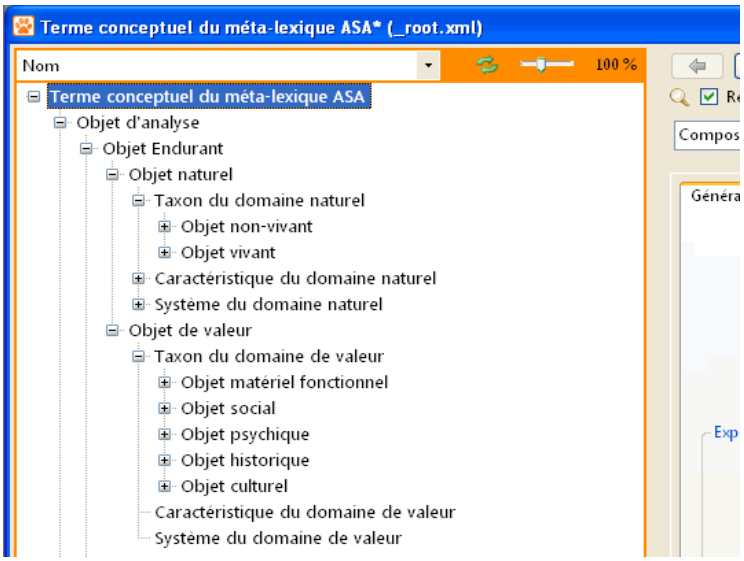

Figure 13.1. *Extrait de la branche [Endurant] avec les taxa organisant les sous-branches [Objet naturel] et [Objet de valeur]*

---

transformerait un objet matériel quelconque en un *objet de valeur* : dans sa fonction d'*être un matériau de construction*, la caillasse la plus ordinaire est vraisemblablement un *objet de valeur plus apprécié* que le diamant bleu de la Couronne de France.

Voyons maintenant la branche du méta-lexique ASA qui débute avec le terme [Objet de valeur]. Comme le montre la figure 13.1, le niveau taxinomique du terme [Objet de valeur] n'est pas réellement déployé dans la version actuelle du méta-lexique ASA. On y prévoit, comme pour *tout* domaine taxinomique de connaissance (peu importe son niveau de généralité ou de spécialisation) faisant partie de l'*univers du discours** ASA, l'éventualité d'avoir besoin, outre le terme conceptuel [Taxon du domaine de valeur], de deux autres termes conceptuels [Caractéristique du domaine de valeur] et [Système d'objets de valeur][62]. Ces deux termes conceptuels pourront être utilisés pour des descriptions très peu différenciées, très générales ne souhaitant pas ou ne pouvant pas faire des distinctions, par exemple, entre domaines sociaux, historiques ou culturels.

La branche en question prévoit, dans sa version actuelle, *cinq grands domaines taxinomiques de connaissance*[63]. Le premier domaine taxinomique de connaissance est représenté par le terme conceptuel [Objet matériel fonctionnel] et recouvre, entre autres, ce qu'on appelle la *culture matérielle* (d'un groupe social, d'une civilisation, etc.). Ce taxon réunit tous les termes conceptuels qui nous servent à rendre compte des objets physiques qui se définissent par rapport à un sujet (un observateur, un utilisateur, un professionnel, etc.). Il s'agit, par exemple, d'objets tels que *constructions*, *produits* et *outils* mais aussi réalisations et *expressions* (*de la parole*, *musicales*, etc.) ou encore possédant un *statut culturel spécifique* (tels que *monuments*, *objets de souvenir*, objets tels que *vestiges* ou *fossiles* constituant des *traces physiques d'un passé*).

Le *deuxième domaine taxinomique de connaissance* est représenté par le terme conceptuel [Objet social]. Sous ce taxon, on trouve l'ensemble des termes conceptuels qui nous servent à la description des topiques thématisant des aspects du monde social : *vie sociale*, *acteurs*, *régimes* et *systèmes sociaux*, etc.

Le *troisième domaine taxinomique de connaissance* est représenté par le terme conceptuel [Objet psychique]. Ce taxon – très peu développé, pour le moment – réunit les termes dont nous avons besoin pour décrire des domaines de connaissances thématisé dans un texte ou corpus de textes (audiovisuels) relatifs aux *comportements*, aux *attitudes*, *aux capacités cognitives*, etc.

Le *quatrième domaine taxinomique de connaissance* est représenté par le terme conceptuel [Objet historique]. Ce taxon identifie les termes conceptuels qui nous servent à décrire des sujets qui thématisent le *passé* ou telle ou telle *époque*, telle ou telle *civilisation*, tel ou tel *type d'histoire* (*histoire d'un pays*, *histoire d'un secteur d'activité*, *histoire d'un groupe social*, *histoire d'une discipline scientifique*, etc.).

---

[62] Au sujet du triplet [Taxon du domaine de valeur], [Caractéristique du domaine de valeur] et [Système d'objets de valeur], voir section 13.3

[63] Au sujet de l'expression "domaine taxinomique de connaissances", voir la section 13.3

Le *cinquième domaine taxinomique de connaissance* est représenté par le terme conceptuel [Objet culturel]. Sous ce taxon nous réunissons, enfin, les termes conceptuels qui nous permettent de décrire des topiques référant aux objets symboliques (lato sensu). Il s'agit des *langages* (dont les *langues*) et des *cultures* au sens de modèles et de cadres de référence cognitifs. Comme nous le verrons encore plus loin, nous distinguons ici, par exemple, entre cultures épistémiques (dont fait partie la culture scientifique), cultures axiologiques (dont fait partie la culture juridique), cultures délibératives (dont fait partie la culture politique), et encore cultures d'art (ici au sens d'un *« savoir- faire technique, d'une tekhnè »*).

Les cinq domaines taxinomiques en question sont construits d'une manière identique aux domaines taxinomiques référant au monde naturel. Cependant, certains se présentent d'une manière plus élaborée que d'autres. Etant donné le cadre de notre travail de recherche, ce sont notamment les domaines représentés par les termes conceptuels [Objet culturel], [Objet social] et [Objet physique fonctionnel] que nous avons pu élaborer plus en avant, tester et valider sous forme d'analyses concrètes de corpus audiovisuels.

Avant de présenter rapidement une petite sélection de domaines taxinomiques de connaissances, voyons d'abord encore quelques questions générales relatives à la terminologie et à l'organisation du vocabulaire des termes conceptuels des objets d'analyse de l'univers du discours ASA – questions qui valent d'ailleurs aussi bien pour le vocabulaire des termes conceptuels représentant les procédures d'analyse (chapitre 14) que pour les bibliothèques des schémas et des séquences de description (chapitre 16).

13.3. Questions d'organisation du méta-lexique ASA

En réalisant le travail de définition et de classification à proprement parler, nous avons du prendre en considération deux aspects complémentaires :

– la classification des termes conceptuels selon la logique *taxinomique* du *plus général au plus spécifique ;*
– le recours à des *dimensions sémantiques* (ou *classèmes*, dans la terminologie de la sémantique structurale de Greimas [GRE 67]) qui motivent et explicitent la classification taxinomique d'un ensemble de termes sous la « tête » d'un terme plus général.

Rappelons d'abord encore une fois que l'expression *terme conceptuel** signifie bien *expression (normalisée, contrôlée) d'un concept* (d'une *notion* ou d'un *thème**). Dans le méta-lexique ASA, un terme conceptuel peut être identifié comme suit :

– un terme conceptuel occupe une *position propre* dans la hiérarchie taxinomique de tous les termes conceptuels qui composent le méta-lexique ;
– un terme conceptuel fait partie de l'un des *trois éléments du triplet canonique* organisant un domaine taxinomique de connaissance (voir infra) ;
– un terme conceptuel se différencie des autres termes conceptuels du même niveau hiérarchique (et faisant partie comme lui d'un des trois éléments du triplet canonique) en référence à au moins un *classème*.

Un *canon* (ou une *base canonique*) de termes conceptuels est un ensemble de termes qui forment la base du méta-lexique ASA lui-même ou encore de tel ou tel domaine taxinomique de connaissance. Ainsi tout domaine taxinomique de connaissance comporte *en principe* le triplet de termes conceptuels suivants :

1. *[Taxon]* : terme conceptuel canonique qui permet d'exprimer le point de vue selon lesquels les objets du domaine taxinomique sont identifiés et classés ;
2. *[Caractéristique] :* terme conceptuel canonique qui permet l'identification des objets dans un domaine taxinomique donné dont l'existence dépend d'un « objet support ». Il s'agit typiquement des objets qui sont des qualités, des quantités, des attributs, des rôles, des statuts, et ainsi de suite.
3. *[Système]* : terme conceptuel canonique qui permet l'identification, dans un domaine taxinomique donné, des objets qui forment ou qui sont appréhendés sous forme de systèmes organisés, de totalités ou de simples assemblages, etc.

Ce triplet articule deux grands classèmes, deux dimensions sémantiques [RAS 87] qu'on retrouve en principe sur chaque niveau du méta-lexique des termes conceptuels ASA :

– la *première dimension* distingue entre des *objets indépendants* et des *objets dépendants* (c'est-à-dire qui, pour exister, dépendent de l'existence d'un « objet-support ») ;
– la *seconde dimension* distingue entre des objets pris comme un *tout* et, d'autre part, des entités qui forment ou qui sont perçues comme des *entités plurielles – structurées ou non, composées ou massives, etc.* (voir Brøndal et Greimas [BRG 86] qui opèrent la distinction entre *totus* et *omnis*).

Chaque niveau de spécialisation de la branche [Objet d'analyse] du méta-lexique ASA prévoit cette double distinction sémantique même si elle n'est pas systématiquement réalisée dans la version actuelle de cette branche. Par exemple, la figure 13.1 nous montre que le domaine taxinomique de connaissance initié par le terme conceptuel [Objet de valeur] ne déploie pas totalement sa base canonique. La raison en est purement pratique : nous n'avons pas eu besoin d'un domaine taxinomique si général, si peu différencié pour l'analyse de nos corpus audiovisuels.

Comme déjà dit ci-dessus, un terme conceptuel fait partie de l'un ou de l'autre des trois termes canoniques et structurants d'un domaine taxinomique de connaissance. Dans la version actuelle du méta-lexique, un terme conceptuel désigne donc :

– soit un *objet classifié selon un certain point de vue* du domaine taxinomique en question ;
– soit un *élément* (qualité, attribut, caractéristique, partie spécifique, etc.) caractéristique du domaine taxinomique ;
– soit encore un *système*, une *totalité*, une *masse* ou simplement un *ensemble (non-structuré, aléatoire, etc.)* d'objets du domaine taxinomique.

Considérons maintenant l'expression *domaine taxinomique de connaissance*. C'est un champ conceptuel (de termes conceptuels) qui possède une certaine *homogénéité sémantique* grâce au terme conceptuel qui lui sert de *tête la plus proche* de la branche dans laquelle il (c'est-à-dire le champ conceptuel en question) se trouve. Par exemple, le terme [Objet naturel] est la *tête la plus proche* du domaine taxinomique (dans le cadre du méta-lexique ASA, d'un niveau très général) dont la base canonique est le triplet [Taxon du domaine naturel], [Caractéristique du domaine naturel], [Système du domaine naturel].

Le domaine taxinomique le plus général d'objets de connaissance dans l'univers du discours ASA est celui dont la tête est le terme conceptuel [Objet d'analyse]. Dans la version actuelle du méta-lexique ASA, il n'est pas déployé complètement mais si besoin est, il est facile de le faire. En tout cas, le méta-lexique ASA forme une hiérarchie de domaines taxinomiques de connaissance des plus généraux aux plus circonscrits.

Comme déjà signalé dans le chapitre 12, la version actuelle du méta-lexique des termes conceptuels désignant les objets d'analyse de l'*univers du discours** ASA distingue quelques 85 domaines taxinomiques de connaissance. Ceux-ci forment la partie commune du méta-lexique, c'est-à-dire partagée par tous les modèles de description élaborés en vue de l'analyse de corpus audiovisuels les plus divers. Cette partie commune du méta-lexique ne peut être modifiée que par le responsable (l'administrateur ou le comité de pilotage) du méta-lexique ASA. Cependant, chaque domaine taxinomique de connaissance peut être complété par des nouveaux termes conceptuels et/ou de nouveaux domaines taxinomiques de connaissance. Ces ajouts forment la partie modifiable et propre au « concept-designer » (ingénieur de connaissances) de l'archive pour laquelle il travaille.

Concrètement parlant, si par exemple un anthropologue-chercheur travaillant sur l'héritage culturel immatériel d'un peuple ou d'une région développe ses propres modèles de description (c'est-à-dire sa propre bibliothèque de modèles de description), il/elle :

– s'appuie (entre autres) sur le méta-lexique des termes conceptuels désignant les objets d'analyse de l'univers du discours ASA et identifie ceux dont il/elle a besoin pour définir la ou les structures topiques* relatives à l'univers du discours de son archive audiovisuelle ;

– si, dans un domaine taxinomique de connaissance déjà existant, un terme conceptuel est absent ou ne figure pas, le chercheur l'ajoute sous forme d'un terme conceptuel spécifique à un groupe d'utilisateurs ;

– si un domaine taxinomique de connaissance manque dans le méta-lexique ASA, le chercheur l'ajoute sous forme d'un *nouvel objet d'analyse** (c'est-à-dire [Objet …]) propre à « son » groupe d'utilisateurs – objet d'analyse qui se déploie sous forme du triplet canonique [Taxon de …], [Caractéristique de …], [Système de …].

Soulignons encore qu'il ne faut pas confondre *domaine taxinomique de connaissance* et *structure topique** (ni, *a fortiori*, *thématique** ; voir chapitre 5.3) définissant un domaine de connaissance qui peut être thématisé (d'une manière ou d'une autre) dans un texte ou un corpus de textes audiovisuels. Dans le premier cas, il s'agit, comme nous venons de le voir, d'un champ constitué de termes conceptuels qui possède une certaine *homogénéité sémantique* grâce au terme conceptuel qui lui sert de *tête la plus proche* dans la branche dans laquelle se trouve le champ conceptuel en question. Dans le deuxième cas, il s'agit d'une configuration qui modélise un domaine de connaissance et, ce faisant, sélectionne des termes conceptuels qui lui sont pertinents – peu importe s'ils proviennent d'un ou de plusieurs domaines taxinomiques du méta-lexique ASA. Nous l'avons déjà vu à plusieurs reprises dans les chapitres précédents de ce livre, les termes conceptuels sélectionnés dans une *configuration topique** proviennent souvent de *domaines taxinomiques différents* et entretiennent entre eux des *relations autres que simplement hiérarchiques* au sens taxinomique du terme.

Revenons encore une fois sur l'organisation du vocabulaire des termes conceptuels ASA. Celle-ci repose essentiellement sur la relation de la *dépendance unilatérale* entre termes conceptuels composant le vocabulaire et sur celle de la *dépendance réciproque* entre ceux-ci. La première relation se manifeste dans le positionnement entre termes conceptuels généraux et termes conceptuels spécialisés (ou encore entre termes conceptuels possédant une *densité* sémantique variable). On parle ici de termes conceptuels *hyperonymes* et de termes conceptuels *hyponymes* (l'*hyperonyme* est le terme conceptuel plus général, l'*hyponyme* est le terme conceptuel plus spécifique). La deuxième relation instaure le positionnement entre deux ou plusieurs termes conceptuels qui dépendent tous d'un même terme conceptuel et qui sont donc partiellement similaires (ou, si l'on veut, contraires). Cette deuxième relation permet de distinguer entre des termes conceptuels

antonymes ou, plutôt, *contraires* (« contraire » au sens d'une *opposition graduelle* ; à distinguer de la *contradiction* qui est une opposition d'*exclusion réciproque*).

Enfin, comme déjà dit ci-dessus, outre les deux relations de base de construction de tout vocabulaire hiérarchique de termes conceptuels (c'est-à-dire celle de la *spécialisation hyperonymique/hyponymique* et celle de l'*opposition graduelle* ou *exclusive*), nous utilisons également un ensemble d'autres catégories sémantiques appelées *classèmes* par les sémanticiens structuralistes comme Greimas [GRE 66] pour identifier et distinguer des groupes ou des sous-ensembles sémantiques de termes conceptuels parmi l'ensemble des termes qui composent le méta-lexique ASA. Une catégorie spécifique de ces classèmes est composée par des classèmes que F. Rastier appelle *dimension sémantique* [RAS 87] – type de catégories qui forment une petite *classe fermée* d'oppositions (graduelles).

Sans vouloir ouvrir ici une discussion trop complexe (touchant au problème du statut des classèmes et d'une théorie de classèmes) mentionnons seulement quelques-unes de ces dimensions sémantiques qui jouent un rôle dans l'organisation du méta-lexique des termes conceptuels ASA.

Nous avons déjà introduit et discutées ci-dessus les deux dimensions sémantiques tout à fait capitales pour l'organisation de notre méta-lexique. Il s'agit d'une part de la dimension sémantique de la *dépendance/indépendance existentielle* des objets et d'autre part de celle de la distinction entre *objets pluriels ou singuliers* (c'est-à-dire « singulier » au sens de « unitaire »). Parmi les autres dimensions sémantiques qui jouent un rôle plus ou moins important dans l'identification (ou la circonscription) de sous-ensembles de termes conceptuels ASA nous comptons notamment les trois suivantes :

– la *dimension du support* physique selon lequel se distingue les uns des autres les objets d'analyse. Ainsi identifions nous un sous-ensemble de termes conceptuels qui désignent des actions somatiques ou des actions mentales (= objets d'analyse désignés par des termes conceptuels appartenant à la branche [Objet Perdurant]), des objets physiques, des objets sociaux ou des objets mentaux (= objets d'analyse désignés par des termes conceptuels appartenant à la branche [Objet Endurant]) ou encore des espaces géométriques, des espaces géographiques ou des espaces représentés (= objets d'analyse désignés par des termes conceptuels appartenant à la branche [Objet Région]). Un autre sous-ensemble de termes conceptuels est réservé pour désigner des objets qui diffèrent du point de vue de leur *spécificité perceptive*. Ainsi, nous pouvons identifier et distinguer, par exemple, entre actions gestuelles et actions visuelles (= objets d'analyse désignés par des termes conceptuels appartenant à la branche [Objet Perdurant]), entre objets visuels et objets acoustiques (= objets d'analyse désignés par des termes conceptuels appartenant à la branche [Objet Endurant]) ;

– la *dimension de la fonction* que peuvent occuper les objets d'analyse. Grâce à cette dimension sémantique, nous identifions les termes conceptuels qui réfèrent à des objets d'analyse qui jouent un *rôle spécifique* dans un domaine d'analyse donné. Nous identifions ainsi des termes conceptuels qui réfèrent aux activités d'extraction, de production et de construction, aux activités de transformation et de communication ou encore aux activités de consommation (= objets d'analyse désignés par des termes conceptuels appartenant à la branche [Objet Perdurant]). Nous pouvons également identifier et désigner des objets à valeur qui possèdent un statut spécifique pour un acteur (des objets de mémoire, par exemple, ou encore des objets fétiche) et des objets qui servent à un acteur et qui constituent pour lui des ressources, des outils, des instruments, etc. (= objets d'analyse désignés par des termes conceptuels appartenant à la branche [Objet Endurant]) ;

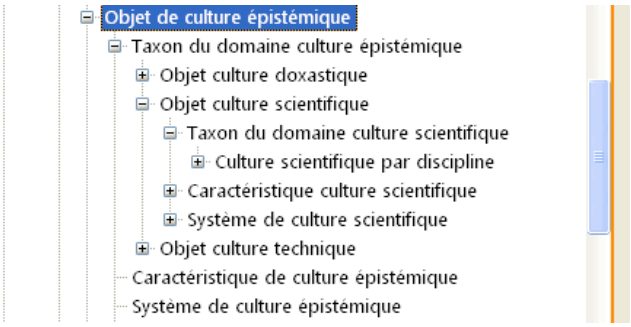

Figure 13.2. *Extrait de la branche du méta-lexique ASA représentant le domaine taxinomique [Culture épistémique]*

– la *dimension modale* (au sens de Greimas [GRE 83]) des objets d'analyse (dont notamment les objets de valeur). Cette dimension réunit un sous-ensemble de termes conceptuels qui réfèrent à des objets d'analyse possédant un *statut épistémique spécifique* (comme, par exemple, les objets et les espaces de fiction, les époques imaginaires, etc.). Elle permet également de circonscrire, dans le méta-lexique ASA, un sous-ensemble de termes conceptuels qui réfèrent aux activités, entités ou régions pourvues d'un *statut moral* (ou *éthique*) *défini*. D'autres sous-ensembles de termes, identifiables grâce à cette dimension sémantique, sont réservés pour désigner des activités ou entités pourvu d'un statut *esthétique*, d'un statut *déontique* ou, enfin, d'un statut *phorique* (affectif, émotionnel).

13.4. Comment tenir compte de classifications différentes ?

Revenons encore une fois sur notre présentation des deux domaines taxinomiques généraux [Objet non-vivant] et [Objet vivant]. Comme on le sait bien,

il existe des classifications très divergentes des éléments ou entités qui peuplent le monde inanimé ou animé.

Certaines classifications font référence à des standards scientifiques, d'autres à des connaissances pratiques séculaires, à des croyances populaires ou à des objectifs pratiques et utilitaires. Tenir compte de ces différents points de vue est une chose délicate mais, en principe, possible dans le cadre du vocabulaire des termes conceptuels ASA. Ainsi, l'appellation même, les principes et les objectifs d'un système de classification spécifique peuvent être traités grâce à et dans le domaine taxinomique [Culture épistémique] du méta-lexique ASA – domaine taxinomique qui fait partie de la branche [Objet de valeur].

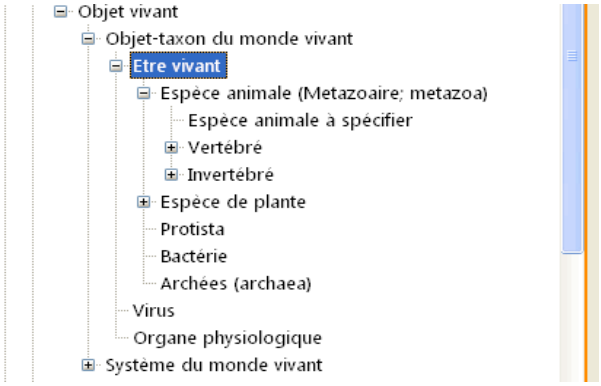

Figure 13.3. *Extrait de la branche du méta-lexique ASA représentant le domaine taxinomique [Etre vivant]*

Comme le montre la figure 13.3, l'objet-taxon de ce domaine permet de catégoriser une classification (de la faune, de la flore ou du monde non-vivant) comme étant, selon l'analyste, une classification scientifique (qui se base sur des recherches biologiques ou géologiques, selon le cas), une classification faisant appel à la doxa (aux connaissances tacites, traditionnelles, etc.) d'un groupe ou une classification pratique et utilitaire faisant appel à une culture technique. Par contre, il faut utiliser le domaine taxinomique [Etre vivant] pour pouvoir décrire les éléments (vivants) classés en types d'éléments par telle ou telle classification (scientifique, populaire ou autre).

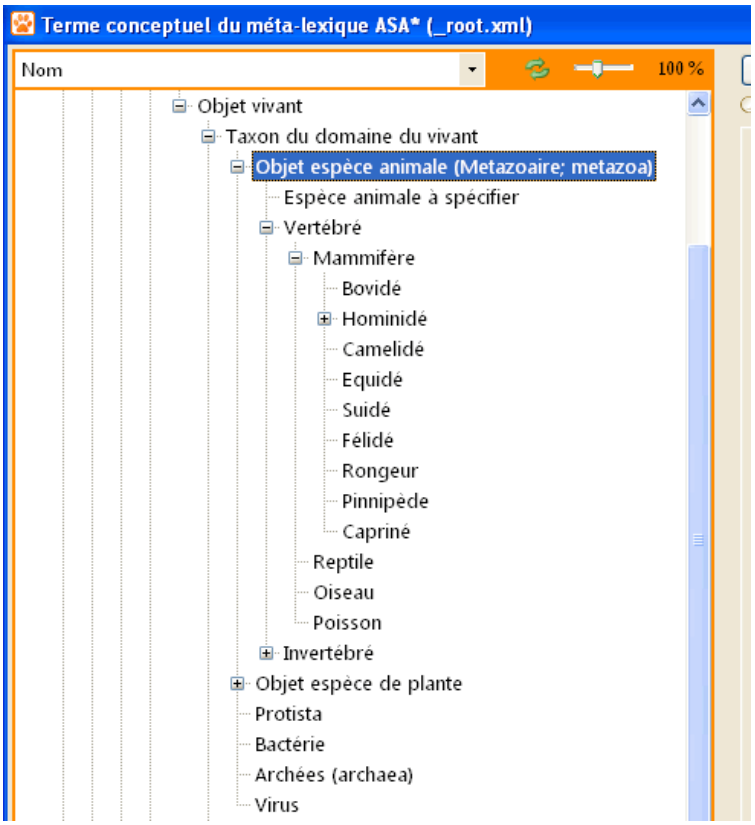

Figure 13.4. *Extrait de la branche du méta-lexique ASA représentant le domaine taxinomique [Objet espèce animal]*

La figure 13.4 montre un extrait plus développé du domaine taxinomique [Etre vivant] constitué, entre autres, par la branche [Objet espèce animale]. Choix et organisation des termes conceptuels peuplant la branche [Objet espèce animal] montrent que nous avons suivi une classification assez standard du monde animal. Toujours est-il que les termes conceptuels retenus ni leur organisation ne satisferont pas nécessairement à la spécificité référentielle de telle ou telle classification.

Dans son état actuel, la branche [Objet espèce animale] ne pourra pas satisfaire aux exigences pointues d'une description de topiques qui thématisent, par exemple, la classification cladistique du vivant. Elle ne pourra pas non plus satisfaire à la description systématique et détaillée de telle ou telle ethno-classification de la faune ou de la flore. Pour remédier à cela, deux possibilités se proposent à nous.

La première est de créer un nouveau domaine taxinomique de connaissance à l'intérieur de la branche [Etre vivant] (voir figure 13.3) qui correspond aux exigences d'un certain type de classification. Ce nouveau domaine conceptuel peut soit faire partie du vocabulaire commun du méta-lexique des termes conceptuels ASA, soit constituer un *module propre* à un *utilisateur* ou *groupe d'utilisateurs* (voir nos explications ci-dessus, section 13.3).

Un tel module pourrait d'abord être déclaré comme *ajout ontologique au méta-lexique ASA*, nécessaire pour couvrir d'une manière satisfaisante l'univers de discours des archives d'une communauté donnée. Ensuite, pourrait être assigné au module en question une place précise dans l'économie générale du méta-lexique ASA. Dans notre exemple, ce serait un « fils » direct du terme conceptuel de base [Etre vivant]. A l'heure actuelle, cette option d'ajouts ontologiques sous forme de modules utilisateurs n'est pas encore réalisée au niveau informatique.

Une deuxième option que nous avons développée plus en avant dans notre travail de recherche est celle de l'utilisation de *micro-thesaurus* pour satisfaire aux besoins divergents d'identifier, de décrire et de classer des faits naturels ou sociaux. Nous en avons discuté rapidement un exemple d'usage de deux micro-thesaurus différents pour décrire et classer des faits culturels : l'un des deux micro-thesaurus est une version simplifiée du thesaurus de l'UNESCO portant sur le domaine de la culture[64], l'autre a été élaboré dans le cadre du projet ASA-SHS. Rien n'empêche, bien sûr d'ajouter d'autres micro-thesaurus.

13.5. Le domaine conceptuel représenté par le terme [Objet matériel fonctionnel]

La figure 13.5 nous montre que le terme conceptuel [Objet matériel fonctionnel] forme un des taxa qui spécifient le terme conceptuel [Objet de valeur]. Son domaine taxinomique est composé du *triplet canonique* [Taxon du domaine matériel fonctionnel], [Caractéristique matérielle fonctionnelle] et [Système matériel fonctionnel].

Le domaine taxinomique représenté par ce triplet de termes conceptuels est le plus général de la branche [Objet matériel fonctionnel] : les caractéristiques identifiées à ce niveau (via le terme conceptuel [Caractéristique matérielle fonctionnelle]) ainsi que les systèmes ou ensembles (via le terme conceptuel [Système matériel fonctionnel] sont donc invariants ou non-spécifiques à tel ou tel domaine taxinomique plus spécialisé et dont la figure 13.6 nous montre l'organisation générale. Ainsi au niveau du domaine taxinomique de connaissance [Objet matériel fonctionnel] (figure 13.5), on trouvera des termes conceptuels tels que [Fonction] (de tel ou tel objet matériel), [Statut] (de tel ou tel objet matériel),

[64] Voir http://databases.unesco.org/thesfr/

[Degré de technicité] (de tel ou tel objet matériel), etc. Du côté [Système matériel fonctionnel], on trouvera, par exemple, les termes conceptuels [Culture matérielle] et [Patrimoine matériel] qui permettent la description générale et non-spécifique de systèmes ou d'ensembles d'objets matériels en relation avec tel groupe social, telle communauté, telle époque, telle civilisation, et ainsi de suite.

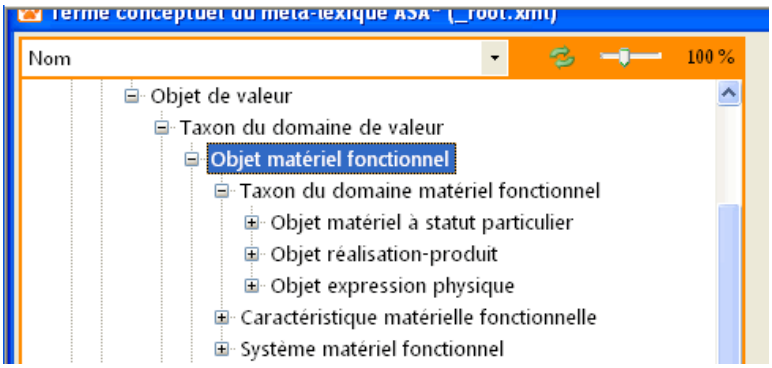

Figure 13.5. *Le domaine conceptuel [Objet matériel fonctionnel]*

Comme le montre la figure 13.5, la branche débutant avec le terme conceptuel [Objet matériel fonctionnel] se différencie en trois domaines taxinomiques plus spécialisés que nous avons élaborés en référence à la fois à notre corpus de travail et à toute une série de thesaurus spécialisés[65] :

[65] Car, comme le dit fort justement Bruno Bachimont [BAC 05], tout n'est pas dans le texte ; il faut donc repérer et utiliser d'autres sources d'information pour bâtir un méta-lexique conceptuel qui « tient la route » … Ainsi, toute une série de thesaurus, terminologies, glossaires et autres vocabulaires nous ont été utiles pour mettre en place la version actuelle du vocabulaire conceptuel composant le domaine taxinomique dont la tête est constituée par le terme conceptuel [Objet matériel fonctionnel]. Citons à titre d'exemple seulement l'*AAT (Art & Architecture Thesaurus)* du Getty Research Institute (http://www.getty.edu/research/tools/vocabularies/), le *Materials Thesaurus* du British Museum (http://www.collectionslink.org.uk/assets/thesaurus_bmm/matintro.htm), le thesaurus *PACTOLS* du Centre Camille Jullien du CNRS (http://biaa.univ-provence.fr/ListRecord.htm?list=table&table=5), le thesaurus de la désignation des œuvres architecturales et mobilières de la Direction de l'Architecture et du Patrimoine du Ministère (français) de la Culture et de la Communication (http://www.culture.gouv.fr/culture/inventai/patrimoine/), etc.

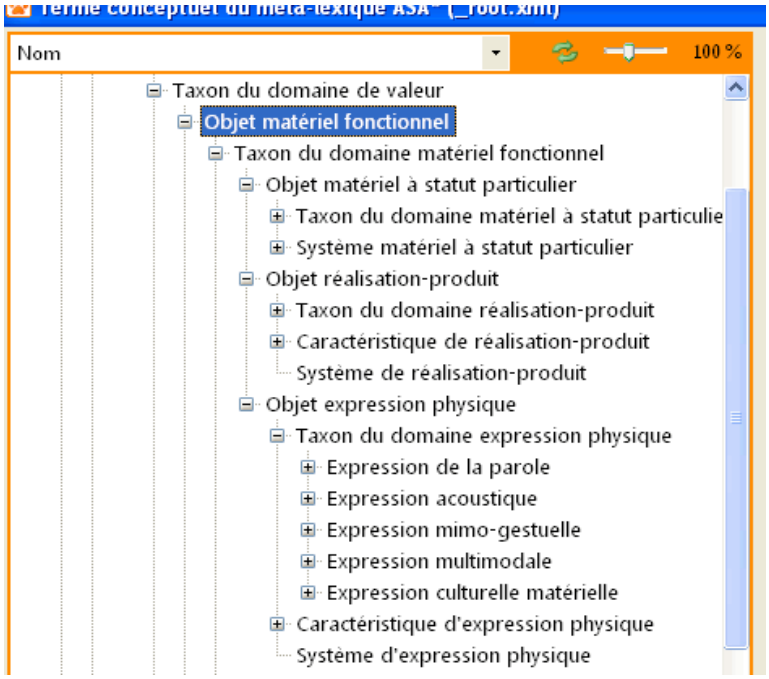

Figure 13.6. *La différenciation de la branche [Objet matériel fonctionnel] en domaines taxinomiques de connaissance plus spécialisés*

– le domaine taxinomique des *objets physiques – non-vivants ou vivants – qui occupent une position particulière* dans le monde de vie d'un acteur (voir figure 13.5 la branche débutant avec le terme [Objet matériel à statut particulier]. Il s'agit aussi bien de personnes physiques, des populations (humaines ou autres) que d'animaux (domestiques, de production, etc.) ou encore d'objets et de collections d'objets de toute sorte documentant un passé ou permettant de reconstruire un passé. Le vocabulaire conceptuel composant cette branche spécialisée réunit donc ici tous les termes conceptuels qui permettent pour analyser des objets (ou ensemble d'objets) matériels possédant une valeur pour l'archéologie, la préhistoire ou la paléontologie. On y trouve, en outre, les termes conceptuels qui réfèrent à des objets physiques à valeur culturelle (historique, mémorielle, etc.) spécifiques dont notamment les monuments et les documents au sens de *pièce* matérielle d'information à caractère historique ou autre ;

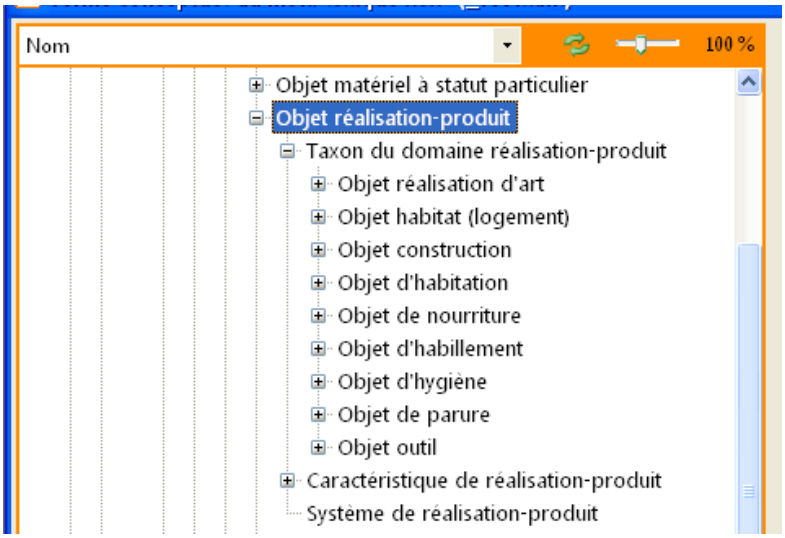

Figure 13.7. *Le domaine taxinomique des artefacts représenté par le terme conceptuel [Réalisation-produit]*

– le domaine taxinomique des *artefacts*, c'est-à-dire des réalisations et des produits tels que les réalisations d'art, les constructions et les habitats, la nourriture, les vêtements, etc. Ce domaine débute avec le terme conceptuel [Objet réalisation-produit]. Etant donné son importance dans l'analyse de corpus audiovisuels mettant en scène les objets qui font partie des pratiques quotidiennes (se loger, se nourrir, se vêtir, se soigner, se présenter, etc.), ce domaine est complètement déployé à l'aide du triplet canonique [Taxon du domaine réalisation-produit], [Caractéristique de réalisation-produit] et, enfin, [Système de réalisation-produit]. Le domaine taxinomique en question se différencie en des domaines taxinomiques plus spécialisés, plus circonscrits tels que celui des aliments, celui de l'habitat, celui de l'habillement ou encore celui de la construction (voir figure 13.7). A leur tour, chacun de ces domaines taxinomiques spécialisés peut, si nécessaire et utile, se différencier en des domaines taxinomiques encore plus spécialisés. Ainsi est-il tout à fait possible d'intégrer dans le méta-lexique ASA, sous forme de termes conceptuels plus spécialisés, des distinctions conceptuelles fines opérées par exemple par des thesaurus très spécialisés tels que celui de la désignation des œuvres architecturales et mobilières de la *Direction de l'Architecture et du Patrimoine du Ministère (français) de la Culture et de la Communication*[66] ;

– enfin, le domaine taxinomique des *expressions physiques* – expressions verbales ou écrites, expressions musicales, expressions audiovisuelles, expressions

[66] Voir http://www.culture.gouv.fr/culture/inventai/patrimoine/

gestuelles, etc. (voir figure 13.6). Ce domaine qui débute par le terme conceptuel [Objet expression physique], joue un rôle prépondérant dans la description et classification des corpus audiovisuels documentant, par exemple, des échanges verbaux, des réalisations musicales, des environnements sonores ou encore des actions et programmes d'action.

13.6. Le domaine conceptuel représenté par le terme [Objet social]

Considérons maintenant la branche du méta-lexique ASA qui débute avec le terme conceptuel [Objet social]. Celle-ci recouvre l'ensemble du vocabulaire des termes conceptuels que nous utilisons pour décrire des endurants sociaux, c'est-à-dire – *grosso modo* – des entités, objets ainsi que des ensembles d'objets du monde social.

Le vocabulaire des termes conceptuels de cette branche permet, dans son état actuel, de décrire et d'analyser des thèmes ou topiques parlant de personnes morales, d'institutions de tout type, de peuples, de groupes sociaux, de mouvements sociaux, de modes de vie, de régimes sociaux, etc.

La figure 13.8 nous montre l'organisation de ce domaine taxinomique complexe qui est composé de toute une variété de domaines plus spécialisés. Au niveau taxinomique de base du domaine en question, nous trouvons notre habituel triplet canonique [Taxon du monde social] différenciant le dit domaine en une série de domaines plus spécialisés, [Caractéristique du monde social] et [Système social]. Ce dernier terme conceptuel est particulièrement utile dans la mesure où il permet la description de topiques dans lesquelles est question, par exemple, de faits sociaux généraux et souvent assez flous tels que la *société industrielle*, la *société de l'information*, la *société précapitaliste*, etc. – des sujets qui sont aujourd'hui assez prisés et récurrents en sciences humaines et sociales.

Le terme conceptuel [Taxon du monde social] (figure 13.8) se différencie, dans la version actuelle du méta-lexique ASA, en quatre principaux domaines taxinomiques. Le premier domaine conceptuel s'organise autour du terme [Objet vie sociale]. Celui-ci permet de tenir compte des topiques traitant des questions relatives, par exemple, à la vie rurale ou urbaine, à la vie économique et politique, à la vie personnelle et privée, etc.

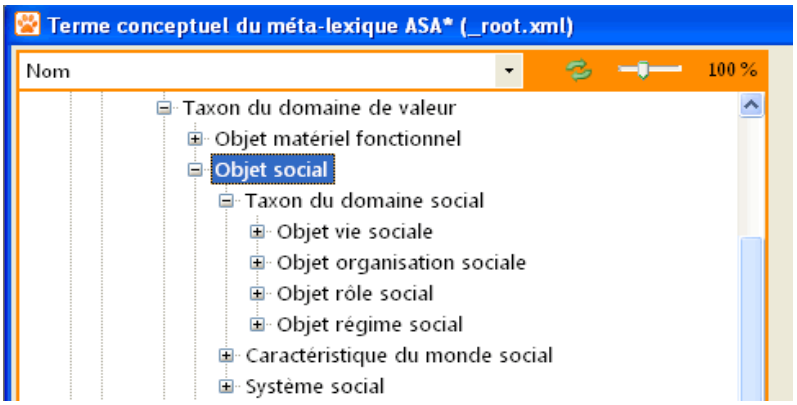

Figure 13.8. *Le domaine taxinomique des objets du monde social représenté par le terme conceptuel [Objet social]*

Le deuxième domaine taxinomique représenté par le terme [Objet organisation sociale] (figure 13.9) recouvre d'abord les différents types de regroupements sociaux (tels que *groupe social*, *mouvement social*, *peuple* (et *ethnie*), *nation*, *communauté* (*ethnique*, *religieuse*, *linguistique*, etc.), *clan* ou *famille*. Il recouvre ensuite divers types d'*organisations institutionnalisées* : *institutions* organisant la vie dans les différents *secteurs d'activités* d'un acteur (d'un groupe social) ; *institutions « traditionnelles »*, *communautaires* et *modernes* ; *institutions* exerçant leur influence sur un *territoire* circonscrit ; *institutions* dotées d'un *pouvoir* (décisionnaire, coercitif, d'influence, etc.). Enfin, un troisième type d'objets d'analyse est identifié par le terme conceptuel [Objet état (politique] qui renvoie aux différentes formes d'organisation politique de la vie sociale.

Le troisième domaine taxinomique qui dépend directement du terme conceptuel [Taxon du domaine monde social] est représenté par le terme [Objet rôle social] (figure 13.10). Ce terme conceptuel réunit tous les termes qui permettent d'analyser les sujets ou topiques qui traitent d'abord les rôles définis dans leur appartenance à un métier ou à une profession.

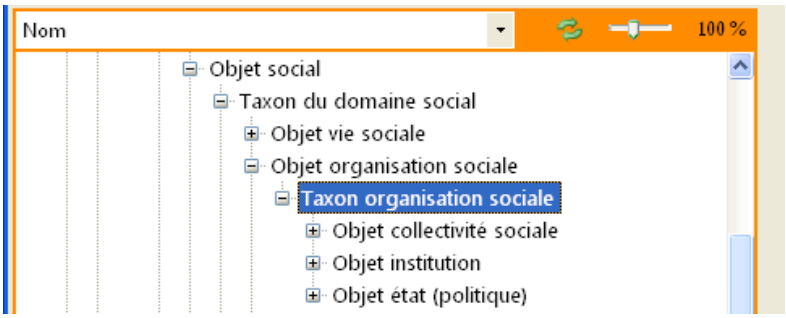

Figure 13.9. *Le domaine taxinomique des objets du monde social représenté par le terme conceptuel [Forme d'organisation sociale]*

Il prévoit également le vocabulaire conceptuel nécessaire pour analyser des rôles sociaux généraux qui engage la cohérence, l'organisation et aussi l'auto-compréhension d'un groupe social dans son ensemble.

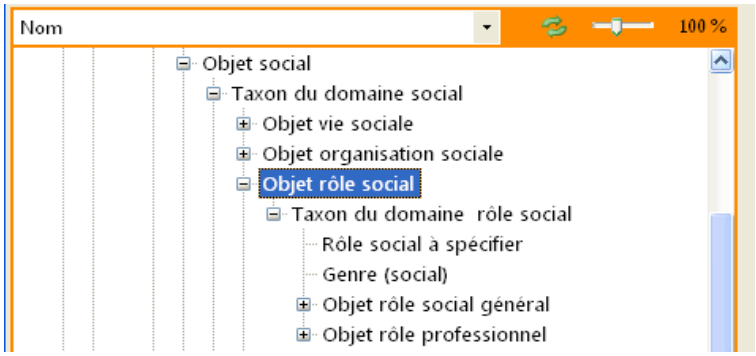

Figure 13.10. *Le domaine taxinomique des objets du monde social représenté par le terme conceptuel [Rôle social]*

C'est le cas lorsque l'on parle, par exemple, d'une *personnalité*, d'une *autorité*, d'un *étranger*, d'un *« leader »* ou encore d'une *victime*, d'un *exclu*, etc. Un troisième objet de connaissance – peu développé dans la version actuelle du méta-lexique ASA – est celui des rôles sociaux liés au sexe qu'on désigne parfois par l'expression *« genre »* (à ne pas confondre, bien évidemment avec son homonyme « genre de discours »).

Enfin, le quatrième domaine taxinomique qui dépend du terme [Taxon du monde social] (figure 13.8), est représenté par le terme conceptuel [Régime social]. Sous ce terme, nous réunissons le vocabulaire dont nous avons besoin pour décrire les textes

audiovisuels qui traitent – seul ou en relation avec d'autres sujets – des questions relatives aux régimes politiques, économiques, religieux, etc. dans tel ou tel pays, telle ou telle région du monde, telle ou telle époque historique.

13.7. Le domaine conceptuel représenté par le terme [Objet culturel]

Etant donné la spécificité référentielle de nos corpus de travail, le domaine taxinomique initié par le terme [Objet culturel] est, dans la version actuelle du méta-lexique ASA, le domaine taxinomique le plus élaboré de la branche [Objet Endurant].

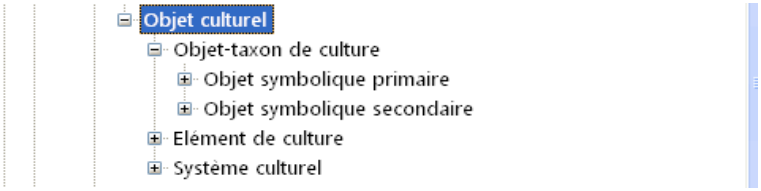

Figure 13.11. *Le domaine taxinomique représenté par le terme conceptuel [Objet culturel]*

Comme le montre la figure 13.11, le méta-lexique ASA distingue entre deux domaines taxinomiques principaux : le premier s'organise autour du terme conceptuel [Objet symbolique primaire] et le second autour du terme [Objet symbolique secondaire]. Nous avons emprunté la distinction entre monde symbolique primaire et monde symbolique secondaire (tout en l'adaptant à notre besoin spécifique) à J. Lotman [LOT 72] (voir aussi [TOR 99], éminent sémioticien des cultures, pour pouvoir réserver une place particulière aux faits culturels qui, d'une manière ou d'une autre, présupposent et (ré-)utilisent le langage (dont la langue naturelle) à leurs fins spécifiques.

Ainsi, comme le montre la figure 13.12, le domaine taxinomique qui dépend du terme [Objet symbolique primaire] se différencie, pour le moment, en quatre domaines taxinomiques plus spécialisés représentés, chacun, par les termes conceptuels de base [Objet langue], [Objet discours], [Objet texte] et [Objet écriture].

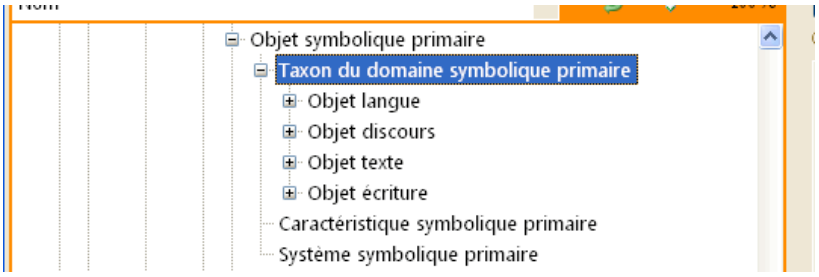

Figure 13.12. *Le domaine taxinomique représenté par le terme conceptuel [Objet symbolique primaire]*

Il s'agit ici davantage d'une liste ouverte que d'une véritable classification de domaines plus spécialisés du monde du langage *lato sensu*, c'est-à-dire du monde des systèmes de signes élaborés par l'homme (et d'une manière plus ou moins rudimentaire par d'autres espèces vivantes) pour exprimer, communiquer, partager, transmettre, conserver et aussi – bien sûr – utiliser des valeurs et des connaissances relatives à lui-même, le monde et l'au-delà transcendant.

L'identification et l'élaboration des domaines taxinomiques composant le terme [Taxon du domaine symbolique primaire] ont été dictées – encore une fois – par les corpus de textes audiovisuels avec lesquels nous avons travaillé pendant ces dernières années. Néanmoins le caractère modulaire du méta-lexique s'appuyant sur la notion centrale du domaine taxinomique et de sa définition « de base » à l'aide de « notre » triplet canonique de termes conceptuels, facilitera la tâche de transformer cette liste en une forme de classification plus élaborée.

Cette transformation devra tenir compte, sans aucun doute, d'une réflexion plus poussée sur l'organisation des différents systèmes de signes dont dispose un agent pour communiquer sur des faits qui l'entourent.

La figure 13.13 nous montre les principaux domaines taxinomiques composant le monde des systèmes symboliques qui présupposent le langage *lato sensu*. Pour le moment, nous distinguons trois grands groupes taxinomiques :

– le premier groupe de domaines taxinomiques de nature *transversale* (c'est-à-dire dont certaines caractéristiques se retrouvent d'une manière ou d'une autre dans tous les autres domaines taxinomiques) est composé par les trois domaines qui s'organisent autour des termes [Objet épistémique], [Objet axiologique] et [Objet de croyance (de transcendance)] ;

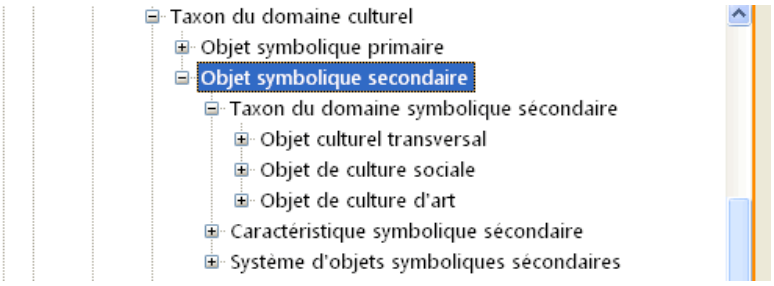

Figure 13.13. *Les trois groupes de domaines taxinomiques composant la branche [Objet symbolique secondaire]*

– le deuxième groupe réunit les domaines taxinomiques permettant de décrire des sujets relatifs aux connaissances, visions, idées, règles, normes, etc. qui façonnent et/ou qui répondent à des problèmes et interrogations relevant du monde social ou d'un de ses champs spécifiques (c'est le cas, par exemple, des domaines taxinomiques qui servent à décrire des sujets relevant de la culture politique, de la culture éducative, de la culture économique, etc.) ;

– enfin, le troisième groupe est composé des domaines taxinomiques qui nous servent à décrire des sujets relatifs à ce qu'on appelle les arts et la littérature.

13.8. Domaines taxinomiques faisant partie de la branche [Objet symbolique primaire]

Voyons maintenant rapidement les domaines taxinomiques qui spécifient le terme conceptuel [Objet symbolique primaire]. Parmi les quatre domaines conceptuels spécifiques faisant partie de l'univers du discours ASA, les trois domaines [Objet langue], [Objet discours] et [Objet texte] sont particulièrement importants pour notre travail de recherche portant sur l'analyse de corpus audiovisuels :

– le domaine [Objet langue] parce qu'il constitue le domaine de référence pour un grand nombre de textes audiovisuels soumis à l'analyse dans le cadre du projet ASA-SHS (ce sont souvent des entretiens et des conférences de chercheurs consacrés, par exemple, à une ou plusieurs familles de langues, à une langue particulière, à telle ou telle structure linguistique, à tel ou tel phénomène sociolinguistique, etc.) ;

– les domaines [Objet discours] et [Objet texte] parce qu'ils renvoient au contexte interne et propre à toute thématisation d'un objet de connaissance (telle qu'une langue ou une famille de langues) : le discours en tant

qu' « outil », en tant que moyen de l'auteur pour parler et communiquer sur un objet de référence, le texte en tant qu'outil d'enregistrement et d'organisation formelle et physique du discours (de l'auteur) sur un support donné.

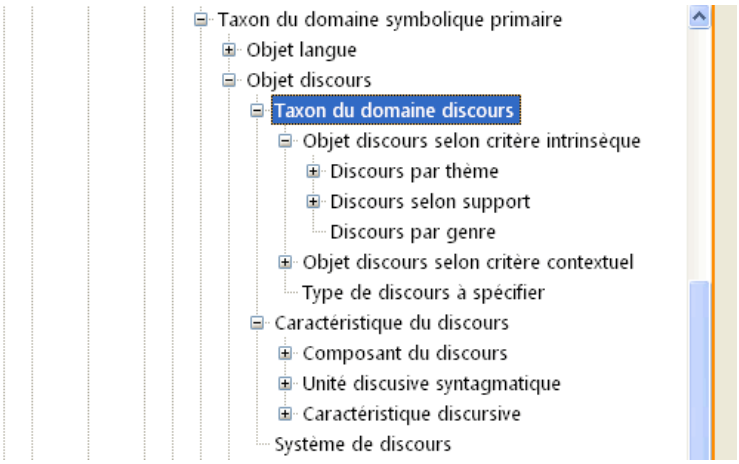

Figure 13.14. *Le domaine taxinomique représenté par le terme conceptuel [Objet discours]*

Le domaine taxinomique représenté par le terme [Objet discours] possède, par ailleurs, une importance particulière pour l'approche développée ici de la description et indexation du contenu d'un texte, d'un corpus ou de toute une archive audiovisuelle. Il réunit en effet le vocabulaire conceptuel dont nous avons besoin pour analyser d'une part le type de discours (descriptif, narratif, explicatif, polémique, etc.) employé par un auteur pour « parler » d'un sujet donné et d'autre part les caractéristiques discursives de la thématisation d'un sujet – caractéristiques telles que le *propos* d'un sujet, le *thème* (l'isotopie thématique) *principale*, le *point de vue* de l'auteur, l'*hiérarchisation thématique* dans le discours, etc. Autrement dit, on trouve ici catégorisée cette classe spécifique d'objets d'analyse que nous avons appelée les *objets discursifs* (voir chapitres 5 et 7).

La figure 13.14 montre un extrait du domaine taxinomique qui s'organise autour du terme conceptuel de base [Objet discours]. Pour le moment, nous nous sommes surtout concentrés à l'identification des principaux types de discours et à un ensemble récurrent de caractéristiques discursives dont nous avons eu besoin pour nos analyses concrètes. Ceci étant, pour structurer ce domaine taxinomique sous forme d'un vocabulaire élémentaire de termes conceptuels, nous nous sommes inspirés aussi bien des recherches en rhétorique et en analyse du discours (voir par exemple [JAC 99, MAC 02, MAI 96]) que des différentes terminologies et ontologies descriptives consacrées plus particulièrement l'objet « discours ».

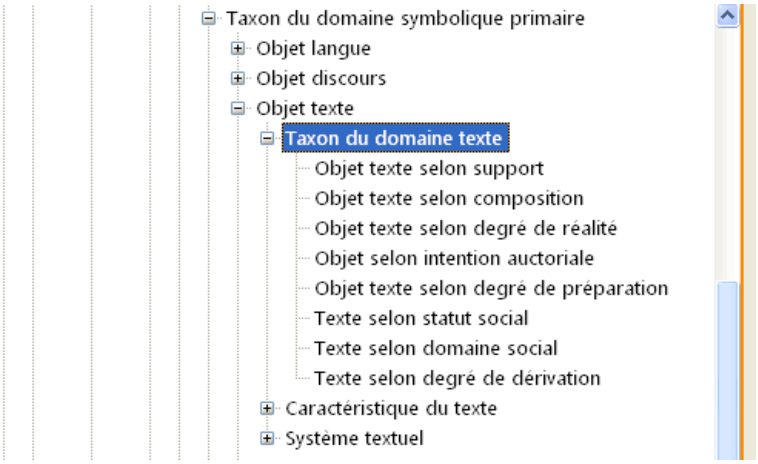

Figure 13.15. *Le domaine taxinomique représenté par le terme conceptuel [Objet texte]*

Le deuxième domaine taxinomique central est, comme déjà dit, celui qui s'organise autour du terme conceptuel [Objet texte]. La figure 13.15 montre un extrait de son organisation et du vocabulaire élémentaire que nous utilisons actuellement pour rendre compte des différents objets de la segmentation et de la description du texte audiovisuel lui-même. Comme nous l'avons vu dans le chapitre 2, la segmentation d'un texte audiovisuel sert à la localisation et à l'identification des passages les plus pertinents par rapport aux objectifs de l'analyste.

La description du texte audiovisuel peut concerner le texte dans son ensemble (par exemple, un enregistrement d'une conférence, un documentaire, etc.). Elle ne peut concerner que tel ou tel passage du texte audiovisuel (par exemple, un ou plusieurs extraits d'une conférence, une séquence dans un documentaire, etc.). Enfin, elle peut concerner un ensemble de textes et/ou passages textuels faisant partie d'un corpus audiovisuel (par exemple des extraits provenant de plusieurs conférences et séminaires consacrés à un sujet précis).

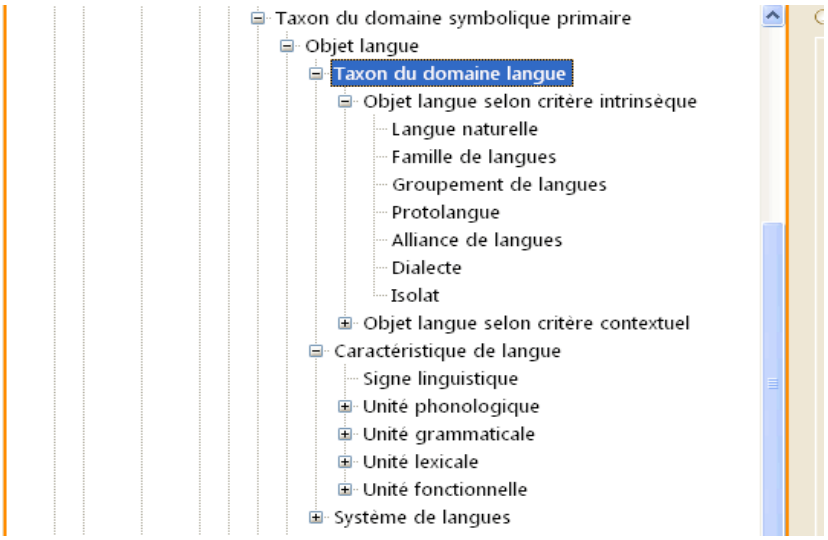

Figure 13.16. *Le domaine conceptuel représenté par le terme conceptuel [Objet langue]*

Les recommandations de la *Text Encoding Initiative* (T.E.I.)[67] nous ont servi de références centrales pour l'élaboration de ce domaine taxinomique dans la mesure où elles s'avéraient pertinentes pour l'analyse d'un corpus de textes audiovisuels. C'est le cas, par exemple, des entités conceptuelles telles que *texte unitaire* ou *groupement de textes*, *corps du texte* et *annexe du texte* (par exemple dans le cadre de discussions filmées qui suivent une conférence ou une séance de séminaire), *découpage du texte en segments* (*paragraphe*, *région*) ou encore *interprétation du texte* (sous forme de commentaires appartenant à des types de commentaires spécifiques).

Enfin, encore quelques mots au sujet du domaine taxinomique [Objet langue] (figure 13.16) qui est assez souvent thématisé dans le corpus composant les deux archives expérimentales ARC et ALIA. Les principaux termes conceptuels représentant le domaine taxinomique [Langue] ont été identifiés et définis, dans la mesure du possible, en référence à l'ontologie GOLD (*General Ontology for Linguistic Description*)[68] qui est une des ontologies de référence pour l'analyse linguistique à proprement parler.

Uniquement un petit sous-ensemble de l'ontologie GOLD est effectivement intégré dans la version actuelle du méta-lexique ASA. En effet, l'analyse des

[67] Pour plus d'informations relatives au codage électronique de textes à l'aide du TEI, voir http://www.tei-c.org/Guidelines/Customization/Lite/teiu5_fr.html
[68] Voir http://linguistics-ontology.org/version

langues et des usages linguistiques n'ont pas été l'objectif de notre travail. L'objectif principal a été plutôt l'analyse des sujets relatifs au *travail linguistique* – sujets abordés, par exemple, dans des entretiens, des conférences et de colloques filmés par le programme des *Archives Audiovisuelles de la Recherche* (AAR)[69]. Une telle analyse ne rentre pas dans les détails de la description d'une langue ou d'une famille de langues ni dans ceux des usages concrets d'une langue par ses locuteurs.

Par contre, la place des termes principaux de GOLD étant identifiée dans la structure taxinomique du méta-lexique ASA, il n'est pas très compliqué d'intégrer le reste de cette ontologie afin de la rendre disponible aux analystes intéressés.

Autrement dit, l'Atelier de Description ASA pourrait être parfaitement bien utilisé pour une description systématique et fine (c'est-à-dire correspondant au niveau de la précision de l'ontologie GOLD) de corpus audiovisuels linguistiques : analyses conversationnelles, lexico-grammaticales, phonétiques, etc.

13.9. Domaines taxinomiques faisant partie de la branche [Objet symbolique secondaire]

Le domaine taxinomique s'organisant à partir du terme conceptuel [Objet symbolique secondaire] recouvre toute une série de domaines taxinomiques plus spécialisés permettant d'analyser des sujets relatifs, par exemple, aux croyances et aux religions, à la recherche scientifique, à la littérature, à la musique et aux arts visuels, aux idéologies et aux visions de nature politique, au savoir et savoir-faire techniques ou encore aux systèmes de valeurs, aux normes ou règles régissant, par exemple, un groupe social.

Il s'agit ici de domaines empiriques centraux thématisés dans les corpus audiovisuels qui nous ont servi comme objet d'analyse. Comme déjà expliqué ci-auparavant (voir section 13.4 et figure 13.13), les termes faisant partie du vocabulaire de la branche [Objet symbolique secondaire] sont regroupés, en trois groupes centraux (voir figure 13.17) formant ou susceptibles de former des domaines taxinomiques plus spécifiques.

[69] Url du portail : http://www.archivesaudiovisuelles.fr/FR/

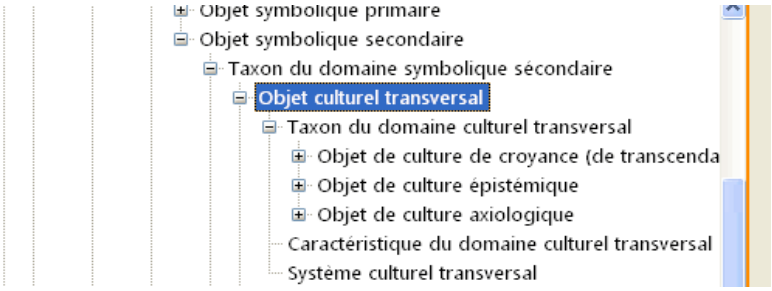

Figure 13.17. *Le domaine taxinomique dont la base est le terme conceptuel [Objet culturel transversal]*

Ainsi, un *premier groupe* réunit tous les termes conceptuels dont nous avons besoin pour décrire des objets de connaissance relatifs, par exemple, aux croyances, aux idéologies, aux pensées de groupe, aux valeurs et normes, aux coutumes et traditions mais aussi aux savoir et savoir-faire dont, notamment, le savoir et savoir-faire scientifique.

Un *deuxième groupe* est composé par l'ensemble des termes dont nous servons pour décrire ce qu'on appelle parfois les *cultures sociales*, c'est-à-dire les *cadres de référence* (au sens de Taylor [TAY 98] ; voir aussi [STO 11c]) définissant les identités et spécificités (cognitives, axiologiques, normatives, etc.) des différents domaines et secteurs de la vie sociale comme, par exemple, ceux de l'éducation, de la santé, du politique, de l'économie, et ainsi de suite.

Enfin, un *troisième groupe* des termes conceptuels nous sert à traiter la thématisation d'objets de connaissance relatifs non pas aux œuvres artistiques ou littéraires mais bien relatifs aux cadres de référence, aux cultures artistiques, littéraires ou encore musicales.

Voyons rapidement quelques exemples du *premier* et du *troisième groupe* ; le deuxième groupe ne se présentant actuellement que sous une forme rudimentaire, nous préférons le traiter ultérieurement.

Comme le montre la figure 13.17, le premier groupe de termes conceptuels faisant partie de la grande branche [Objet symbolique secondaire] réunit les domaines taxinomiques qui se forment à partir des termes de base [Objet de culture épistémique], [Objet de culture axiologique] et [Objet de culture de croyance (de transcendance)] :

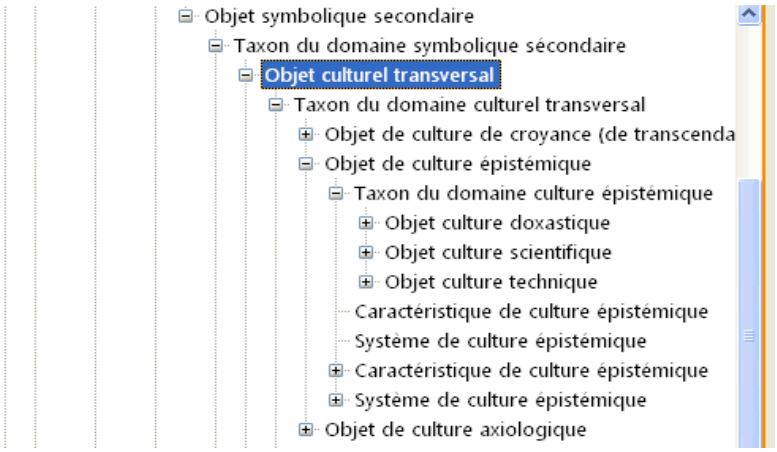

Figure 13.18. *Le domaine taxinomique dont la base est le terme conceptuel [Objet de culture épistémique]*

– le vocabulaire qui dépend du terme conceptuel de base [Objet de culture épistémique] permet la description des systèmes de connaissance *scientifique* ou non, *doxastique* (c'est-à-dire définissant les évidences et les vérités pour un acteur social) ou encore *technique* ;

– le vocabulaire qui dépend du terme conceptuel de base [Objet de culture axiologique] sert, comme l'expression « axiologie » le suggère déjà, à la description des valeurs (collectives ou personnelles, générales ou spécifiques à une pratique, à un champ social, etc.) dont, notamment, les normes, règles (explicites ou tacitement admises) et autres lois (juridiques au sens du droit contemporain ou pré-juridiques, traditionnelles, coutumières, sacrées, etc.). Une catégorie de termes conceptuels quelque peu à part (selon von Wright [WRI 63]) mais incluse dans ce vocabulaire des termes conceptuels est celle qui permet la description des règles dites techniques, c'est-à-dire des consignes et de procédés. La figure 13.18 nous montre ce groupe qui s'organise autour des trois domaines taxinomiques principaux que sont [Objet de culture doxastique], [Objet de culture scientifique] et [Objet de culture technique] ;

– enfin, le vocabulaire qui dépend du terme conceptuel de base [Objet de culture croyance (de transcendance)] permet la description des sujets consacrés non seulement aux systèmes et mouvements religieux mais aussi aux croyances dites populaires, à toute forme de croyance « spontanée » de groupe ou encore à cette catégorie de croyances hétérodoxes qu'on taxe, selon le cas, de croyance « de secte », d'hérésie, etc.

On voit bien que les frontières entre ces trois domaines d'analyse sont floues. Après tout, une doxa a naturellement tendance à s'ériger en une norme pour tous ceux qui y croient (« croire » ici au sens d'un rapport fiduciaire, d'une relation de confiance dite primaire, spontanée ou non-réfléchie [GRE 79, RIC 83]) et qui en font un dogme au sens quasi-religieux. Pour saisir ces formes cognitives floues dans ce sens qu'elles mêlent d'une manière quasiment indissociable des aspects aussi bien épistémiques qu'axiologiques ou encore fiduciaire (c'est-à-dire de croyance), l'analyste peut les décrire comme des systèmes symboliques secondaires (sans les spécifier davantage par rapport à tel ou tel domaine conceptuel plus circonscrit).

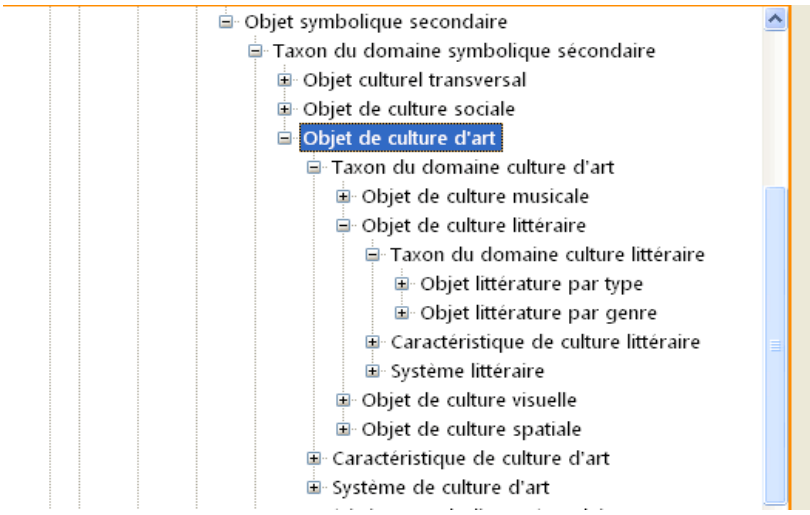

Figure 13.19. *Le domaine taxinomique dont la base est le terme conceptuel [Objet de culture d'art]*

Considérons maintenant la figure 13.19. Celle-ci nous montre un extrait de la branche dont le terme de base est [Objet de culture d'art]. Ce terme ouvre le domaine taxinomique du troisième groupe identifié ci-dessus composant, ensemble avec le groupe d'objets culturels transversaux et le groupe d'objets culturels du monde social, la branche [Objet symbolique secondaire]. Pour le moment, nous avons distingué, à l'intérieur de ce troisième groupe, les domaines taxinomiques plus spécifiques composés de termes conceptuels à l'aide desquels nous pouvons analyser des sujets audiovisuels consacrés aux cultures musicales, littéraires, visuelles et architecturales.

Nous sommes tout à fait conscients qu'il s'agit ici, pour le moment, d'une simple *liste* de domaines taxinomiques relatifs à l'analyse de sujets audiovisuels consacrés à l'art et à la littérature. L'identification et l'élaboration de ces domaines correspondent aux spécificités des corpus audiovisuels sur lesquels nous avons

travaillé. Des analyses sur des corpus audiovisuels plus variés nous montreront les limites de cette organisation actuelle et donc aussi les modifications nécessaires à y apporter.

13.10. Les domaines taxinomiques de la branche [Objet perdurant]

La deuxième grande branche composant la base canonique du méta-lexique ASA qui nous sert pour définit nos objets d'analyse, est celle qui débute par le terme conceptuel [Objet Perdurant]. Comme déjà expliqué au chapitre 11 (section 11.3), cette branche réunit l'ensemble du vocabulaire conceptuel dont nous avons besoin pour décrire des sujets relatifs aux actions, activités, processus, changements ou encore aux phénomènes qui – justement – perdurent davantage dans le temps comme c'est le cas des situations, des états ou des équilibres et déséquilibres.

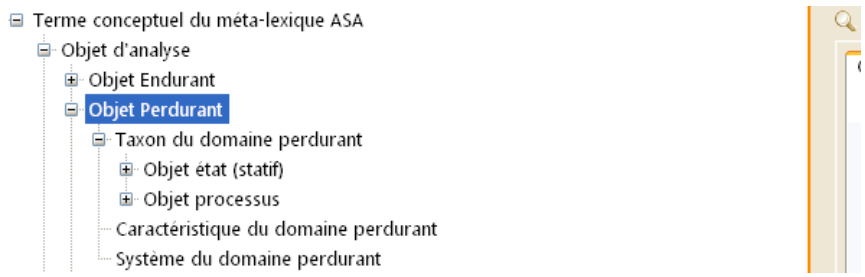

Figure 13.20. *La base de la branche débutant avec le terme conceptuel [Objet Perdurant]-terme canonique du méta-lexique ASA*

La figure 13.20 nous montre les deux domaines taxinomiques de base qui composent la branche [Objet perdurant]. Il s'agit d'une part du domaine taxinomique initié par le terme conceptuel [Objet état] et, de l'autre, du domaine taxinomique initié par le terme conceptuel [Objet processus].

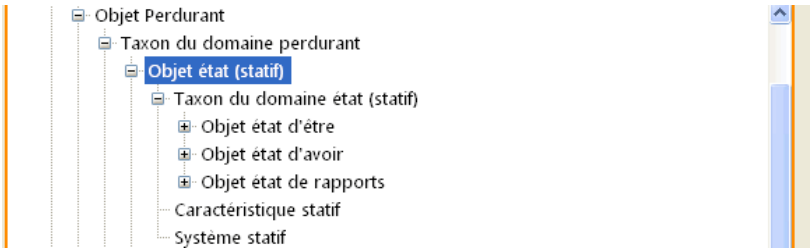

Figure 13.21. *Organisation taxinomique du domaine initié par le terme conceptuel [Objet état (statif)]*

Les objets de connaissance pouvant être représentés et décrits par le vocabulaire conceptuel dont le terme de base est [Objet état] sont des phénomènes qui se distinguent par une forme de *maintien* ou de *maintenance* (qualitative ou quantitative) *dans le temps*. La figure 13.21 montre les trois principaux domaines taxinomiques spécifiant le terme [Objet état (statif)].

Le premier domaine taxinomique s'organise autour du terme conceptuel [Objet état d'être]. Sous cette dénomination quelque peu barbare, on trouve un petit vocabulaire conceptuel spécialisé que nous utilisons pour décrire, par exemple, un état psychique ou un statut social (tel que celui d'*être citoyen* ou *d'être étranger*). D'une manière plus générale, il s'agit donc ici de pouvoir décrire des sujets thématisant un *rapport d'identité* ou *d'identification* (la *jonction*, pour reprendre ici un terme utilisé par Greimas [79], entre un agent (ou un objet) et une *qualité*, une *caractéristique*, un *attribut*, voir un sentiment ou une sensation).

Le terme conceptuel [Objet état d'avoir] initie un deuxième domaine taxinomique spécifiant la branche [Objet état]. Il s'agit ici typiquement de rapports de possession économique ou autre comme par exemple de rapports de richesse ou de pauvreté – sujets fortement développés dans nos corpus audiovisuels consacrés à la mondialisation et à l'immigration.

Enfin, le troisième domaine taxinomique spécifiant le terme conceptuel de base [Objet état] réunit un vocabulaire spécialisé permettant de traiter plus particulièrement des situations dites intersubjectives. Il s'agit, par exemple, de situations consensuelles (paix, entente), de situations polémiques (conflit, crise, blocage), asymétriques (domination, oppression) ou niées, c'est-à-dire refusées (exclusion et négation symbolique ou physique de l'autre).

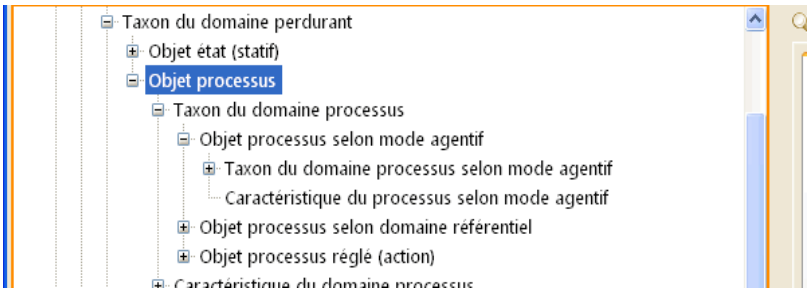

Figure 13.21. *Organisation taxinomique du domaine initié par le terme conceptuel [Objet processus]*

La deuxième grande branche qui constitue le terme canonique [Objet Perdurant] est initiée, comme le montre la figure 13.20, par le terme conceptuel [Objet processus]. Cette branche se différencie, encore une fois, en trois domaines

taxinomiques plus spécifiques mais très inégalement développés dans la version actuelle du méta-lexique ASA.

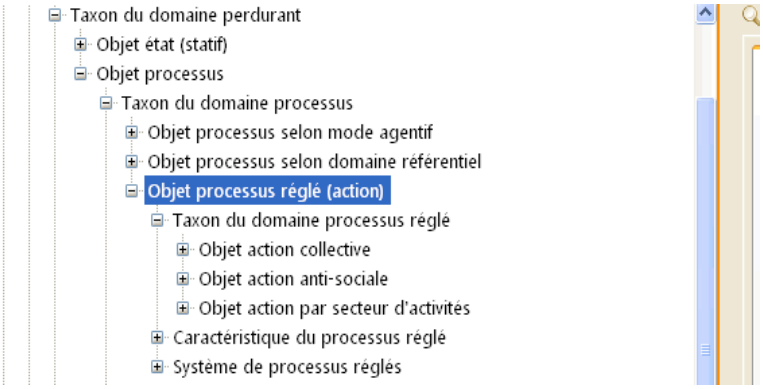

Figure 13.22. *Organisation taxinomique du domaine initié par le terme conceptuel [Objet processus réglé (action)]*

Le premier domaine taxinomique débute par le terme [Objet processus selon mode agentif] et nous permet de distinguer entre des processus causaux et intentionnels. Cette distinction, archi-utilisée en lexicologie, nous sert à identifier d'une part des processus de nature physique ou biologique et d'autre part, des activités dont l'origine – l' « auteur » – est un agent quelconque doté d'une certaine intentionnalité (c'est-à-dire capable d'agir de façon préméditée pour satisfaire un besoin, un désir ou tout autre motif d'action).

Le deuxième domaine taxinomique – débutant par le terme conceptuel [Processus selon domaine référentiel] – réunit le vocabulaire conceptuel nécessaire pour décrire notamment des sujets traitant d'activités plus spécialisées telles que, par exemple, les activités concrètes qui ponctuent la vie quotidienne, la vie privée ou encore la vie professionnelle. Il s'agit ici d'un immense champ empirique, difficile à organiser mais omniprésent en tant que thème visuel ou thème de discours dans tout type de corpus audiovisuels.

Enfin, le troisième domaine taxinomique qui spécifie le terme conceptuel [Objet processus] s'organise autour du terme conceptuel [Objet processus réglé]. Sous cette dénomination, nous réunissons notamment les pratiques individuelles ou collectives, professionnelles ou non qui forment des complexes d'activités se déroulant d'une manière stéréotypée, c'est-à-dire sous forme de routines ou de rituels plus ou moins prévisibles. La figure 13.22 nous montre un extrait de ce troisième domaine taxinomique qui joue en effet un rôle important dans la description du contenu de nos corpus audiovisuels. Le vocabulaire conceptuel réuni ici nous sert à décrire, par exemple, des fêtes, des rituels, des pratiques artistiques et artisanales, des pratiques

de communication, des pratiques d'éducation, des routines de travail agricole ou industriel, et ainsi de suite.

13.11. Les domaines taxinomiques de la branche [Objet région]

Voyons encore rapidement l'organisation taxinomique de la branche initiée par le troisième terme conceptuel canonique du méta-lexique ASA des termes conceptuels désignant les *objets d'analyse** qui font partie de l'*univers du discours** des archives ASA. Il s'agit du terme conceptuel canonique [Objet Région].

Comme le montre la figure 13.23, nous distinguons ici deux grands domaines taxinomiques : le premier est initié par le terme conceptuel [Objet région spatiale], le second par le terme conceptuel [Objet région temporelle].

La branche qui débute par le terme conceptuel [Objet région spatiale] se divise, comme le montre la figure 13.24, en une série de domaines taxinomiques qui attendent encore une meilleure structuration. Pour le moment, nous y trouvons des vocabulaires conceptuels spécialisés qui nous sont fort utiles notamment pour l'activité de la description dont l'objectif principal est de localiser, dans l'espace et le temps, les objets de connaissance thématisés dans un texte ou un corpus de textes audiovisuels.

Figure 13.23. *Organisation générale de la branche débutant par le terme conceptuel canonique [Objet Région]*

Nous trouvons ici, par exemple, le vocabulaire conceptuel dont nous avons besoin pour identifier et décrire une *région géopolitique*, un *lieu de travail*, l'*espace urbain ou rural*, etc.

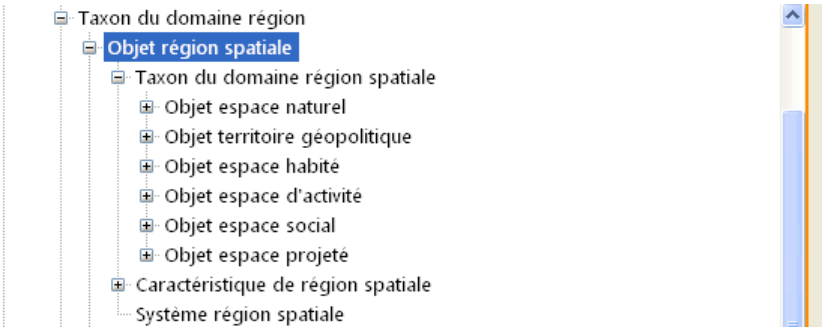

Figure 13.24. *Les domaines taxinomiques actuellement identifiés qui spécifient le terme conceptuel [Objet région spatiale]*

Enfin, la figure 13.25 nous montre l'actuelle organisation taxinomique de la branche initiée par le terme conceptuel [Objet région temporelle]. Nous distinguons, pour le moment et d'une manière provisoire, entre deux grands groupes de domaines taxinomiques.

Le premier groupe est réservé à la définition du vocabulaire conceptuel dont nous avons besoin pour identifier et décrire des *objets temporels de base* avec ou sans durée tels que, par exemple, les *instants* et les *périodes* (temporelles) voire, d'une manière plus concrète, des *unités temporelles* telles que les *jours*, les *mois*, les *années*, les *décennies* les *siècles*.

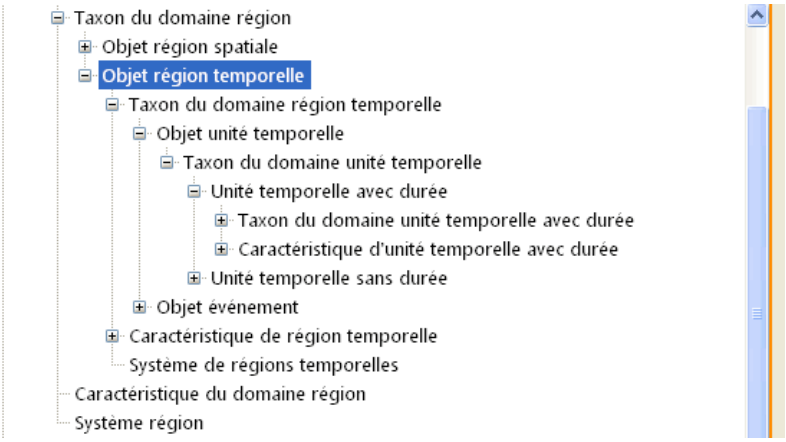

Figure 13.25. *Les domaines taxinomiques actuellement identifiés qui spécifient le terme conceptuel [Objet région temporelle]*

Le deuxième groupe est réservé à la description des *événements* : des événements qui ponctuent la vie biologique ou sociale (naissance, moments de passage, mort), l'évolution historique, l'identité personnelle ou collective ou encore la révérence à et l'union avec une transcendance.

Chapitre 14

Le méta-lexique des activités d'analyse de l'objet textuel

14.1 Introduction

Après avoir discuté le méta-lexique réservé au vocabulaire conceptuel pour décrire les *objets d'analyse** qui font partie de l'*univers du discours**, voyons maintenant celui qui identifie les *activités d'analyse* dont nous nous servons pour décrire, annoter et indexer un texte ou un corpus de textes audiovisuels.

Comme nous le développerons dans la section suivante (section 14.2), nous nous sommes concentrés sur quatre classes particulières d'activités d'analyse qui forment,

ensemble, un *deuxième méta-lexique* de termes conceptuels, c'est-à-dire une ressource métalinguistique à part entière dont se sert l'ingénieur de connaissances – le « concept designer » – pour spécifier et développer les *modèles de description** qui composent la bibliothèque de modèles de description de l'*univers du discours** propre à une archive audiovisuelle donnée. L'analyste se sert de ces procédures pour réaliser son travail d'analyse ou d'expertise sur un texte ou un corpus de textes audiovisuels.

Dans les sections 14.2 à 14.6, nous introduirons les principales classes de procédures d'analyse sur lesquelles nous nous appuyons pour traiter un corpus audiovisuel.

Toute activité d'analyse définie dans le méta-lexique réservé à l'identification des activités de description est expliquée par un ou plusieurs schémas d'indexation – un schéma d'indexation étant un ou un ensemble de champs spécialisés que l'analyste doit ou peut saisir dans son analyse d'un texte ou d'un corpus de textes audiovisuels. Dans la section 14.7, nous discuterons les rapports de définition entre un ou plusieurs schémas d'indexation et un terme conceptuel identifiant une activité (ou plutôt, un type spécifique d'activités) de description. Ensuite, nous présenterons encore très rapidement la bibliothèque des schémas d'indexation que nous avons pu élaborer dans le cadre du projet ASA-SHS.

14.2. Quatre classes d'activités d'analyse textuelle

Souvenons-nous de notre discussion dans le chapitre 13 : tout *modèle de description** (accessible à l'analyste via une interface de travail spécialisé) est composé d'un ou de plusieurs *séquences de description**, chaque séquence de description étant fonctionnellement spécialisée dans l'élucidation progressive du *sujet**, du thème d'un discours audiovisuel (c'est-à-dire d'un discours exprimé et communiqué essentiellement par l'image et le son). Une séquence, à son tour, est composée d'au moins deux *schémas de définition**, c'est-à-dire de deux petites configurations génériques dont l'une spécifie l'*objet d'analyse** dans une *séquence** de description et l'autre les *procédures d'analyse**.

Nous connaissons deux procédures de base pour analyser un texte ou un corpus de textes audiovisuels : la *procédure dite libre** et la *procédure dite contrôlée**. Dans la procédure dite libre, l'analyste « remplit » (comme l'appellation « description libre l'indique déjà) *librement* les champs composant son interface de travail. Ces champs dépendent directement de la structure du modèle de description utilisé et auxquels l'analyste est invité à répondre pour produire une description concrète de son objet d'analyse. Dans le cas de la procédure dite contrôlée, c'est un *thesaurus** (ou, plutôt, un *micro-thesaurus** interprétant le rang des valeurs possibles pour un *terme conceptuel** ou une configuration de termes conceptuels) qui offre à l'analyste la possibilité de choisir, soit une, soit plusieurs valeurs (c'est-à-

dire expressions normalisées ou descripteurs) représentant au mieux la spécificité de son objet d'analyse – d'où l'expression « description contrôlée ». Tout au long de ce livre, nous avons vu des cas concrets illustrant ces deux procédures (voir, par exemple, les chapitres 3, 9 et 10).

Ceci dit, les principaux outils d'analyse que nous avons élaborés et que nous utilisons actuellement pour travailler concrètement sur le contenu d'un texte ou d'un corpus audiovisuel, se divisent comme suit :

– *outils linguistiques (lato sensu)* groupés autour de l'explicitation, la description et l'indexation du contenu d'un texte audiovisuel ou d'un passage dans un texte audiovisuel ;
– *outils audiovisuels* groupés autour de l'explication/illustration du contenu à l'aide d'expressions visuelles ou acoustiques telles que icônes, *emoticons*, *jingles*, etc. qui acquièrent ainsi, dans ce cas concret, le statut d'une donnée métalinguistique décrivant (ou plutôt : illustrant) un objet d'analyse ;
– *outils* groupés autour de la *traduction/adaptation*, explicitation, etc. d'un objet textuel dans une langue cible différente de celle ayant servie à produire l'objet en question ;
– *outils* groupés autour de *l'usage de thesaurus*, socio-taxinomies (« folksonomies »), glossaires, terminologies (personnelles, institutionnelles, etc.) pour identifier et classer l'objet textuel ;
– *outils* servant à rapprocher une analyse ASA à un *standard* ou une norme (tel que LOMFR, Dublin Core, etc.) ;
– *outils* (d'analyse critique) permettant de *commenter*, critiquer, améliorer, enrichir, etc. un contenu.

Avec les outils servant à la segmentation des passages audiovisuels pertinents et avec ceux de la publication à proprement parler d'un corpus audiovisuel analysé et indexé, les outils d'analyse que nous venons de citer, constituent un arsenal riche et diversifié permettant, à notre avis, une appropriation satisfaisante d'un objet audiovisuel quelconque pour en faire une ressource *sui generis* à destination d'un public spécifique (voir nos explications dans le chapitre 1).

La figure 14.1 nous montre l'organisation générale du méta-lexique qui débute par le terme conceptuel primitif [Procédure d'analyse]. Ce méta-lexique est divisé en quatre grandes classes taxinomiques *de procédures d'analyse*.

Une première classe de procédures recouvre toutes les activités qui nous permettent de traiter les différents aspects d'un texte audiovisuel ou d'un corpus de textes audiovisuels. Comme déjà expliqué dans le chapitre 1, nous considérons le texte audiovisuel comme étant constitué d'un ensemble de plans (ou strates) caractéristiques tels que celui de la thématisation référentielle d'un objet de connaissance, celui du développement syntagmatique d'un objet thématisé, celui du cadrage discursif et audiovisuel, etc.

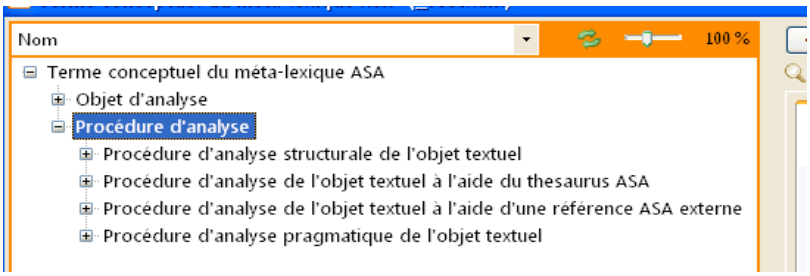

Figure 14.1. *Les quatre classes d'activités d'analyse d'un texte ou d'un corpus de textes audiovisuels*

Outre le fait qu'un texte audiovisuel est un objet constitué d'un ensemble de strates caractéristiques, c'est aussi un objet qui peut être décomposé en des unités plus petites. Il s'agit, d'une part, des *unités textuelles générales* telles que l'unité [Segment], l'unité [Transition] ou l'unité narrative [Séquence] et, d'autre part, des *unités textuelles nommées et typées* (c'est-à-dire représentant un genre spécifique d'unités) telle que l'unité [Générique (d'un film)], [Insert], etc.

Or, l'analyse de ces différentes strates et la possibilité de décomposer un texte audiovisuel en des unités plus petites, est la tâche principale à laquelle doit répondre cette première classe d'activités d'analyse représentée par le terme conceptuel [Procédure d'analyse structurale de l'objet textuel].

Une deuxième classe de procédures (figure 14.1) est représentée par le terme conceptuel de base [Procédure d'analyse de l'objet textuel à l'aide du thesaurus ASA]. Elle réunit toutes les activités d'analyse qui s'exécutent à l'aide d'un ou de plusieurs micro-thesaurus faisant partie du thesaurus commun ASA (voir aussi les chapitres 10 et 15).

La classe suivante – la troisième (figure 14.1) représentée par le terme conceptuel [Procédure d'analyse de l'objet textuel à l'aide d'une référence ASA externe] – est réservée à toutes les activités d'analyse qui font appel à des thesaurus, terminologies, ontologies ou encore standards qui sont externes au système ASA. En somme, il s'agit ici de permettre à l'analyste, s'il le souhaite, de procéder à des analyses d'un corpus audiovisuel non pas avec le thesaurus commun ASA mais avec un thesaurus, une terminologie de son propre choix.

Enfin, la quatrième classe de procédures (figure 14.1) réunit toutes les activités d'analyse qui permet d'adapter l'objet textuel à un public (à un destinataire) ou encore à un usage particulier. Elle débute par le terme conceptuel de base [Procédure d'analyse pragmatique de l'objet textuel].

Dans la prochaine section, nous discuterons d'une manière plus détaillée ces quatre grandes classes qui sont développées, à l'heure actuelle, d'une manière assez inégale. Ce sont les activités des deux premières classes – [Procédure d'analyse structurale de l'objet textuel] et Procédure d'analyse de l'objet textuel à l'aide du thesaurus ASA] – qui sont les plus avancées. En ce qui concerne les activités de la troisième classe – [Procédure d'analyse de l'objet textuel à l'aide d'une référence ASA externe], nous nous sommes contentés de quelques expérimentations ponctuelles de faisabilité pratique que d'un développement réel de tout un arsenal d'activités plus spécifiques. Enfin, la quatrième classe n'a pas constitué un enjeu central de recherche et développement dans le cadre du projet ASA-SHS.

14.3. La classe des activités [Procédure d'analyse structurale de l'objet textuel]

Les deux figures 14.2 et 14.3 nous montrent un extrait du vocabulaire conceptuel identifiant les principaux types d'activités d'analyse de l'objet textuel en tant qu'unité structurale.

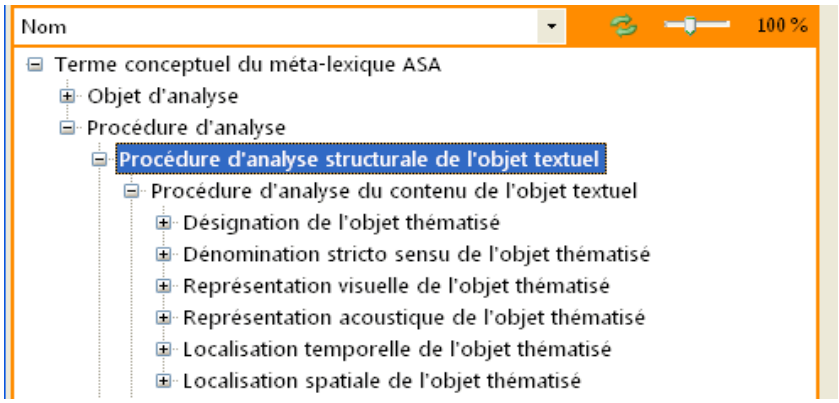

Figure 14.2. *La procédure d'analyse du contenu de l'objet textuel*

Ainsi, nous pouvons y distinguer trois domaines taxinomiques centraux représentés par les termes conceptuels [Procédure d'analyse du contenu de l'objet textuel] (voir figure 14.2), [Procédure d'analyse de l'expression multimodale de l'objet textuel] (voir figure 14.4) et [Procédure d'analyse du support textuel] (voir également figure 14.4).

Comme on peut le constater, nous ne nous sommes pas appuyés pour développer ce vocabulaire conceptuel sur notre habituel triptyque canonique [Taxon de …], [Caractéristique de …] et [Système de…]. La raison de cette omission est de nature

purement pragmatique : nous n'en avons pas eu besoin, la hiérarchie des termes conceptuels pour ce vocabulaire-ci étant bien moins complexe que celui représentant les *objets d'analyse** de l'*univers du discours** ASA (voir chapitre 13). Ceci dit, à terme ce vocabulaire sera également transformé, d'une manière tout à fait similaire au vocabulaire conceptuel représentant les objets d'analyse, en une *hiérarchie de domaines taxinomiques*.

En considérant de plus près la figure 14.2, on voit bien que le terme conceptuel [Procédure d'analyse du contenu de l'objet textuel] se différencie en toute une série d'activités spécialisées.

Les deux premiers groupes de ces activités plus spécialisées sont identifiés par les termes conceptuels [Désignation de l'objet thématisé (dans un texte audiovisuel)] et [Dénomination *stricto sensu* de l'objet thématisé]. Ils sont réservés à l'activité linguistique de l'indexation (libre) pour identifier et nommer un objet sujet dans un texte audiovisuel soumis à l'analyse. Nous en avons rencontré des exemples concrets lors de notre discussion sur la procédure de *description libre**.

Sans vouloir entrer dans une présentation détaillée, signalons au moins que ces deux groupes de procédures d'analyse du contenu réunissent toute une diversité d'activités d'analyse encore plus spécialisées. Ainsi, du côté [Désignation de l'objet thématisé], nous trouvons, entre autre, les activités suivantes que nous avons déjà rencontrées dans nos exemples concrets (voir par exemple, ceux du chapitre 3 ou du chapitre 9) :

– [Désignation minimale de l'objet thématisé] (et de leurs variantes pragmatiques) ;
– [Désignation contextualisée de l'objet thématisé] ou encore ;
– [Désignation de l'objet thématisé par mots-clés].

Du côté [Dénomination *stricto sensu* de l'objet thématisé] (figure 14.2), nous trouvons toutes les opérations d'analyse qui demandent à l'analyste de produire des noms propres appropriés pour identifier par exemple :

– le nom d'un *lieu* (d'un courant d'eau, d'un relief géographique, etc.) ;
– le nom d'un *produit* ou d'une *marque* ;
– le nom et l'acronyme d'une *institution* ou encore ;
– l'ethnonyme d'un *peuple* ou d'un *groupe ethnique*.

Enfin, on trouve ici aussi l'activité d'analyse [Dénomination anthroponymique] (et ses différentes variantes) pour identifier une *personne* par son ou ses différents noms (nom de famille, prénom, patronyme, matronyme, surnom, etc.).

Deux autres groupes d'activités qui font partie du domaine taxinomique [Procédure d'analyse du contenu de l'objet textuel] (figure 14.2), sont représentés par les termes conceptuels [Représentation visuelle de l'objet thématisé] et [Représentation acoustique de l'objet thématisé]. Les activités réunies dans ce

domaine taxinomique offrent à l'analyste la possibilité d'indexer le contenu d'un texte audiovisuel non pas à l'aide d'expressions linguistiques (éventualité offerte à l'analyste par les activités d'analyse faisant partie des deux premiers groupes cités ci-dessus) mais plutôt à l'aide d'expressions *visuelles* et/ou *acoustiques*. Un analyste habile et expérimenté peut ainsi créer toute une *méta-iconographie* propre à « son » archive pour représenter tel ou tel type de sujets, tels ou tels thèmes de discours – une façon certainement très originale et inventive de réaliser des interfaces visuelles et/ou acoustiques d'accès au contenu d'une archive…

Figure 14.3. *Accès au fonds filmique d'une archive audiovisuelle via carte géographique Google (exemple : les archives ADA – Atelier des Arkéonautes)*

Les deux derniers groupes du domaine taxinomique [Procédure d'analyse du contenu de l'objet textuel] (figure 14.2) sont représentés par les termes conceptuels [Localisation temporelle de l'objet thématisé] et [Localisation spatiale de l'objet thématisé]. Ils réunissent, comme leurs appellations l'indiquent déjà, une série d'activités permettant à l'analyste d'identifier, si pertinent, pour un objet thématisé ses coordonnées géographiques et les dates (précises ou approximatives) qui ponctuent son existence dans le temps.

Les données – géographiques et temporelles – collectées par l'analyste ou générées d'une manière automatique par le système ASA[70] (pour plus d'informations, voir [GUE 11]) sont utilisées, entre autres, pour la réalisation d'accès au fonds filmique d'une archive audiovisuelle via cartes géographiques et frises temporelles interactives. Ainsi, par exemple, une bonne partie du fond audiovisuel des archives développées dans le cadre du projet ASA-SHS est accessible via des frises temporelles et des cartes géographiques Google (voir la figure 14.3 montrant l'accès au fonds audiovisuel des archives *ADA – Atelier des Arkéonautes via* la carte géographique Google).

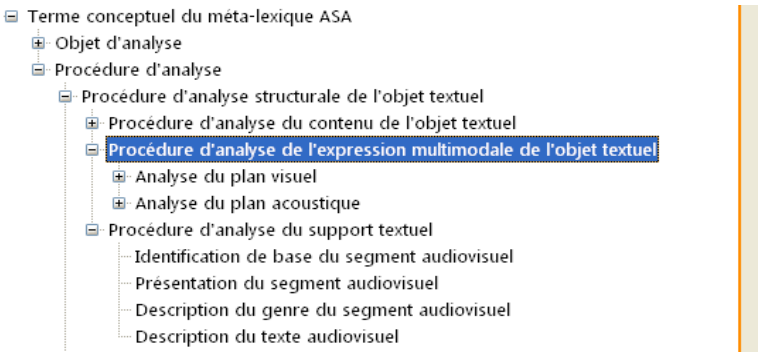

Figure 14.4. *La procédure d'analyse de l'expression multimodale et du support de l'objet textuel*

La figure 14.4 nous montre que le domaine taxinomique initié par le terme conceptuel [Procédure d'analyse structurale de l'objet textuel] se différencie encore – outre l'ensemble de procédures d'analyse consacrées au contenu d'un texte audiovisuel – en deux autres ensembles d'activités d'analyse, à savoir en celui représenté par le terme conceptuel [Procédure d'analyse de l'expression multimodale de l'objet textuel] et en celui représenté par le terme conceptuel [Procédure d'analyse du support textuel]. L'ensemble des activités représenté par le terme conceptuel de base [Procédure d'analyse de l'expression multimodale de l'objet textuel] nous sert à l'analyse du langage visuel et/ou audiovisuel. Il s'agit ici, par exemple, de l'analyse du cadrage visuel (ou acoustique) d'un objet thématisé, du champ et de l'hors-champ visuel, des plans visuels (et acoustique), etc.

Les activités réunies sous le terme conceptuel de base [Procédure d'analyse du support textuel] nous servent à l'identification de segments pertinents dans un texte

[70] A l'aide d'une application invisible à l'analyste qui récupère de Wikipédia plus particulièrement les coordonnées géographiques nécessaires pour générer une localisation de vidéos entières et/ou de passages choisis dans une vidéo sur des cartes géographiques Google.

audiovisuel, au découpage du texte audiovisuel en des unités plus petites et en la description et le classement de ces unités.

14.4. La classe des activités [Procédure d'analyse de l'objet textuel à l'aide du thesaurus ASA]

Dans la figure 14.5 nous voyons l'extrait de l'organisation assez complexe de l'ensemble des activités qui composent la *procédure de description contrôlée** (voir chapitre 10), c'est-à-dire la procédure qui fait appel, contrairement à celle de la *description libre** (voir chapitre11), à l'usage d'un ou de plusieurs *thesaurus** (micro-thesaurus).

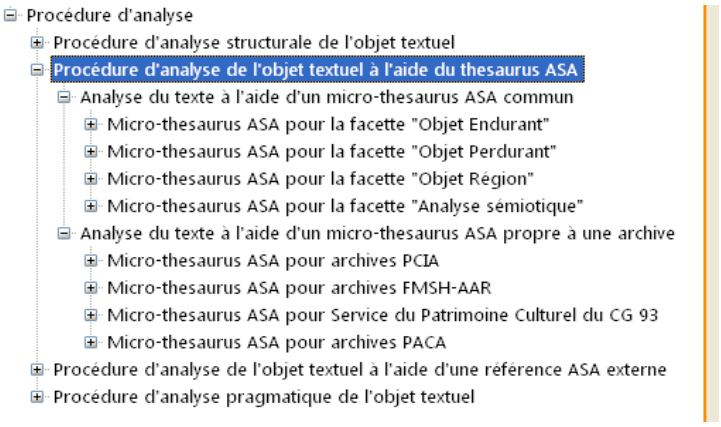

Figure 14.5. *La procédure d'analyse de l'objet textuel à l'aide du thesaurus ASA*

Dans la dite figure, nous observons une organisation bipartite qui se base sur la notion de *micro-thesaurus* et sur la distinction entre *micro-thesaurus communs* à l'ensemble des archives audiovisuelles se servant des *ressources métalinguistiques ASA** et *micro-thesaurus propres* à une archive audiovisuelle, c'est-à-dire à l'*univers du discours** d'une archive audiovisuelle spécifique. Un micro-thesaurus est formé :

– par une *facette*, c'est-à-dire un *trait sémantique* (ou, comme nous préférons le dire, un *classème* au sens de la sémantique structurale [GRE 67]) et ;
– une liste (hiérarchique) d'*expressions contrôlées* (c'est-à-dire une terminologie) qui constituent les valeurs de la facette.

Ainsi, par exemple, l'expression contrôlée <XVIIIe siècle> peut faire partie, entre autres, d'un micro-thesaurus qui réunit simplement une liste d'unités

temporelles découpant l'axe chronologique en des périodes arbitraires de cent ans ; mais il peut faire partie également d'un micro-thesaurus des époques de l'histoire de France où il désigne plutôt une époque qu'un découpage chronologique arbitraire.

Toujours est-il que nous utilisons toute une série de micro-thesaurus pour pouvoir décrire et indexer des objets thématisés dans l'*univers du discours** d'une archive audiovisuelle donnée. Ainsi, disposons-nous de toute une série de micro-thesaurus pour localiser chronologiquement un *objet thématisé** dans des *corpus audiovisuels** les plus variés.

Cependant, il est évident que l'univers du discours d'une archive audiovisuelle donnée possède ses spécificités sémantiques référentielles ou discursives dont il faut tenir compte. Par exemple, une archive audiovisuelle consacrée au patrimoine immatériel andin (au Pérou et en Bolivie)[71], a besoin d'un micro-thesaurus assez particulier de localisation géographique (territoriale, administrative, ethnique, linguistique, etc.) qui puise dans le vocabulaire du thesaurus ASA mais qui les classe selon une facette spécifique à l'univers de discours en question. Il se peut que ce micro-thesaurus soit utile pour une autre archive. Dans ce cas-là, il sera mis à disposition de cette nouvelle archive mais on considère qu'il fait plutôt partie de besoins propres à un groupe d'utilisateurs ou à une archive spécifique, voire à une communauté spécifique d'archives. L'activité d'analyse qui s'en sert est donc rangé sous le terme conceptuel de base [Analyse du texte à l'aide d'un micro-thesaurus ASA propre à une archive] (voir figure 14.5).

14.5. La classe des activités [Procédure d'analyse de l'objet textuel à l'aide d'une référence ASA externe]

La figure 14.6 nous montre la troisième classe de procédures intitulée [Procédure d'analyse de l'objet textuel à l'aide d'une référence ASA externe]. Il s'agit ici d'une classe d'analyse qui possède une fonction très stratégique car elle offre à l'analyste la possibilité d'utiliser l'environnement de travail ASA (c'est-à-dire le *Studio ASA* composé des trois ateliers de *Segmentation*, de *Description* et de *Publication* de corpus audiovisuels) tout en procédant à des analyses concrètes en référence à des standards, thesaurus ou autres langages externes à l'univers ASA. Ainsi, un analyste peut réaliser une description qui tient compte de la norme LOMFR ; un autre peut produire une description conforme au standard *Dublin Core*, et ainsi de suite.

[71] Nous référons ici au très beau projet de Valérie Legrand-Galarza *Patrimoine Culturel Immatériel Andin* réalisé dans le cadre de deux projets de R&D : le projet français ASA-SHS et le projet européen Convergence.

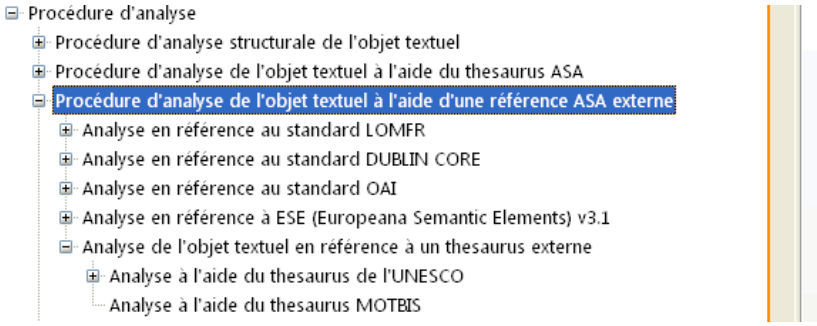

Figure 14.6. *La procédure d'analyse de l'objet textuel à l'aide d'une référence externe*

Prenons l'exemple de la description d'un texte ou d'un corpus audiovisuel à l'aide d'un *modèle de description** qui intègre la norme LOMFR. Une telle description sert, entre autres, à expliciter et à classer des textes audiovisuels (ou encore des passages spécifiques dans un texte audiovisuel) selon un point de vue éducatif et en prenant en considération l'intérêt potentiel de l'objet analysé pour un usage pédagogique spécifique. Autrement dit, et en adoptant le point de vue ASA, il s'agit de produire :

– *d'abord* une analyse soit du contenu (du *sujet**) d'un texte (ou d'un passage spécifique dans un texte) audiovisuel soit du langage audiovisuel (c'est-à-dire des plans visuels et acoustiques);
– et *ensuite* une évaluation du contenu ou du langage audiovisuel analysé selon les critères propres à la norme LOMFR.

La figure 14.7 nous montre ainsi un extrait de la première partie de cette analyse qui consiste en la description du contenu d'un passage audiovisuel thématisant la fabrication traditionnelle du pain dans le nord-est du Portugal[72].

[72] Documentation audiovisuelle réalisée par Mouette Barboff (Association *L'Europe, Civilisation du Pain*), Elisabeth de Pablo et Richard Fillon (ESCoM-FMSH) en 2006 dans le cadre du Programme Archives Audiovisuelles de la Recherche

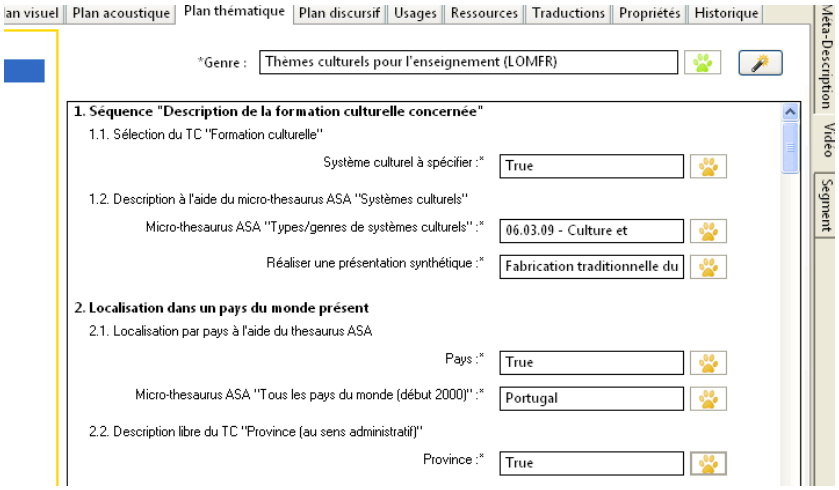

Figure 14.7. *Première partie de l'interface de travail permettant de saisir une description en référence à la norme LOMFR : description de l'objet*

En ajoutant à cette première description une deuxième qui consiste en une procédure normalisée d'évaluation de la pertinence éducative ou pédagogique spécifique du contenu analysé, l'analyste contribue non seulement à une meilleure diffusion/accessibilité du texte audiovisuel analysé mais aussi à sa meilleure appréciation en tant que ressource pédagogique pertinente par les acteurs (enseignants, formateurs, etc.) concernés. C'est ce que nous montre la figure 14.8 dans laquelle nous voyons seulement le début de la saisie des informations à collecter en vue de rendre conforme une description à la norme LOMFR.

Ainsi, pour reprendre notre exemple, l'interface de travail représentée partiellement par les deux figures 14.7 et 14.8 permettrait une mise en place assez facile d'une collection pédagogique conforme à la norme LOMFR qui proposerait à un public intéressé (enseignants, formateurs, etc.) des sujets culturels sous forme de documents audiovisuels à utiliser dans le cadre de tel ou tel enseignement, correspondant à tel ou tel niveau, pour telle ou telle catégorie d'apprenants, favorisant tel ou tel style pédagogique, et ainsi de suite.

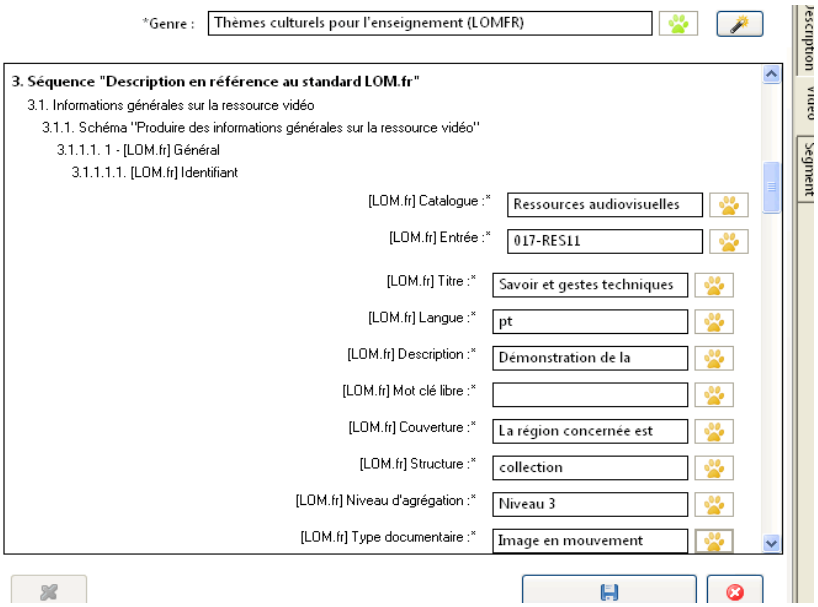

Figure 14.8. *Deuxième partie de l'interface de travail permettant de saisir une description en référence à la norme LOMFR : début de la description selon la norme LOMFR*

En revenant encore une fois à la figure 14.6, on voit bien que la classe d'activités de description actuellement disponibles pour procéder à des analyses conformes à une référence externe au système ASA, n'est pas très développée. Comme nous l'avons déjà dit, il s'agissait pour nous avant tout d'étudier et de tester la possibilité d'intégrer, dans le système ASA, des *modèles de description** qui tiennent compte de telles références. L'avenir montrera comment cette classe d'activités d'analyse évoluera. Outre le cas particulier de la norme LOMFR, nous sommes pour le moment (et comme le montre la figure 14.6) surtout intéressés à fournir à l'analyste la possibilité de réaliser des descriptions de contenus audiovisuels susceptibles d'être (re-)diffusées sur les grands portails de patrimoines culturels et scientifiques qui intègrent souvent des standards de type *Dublin Core* ou OAI.

14.6. La classe des activités [Procédure d'analyse pragmatique de l'objet textuel]

Enfin, la figure 14.9 nous montre les trois groupes principaux d'activités d'analyse qui composent le domaine taxinomique [Procédure d'analyse pragmatique de l'objet textuel]. Sous le terme conceptuel [Procédure de traduction de l'objet textuel], nous réunissons toutes les activités qui servent à rendre accessible un texte

audiovisuel à un public qui ne possède pas les compétences linguistiques nécessaires pour le comprendre.

Comme nous l'avons déjà expliqué ailleurs [STO 08] et [STO 10], il s'agit ici d'ouvrir un texte audiovisuel qui est en général monolingue à un « marché de connaissances » qui, lui, est intrinsèquement multilingue. Il ne s'agit pas nécessairement de produire des traductions professionnelles (sous forme de doublages, de sous-titrages, etc.) de chaque texte audiovisuel ce qui est, au moins d'un point de vue économique, pratiquement impossible. Il s'agit plutôt d'offrir des services de *médiation linguistique* de type *traduction relais* ou *traduction communautaire* permettant à un analyste-traducteur de produire, dans la langue du public visé, des notes et résumés explicatifs, des traductions en *style télégraphique*, voire des versions plus ou moins libres.

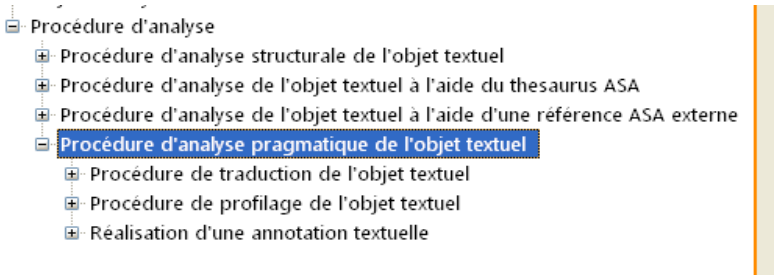

Figure 14.9. *La procédure d'analyse pragmatique de l'objet textuel*

Le terme conceptuel de base [Procédure de profilage de l'objet textuel] (figure 14.9), réunit les activités qui permettent à l'analyste de jouer le rôle d'un *médiateur culturel* entre l'identité spécifique d'un texte audiovisuel donné et l'identité de son public, le cadre des références culturelles de ce dernier, ses attentes et, enfin, ses compétences cognitives. Concrètement parlant, ces activités se réalisent sous forme de commentaires, d'explications, de modes d'emploi et d'aides, de localisation de références bibliographiques ou webographiques, etc.

Le dernier groupe d'activités représenté par le terme conceptuel de base [Réalisation d'une annotation textuelle], réunit enfin un petit ensemble d'activités de niveau très général dont l'objectif est de permettre à l'analyste de produire, comme l'intitulé du terme conceptuel le suggère déjà, des notices textuelles qui expliquent son travail d'analyse.

Nous retrouvons ici le cas de la fameuse *méta-description* présentée plus en détail ailleurs [CHE 11a]. La méta-description n'est pas une description de l'objet audiovisuel mais plutôt une explication, fournie par l'analyste du contenu et des objectifs de son propre travail d'analyse ou d'expertise. En effet, il existe

évidemment, comme développé dans [CHE 11a], toute une diversité de types d'analyse : de la description la plus simple et la plus rudimentaire jusqu'à la « vraie » expertise d'un texte audiovisuel en passant par les descriptions de type documentaliste et bibliothéconomique. C'est, bien sûr, à l'analyste de tracer le territoire de son propre faire analytique et de fournir les éléments permettant d'apprécier la spécificité de son travail. Le petit groupe d'activités réunies sous le terme conceptuel [Réalisation d'une annotation textuelle] ouvre à l'analyste la possibilité de réaliser un tel travail explicatif.

14.7. Activité d'analyse et schémas d'indexation

Toute activité d'analyse (ou de description) est définie par un ou plusieurs *schémas* dits *d'indexation**. Un schéma d'indexation est composé d'un ou de plusieurs *champs de saisie d'informations*.

La figure 14.10 nous montre cette organisation relativement complexe qui sous-tend *toute* activité d'analyse. Elle nous montre également la similarité partielle et la différence entre un terme conceptuel représentant un *objet d'analyse** (c'est-à-dire faisant partie du vocabulaire conceptuel [Objet d'analyse] et un terme conceptuel représentant une activité d'analyse (c'est-à-dire faisant partie du vocabulaire [Procédure d'analyse]).

Toute activité d'analyse est donc définie par un ou plusieurs schémas d'indexation. Ce fait nous est exemplifié par la figure 14.11 qui nous montre l'activité d'analyse intitulée [Désignation minimale de l'objet thématisé – formulaire standard]. Cette activité est composée de *trois schémas d'indexation* :

i) le schéma d'indexation *Saisie de l'expression minimale* (schéma permettant à l'analyste de saisir « physiquement » l'expression linguistique – habituellement nominale – qui lui semble appropriée pour désigner l'objet thématisé dans son texte) ;

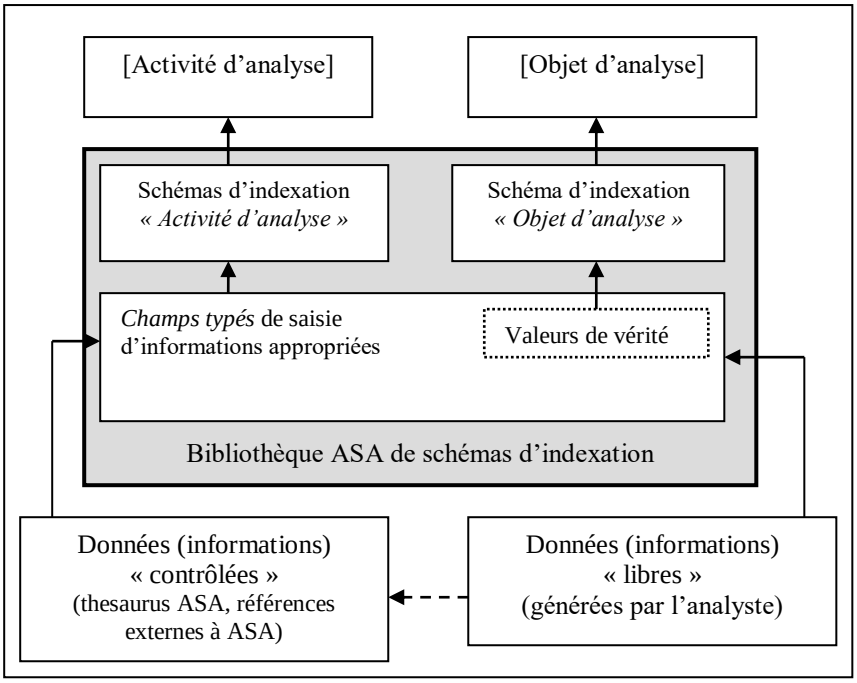

Figure 14.10. *Rapports entre schémas d'indexation et les deux méta-lexiques des termes conceptuels de l'univers du discours ASA*

ii) le schéma d'indexation *Catégorie linguistique (de l'expression minimale)* schéma permettant à l'analyste de préciser si l'expression minimale, saisie par lui, est, par exemple, un nom simple, un nom composé, un syntagme nominal, une expression figée, etc. ;

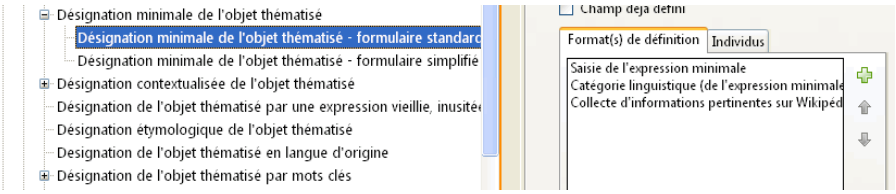

Figure 14.11. *L'activité [Désignation] et son interprétation par des schémas d'indexation*

iii) le schéma d'indexation *Collecte d'informations pertinentes sur Wikipedia* (schéma permettant à l'analyste, s'il le souhaite, d'indiquer l'url d'un article Wikipedia où le système ASA pourra collecter d'autres informations sur l'objet

désigné ; ce service – actuellement seulement disponible en langue anglaise – est particulièrement utile pour collecter, comme déjà expliquée auparavant, des données géographiques et/ou temporelles pouvant être utilisées pour des accès à un fonds audiovisuel via carte géographique et/ou frise temporelle).

La figure 14.12 montre à quoi ressemble l'activité [Désignation minimale de l'objet thématisé – formulaire standard] ainsi définie lorsque l'analyste s'en sert pour réaliser une description concrète.

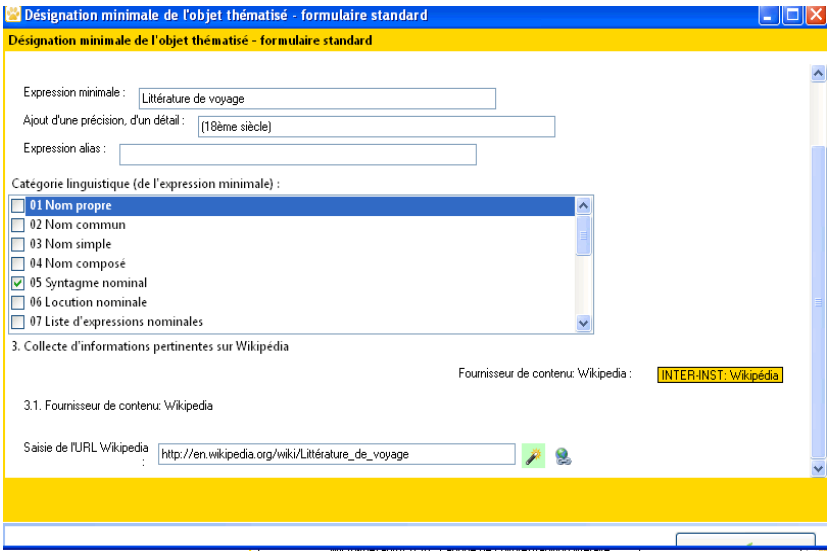

Figure 14.12. *Interface de travail définissant l'activité [Désignation minimale de l'objet thématisé – formulaire standard]*

Retenons donc que le ou les *résultats* d'une opération spécifique d'analyse sont définis par des *schémas dits d'indexation*. Comme déjà exposé dans le premier chapitre (section 1.4), il faut prendre le terme « indexation » ici au sens très large, c'est-à-dire au sens de *tout résultat concret*, de *l' « output » physique d'une analyse* sous forme, par exemple, de mots clés, de paragraphes textuels, d'icônes visuels, de jingles acoustiques, de données numériques, de données de géo-localisation, de termes faisant partie d'une terminologie ou d'un thesaurus, etc.

14.8. La bibliothèque des schémas d'indexation

Nous distinguons bien, dans la figure 14.10, une partie grisée centrale. Celle-ci nous suggère l'existence de toute une bibliothèque de schémas d'indexation dans

laquelle on puise pour définir un *terme conceptuel** faisant partie du méta-lexique ASA.

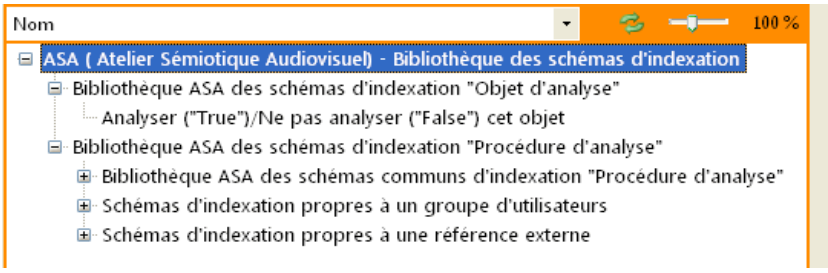

Figure 14.13. *Extrait de la bibliothèque des schémas d'indexation ASA*

La figure 14.13 nous montre un extrait de la *bibliothèque* de schémas d'indexation composée d'un ensemble de *collections* de schémas d'indexation plus spécialisées. Celle-ci se différencie notamment en deux parties parfaitement complémentaires :

– une bibliothèque des schémas d'indexation réservés, comme signalé également dans la figure 14.10, à la *sélection* de tel ou tel *terme conceptuel* (dans un formulaire interactif de travail) qui doit être soumis à l'analyse ;

– et une bibliothèque de schémas d'indexation pour *définir une procédure ou une activité d'analyse*.

La bibliothèque des schémas réservés à l'identification d'un terme conceptuel est réduite, dans notre cas, à un et un *seul schéma de sélection* ou d'identification d'un terme conceptuel. Il s'agit du schéma d'indexation intitulé *Analyser (« True »)/Ne pas analyser (« False ») cet objet*. Cependant, d'autres schémas peuvent venir compléter celui-ci comme, par exemple, des *schémas de quantification* du référent ou encore des *schémas modaux* (de modalités) permettant de mieux définir les conditions et modes d'existence d'un référent représenté par un terme conceptuel et analysé, via des schémas d'indexation appropriés, par l'analyste.

La *bibliothèque des schémas d'indexation* réservés à la *définition des activités d'analyse* est, en revanche, très élaborée et composée de toute une série de schémas d'indexation permettant un suivi très fin et nuancé de cet acte essentiel pour l'analyse qui est *la* saisie appropriée d'informations (verbales, visuelles, acoustiques, symboliques, etc.).

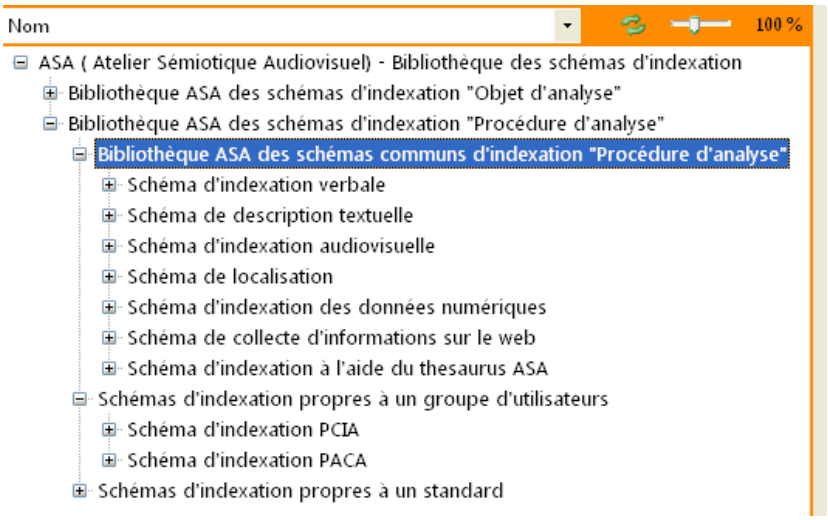

Figure 14.14. *Extrait de la bibliothèque des schémas d'indexation réservés à la définition d'une procédure d'analyse*

Ainsi pouvons-nous distinguer dans la *bibliothèque des schémas d'indexation « Procédures d'analyse »* (figure 14.14) une collection de schémas d'indexation réservés à l'*indexation verbale* d'un terme conceptuel ou d'une configuration de termes conceptuels qui définissent un domaine de connaissances thématisé dans un texte ou un corpus de textes audiovisuels. Cette collection de schémas se divise en une série de collections plus spécialisées (figure 14.15) : une collection de schémas d'indexation réservés à la saisie des noms propres ; une collection de schémas réservés à la saisie de l'expression d'un terme conceptuel en langue d'origine, etc.

Une collection centrale ici est celle qui s'intitule *Schéma d'indexation de l'expression minimale*. C'est le schéma ou, plutôt la collection de schémas d'indexation qui définit le *seul* acte obligatoire à exécuter par l'analyste s'il souhaite produire des informations relatives à un objet donné à l'aide de la *procédure de description libre**. Autrement dit, les schémas de cette collection définissent le type d'activité d'analyse [Désignation minimale de l'objet thématisé] qui connaît, pour le moment, deux versions : une version simplifiée et une version dite standard. L'interface de la version standard est représentée par la figure 14.12. Elle est composée des deux schémas d'indexation (figure 14.15) : *Saisie de l'expression minimale* et *Catégorie linguistique (de l'expression minimale)*. La version simplifiée de l'activité d'analyse [Désignation minimale de l'objet thématisé], en revanche, n'est défini que par le schéma d'indexation [Saisie de l'expression minimale. – formulaire standard].

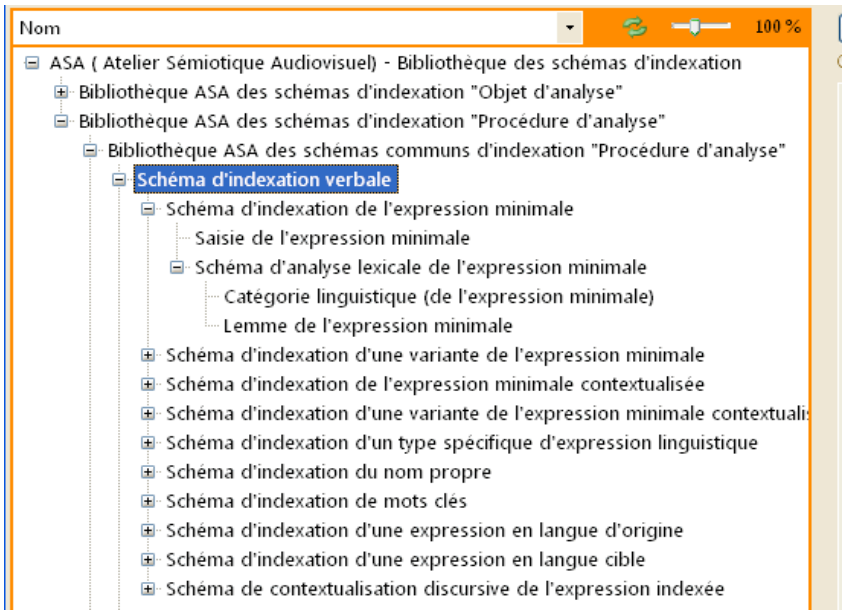

Figure 14.15. *Extrait de la bibliothèque des schémas d'indexation réservés à l'indexation verbale*

Ce petit exemple montre comment nous avons essayé de « marier » des besoins en apparence fort divergents comme par exemple, celui de pouvoir réaliser des descriptions simples et rapides avec celui de pouvoir réaliser, au contraire, des analyses approfondies, sophistiquées, adaptées à des objectifs spécialisés. Les *ressources métalinguistiques ASA** nous permettent, en effet, de spécifier et de développer des modèles de description qui ne sont composés que d'un seul terme conceptuel représentant un seul type d'objet et dont la description est réduite à l'activité [Désignation minimale de l'objet thématisé] dans sa version simplifiée, c'est-à-dire ne demandant à l'analyste de saisir qu'un seul champ pour que l'analyse soit achevée. Elles nous permettent également de développer des modèles très sophistiqués et « sur mesure » en passant par toute une panoplie de modèles plus proches à telle ou telle pratique documentaire ou encore par des modèles de description qui s'appuient sur des références (standards, normes, etc.) externes.

Dans la figure 14.15, nous pouvons distinguer d'autres collections de schémas d'indexation. Une collection de schémas d'indexation est spécialisée, par exemple, dans l'*indexation visuelle* ou *acoustique*, c'est-à-dire dans l'usage d'une méta-iconographie représentant (illustrant, catégorisant à l'aide d'un « lexique » visuel et/ou acoustique) le contenu d'un texte ou d'un corpus de textes audiovisuels.

Une autre collection de schémas d'*indexation dite textuelle* sert à la réalisation d'une annotation textuelle (voir figure 14.15).

Une collection très importante dans cette bibliothèque de schémas d'indexation recouvre tous les *schémas d'indexation* incorporant des termes ou listes de termes prédéfinis parmi lesquelles l'analyste doit faire son choix (contrairement aux schémas d'indexation verbale discutés ci-auparavant qui permettent à l'analyste de produire librement une indexation appropriée). Ces schémas appartiennent à la collection *Schémas d'indexation à l'aide du thesaurus ASA* (figure 14.14).

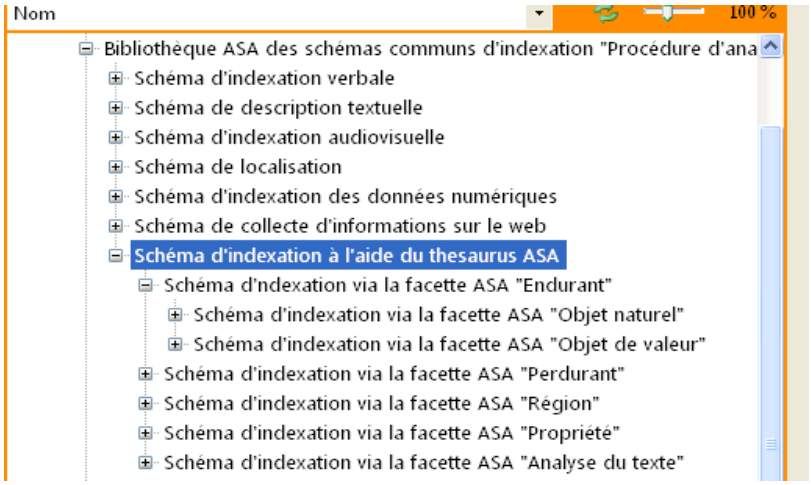

Figure 14.16. *Extrait de la bibliothèque des schémas d'indexation réservés à l'indexation verbale*

La collection intitulée *Schémas d'indexation à l'aide du thesaurus ASA* est, comme le montre la figure 14.16, très diversifiée, et composée d'un nombre important de schémas d'indexation thématiquement spécialisés. En effet, comme on peut le voir dans la figure 14.16, tous les schémas composant cette collection de la bibliothèque des schémas d'indexation ASA, partagent une caractéristique commune qui celle de « pointer » vers un *rang* de *valeurs prédéfinies* dans le thesaurus ASA. Un rang de valeurs prédéfinies est un espace sémantique que nous appelons *facette* et qui correspond à ce que Greimas a appelé un *classème*, c'est-à-dire un trait sémantique qui (l'appellation le suggère déjà) classe *ensemble* un certain nombre de valeurs (d'expressions normalisées, de descripteurs) dans le thesaurus.

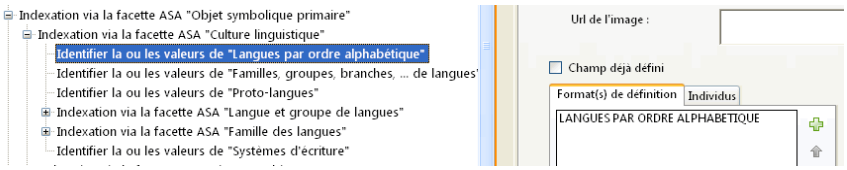

Figure 14.17. *Collection de schémas d'indexation par facette réservées à la description du domaine taxinomique Culture linguistique*

Considérons rapidement la figure 14.17. Elle nous montre une collection spécialisée de schémas d'indexation par facette qui composent un ensemble de ressources métalinguistiques pour décrire des textes audiovisuels développant une problématique de « culture linguistique » : langues, familles de langues, structures et usages d'une langue, systèmes d'écriture, etc.

Comme on peut le voir (figure 14.17), nous avons sélectionné un schéma particulier d'indexation par facette intitulée *Identifier la ou les valeurs de « Langues par ordre alphabétique »*. Ce schéma d'indexation définit un type d'*activité d'analyse** qui fait partie d'une procédure de *description contrôlée** dont on se sert pour élaborer des *modèles de description** pour analyser des textes audiovisuels thématisant une ou plusieurs langue(s) (dans un contexte donné, en rapport à tel ou tel usage, tel ou tel groupe de locuteurs, …) et qui fait partie d'une ou de plusieurs *bibliothèques** de modèles de description définissant, comme nous le savons déjà, l'*univers du discours** d'une archive audiovisuelle donnée.

Dans la partie droite de la figure 14.17, nous voyons l'expression (en majuscules) LANGUES PAR ORDRE ALPHABETIQUE. Celle-ci dénomme le rang de valeurs, c'est-à-dire la facette qui fait partie du thesaurus ASA (voir chapitre 15) et qui sert comme classème pour un certain nombre d'expressions (normalisées).

Par exemple, <Français> et <Nénètse> sont deux expressions (normalisées) qui relèvent – comme bien d'autres expressions (normalisées) du thesaurus ASA – du rang ou de la facette *Langues par ordre alphabétique*. Mais, comme nous l'avons déjà vu dans le chapitre 11, une expression normalisée peut constituer la valeur de deux ou plusieurs facettes. Ainsi, <Nénètse> fait également partie de la facette ASA *Langues et familles de langues ouraliennes*, de même que <Français> qui fait partie de la facette ASA *Langues et familles de langues indoeuropéennes*.

On voit bien que la *facette* exprime une ou une partie d'une classification qui peut être de nature scientifique, pratique, utilitaire, « populaire », etc. Autrement dit, elle permet d'expliciter le trait ou la caractéristique sémantique dont se sert un acteur (une personne, un groupe social, une institution) pour grouper ensemble des choses qu'un autre acteur (une autre personne, un autre groupe social, une autre institution) range dans des catégories différentes.

Chapitre 15
Le thesaurus ASA

15.1. Introduction

Nous avons déjà souligné à plusieurs reprises le rôle central du thesaurus dans le dispositif métalinguistique ASA. Celui-ci réside, entre autre, dans le fait qu'il nous permet à la fois de maintenir à un niveau d'assez grande généralité le *méta-lexique des termes conceptuels** (des *objets d'analyse** et des *activités d'analyse**) et de contenir le nombre tout en pouvant rendre compte non seulement des spécificités référentielles de l'*univers du discours** de telle ou telle archive mais aussi des points de vue et donc des diverses classifications (des diverses « folksonomies ») des objets d'analyse dans un univers du discours donné.

Le thesaurus est par ailleurs un outil indispensable pour la *procédure de description contrôlée** qui, comme nous le savons déjà, constitue l'une des deux procédures de description de base d'un texte ou d'un corpus de textes audiovisuels.

Comme nous avons déjà eu l'occasion de développer les aspects les plus importants du thesaurus ASA (voir notamment les chapitres 10 et les sections 11.4 et 14.4), nous allons nous contenter, dans ce chapitre, d'en présenter d'une manière synthétique et générale son organisation et son fonctionnement.

La section 15.2 nous rappellera encore une fois place et fonction du thesaurus dans le système métalinguistique ASA.

La section 15.3 traitera plus particulièrement un aspect central dans la construction et le fonctionnement interne du thesaurus ASA qui est celui des facettes interprétées ici au sens d'un *trait sémantique* – d'un *classème*, pour prendre un concept utilisé par Greimas dans sa sémantique structurale [GRE 66] – qui constitue une dimension possible du sens (du *signifié*) d'un terme conceptuel et qui est interprétée par un ou une série d'expressions normalisées également appelées *descripteurs*.

15.2. Présentation générale du thesaurus ASA

Dans le système des ressources métalinguistiques ASA les vocabulaires contrôlés jouent une fonction essentielle :

– d'une part, ils permettent de maintenir aussi réduit que possible le noyau du système métalinguistique ASA qui est formé par le méta-lexique des termes conceptuels représentant les *objets d'analyse** de l'univers du discours ASA ;
– et d'autre part, ils offrent une excellente possibilité de satisfaire aux habitudes, traditions, spécificités intellectuelles ou idéologiques des utilisateurs et groupes d'utilisateurs les plus différents face à un domaine de connaissances, un « terrain » ou un patrimoine déjà constitué.

Le thesaurus ASA a été conçu d'abord pour renseigner seulement quelques *termes conceptuels** tels que [Pays] (c'est-à-dire le nom de tel ou tel pays), [Région géopolitique] (c'est-à-dire le nom de telle ou telle région géopolitique), [Langue] (c'est-à-dire le nom de telle ou telle langue) ou encore [Epoque] (c'est-à-dire le nom de telle ou telle époque historique). Il s'agit ici d'une fonction typique du thesaurus qui consiste dans le fait de proposer des listes (alphabétiques, structurées, etc.) de *valeurs prédéfinies* pour un concept ou une configuration conceptuelle, c'est-à-dire un ensemble de concepts positionnés les uns par rapport aux autres dans un schéma ou une séquence de description d'un *modèle de description** (voir chapitre 10 pour plus d'explications)[73].

Même si on adoptait une vision théorique très large en permettant d'inclure tel ou tel pays, telle ou telle langue ou telle ou telle région géopolitique dans le *méta-lexique des termes conceptuels** représentant les *objets d'analyse** d'un *univers du*

[73] Le thesaurus ASA dispose, par exemple d'une facette appelée « Auteurs de la littérature française » qui réunit une liste assez conséquente (et ouverte, c'est-à-dire susceptible d'être complétée à tout moment) des noms de personne remplissant la fonction d'*être un auteur de la littérature française*. Du point de vue conceptuel, « auteur de la littérature française » est une configuration composée des termes conceptuels génériques [Auteur], [Littérature par pays] ainsi que du terme référentiel (c'est-à-dire possédant une seule valeur référentielle) [Pays : France].

*discours** donné, cette solution n'en est guère une. Après tout, pourrait-on argumenter, « France » pourrait bien être considérée comme un terme conceptuel spécialisé du terme conceptuel plus général qui est [Pays]. « France » considéré comme un terme conceptuel pourrait même avoir des valeurs référentielles fort différentes : une première fois, il pourrait signifier la définition constitutionnelle de la France comme état national et indépendant, une deuxième fois, il pourrait signifier la représentation cognitive d'un groupe social de ce qu'est la France, une troisième fois, il pourrait signifier un lieu privilégié de vacances, et ainsi de suite.

Adopter ce point de vue risque de faire en quelque sorte exploser une ontologie ne serait-ce que d'un point de vue purement quantitatif sans parler du fait (à notre avis bien plus crucial) que le métalangage de description se réduirait dans une telle optique à une simple substitution du lexique d'une langue naturelle par un soit disant lexique métalinguistique.

Si on mettait en œuvre une telle optique « radicale », le métalangage de description serait entre autres privé d'un de ses intérêts majeurs, à savoir d'être un outil de *classification raisonnée* et de *raisonnement* (de *solution de problèmes*) s'appuyant sur les descriptions d'un fonds audiovisuel réalisées grâce à une bibliothèque de modèles de description qui forme le *métalangage de description** propre à l'*univers du discours** d'une archive audiovisuelle.

Plus concrètement parlant, en adoptant une telle optique, il faudrait ajouter dans le méta-lexique des objets d'analyse ASA, sous la branche [Pays], les quelques 190 pays actuels du monde (sans parler des pays et autres territoires qui ont pu exister en tant qu'entité politiquement indépendante dans le passé) ; il faudrait ajouter, sous le terme conceptuel [Langue] les quelques milliers des langues actuellement parler dans le monde, et ainsi de suite.

Cependant, l'usage d'un thesaurus possède aussi ses limites évidentes dont les deux les plus importantes nous semblent être celle de l'*exhaustivité empirique* et celle d'imposer à l'analyste une *organisation terminologique* qui n'est pas obligatoirement celle qu'il souhaite utiliser. Tout thesaurus étant en quelque sorte figé, aucun n'échappe au fait qu'il soit *empiriquement limité*.

Par ailleurs, tout thesaurus étant un artefact, un outil conçu pour répondre à un certain type de problèmes de classification de documents ou d'objets (de *realia*) de toute sorte, son organisation interne peut correspondre aux attentes et besoins d'un analyste – mais il peut également ne pas y correspondre.

En tout cas, étant convaincu que l'usage de thesaurus reste à l'ordre de jour aussi dans le cadre du « web sémantique », nous avons réservé à cet outil une place importante dans l'économie générale des ressources métalinguistiques dont nous nous servons pour définir et réaliser les modèles de description d'un texte ou d'un corpus de textes audiovisuels faisant partie du fonds d'une archive.

La figure 15.1 nous montre l'organisation générale du thesaurus ASA développé dans le cadre du projet de recherche ASA-SHS. Nous y distinguons trois parties principales :

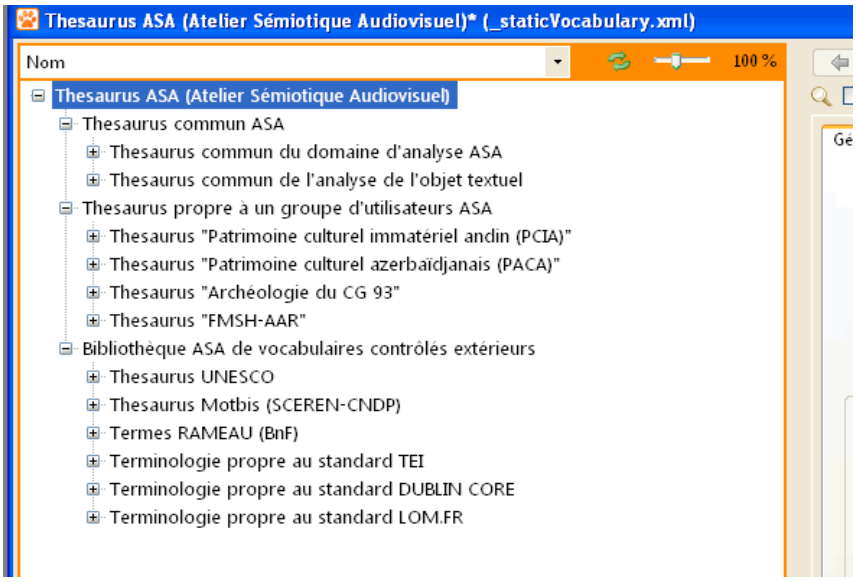

Figure 15.1. *Vue d'ensemble du thesaurus ASA*

1. La partie appelée *Thesaurus commun ASA*. Elle est, comme son nom l'indique déjà, mis à la disposition de tous les analystes de corpus audiovisuels et qui font partie de l'*univers du discours** ASA. Cette partie est elle-même divisée d'une part en un thesaurus spécialisé (et déjà très développé) consacré au domaine d'analyse (c'est-à-dire aux objets et domaines de connaissance de l'*univers du discours** ASA) et d'autre part en un thesaurus spécialisé consacré à l'analyse de l'objet textuel.

2. La partie appelée *Thesaurus propre à un groupe d'utilisateurs ASA*. Dans cette partie, on trouve les thesaurus élaborés pour répondre aux besoins spécifiques d'une archive audiovisuelle donnée. On y trouve plus particulièrement les facettes, c'est-à-dire les rangs de valeurs prédéfinies qui classent des expressions (normalisées) selon le point de vue adopté par un groupe d'utilisateurs pour analyser des corpus audiovisuels.

3. La partie appelée *Bibliothèque ASA de terminologies propres à un standard* ou *une norme*. Sont réunies ici les expressions provenant des divers

thesaurus, terminologies (glossaires, etc.) que nous utilisons directement (c'est-à-dire via l'interface de travail ASA) pour renseigner (indexer) un terme conceptuel ou une configuration de termes conceptuels.

15.3. Facettes et listes d'expressions normalisées

Les deux premières parties du thesaurus ASA (figure 15.1) sont construites de la même façon. Le thesaurus commun et tel ou tel thesaurus propre à un groupe d'utilisateurs sont constitués par un ensemble de facettes et par une liste (hiérarchique) d'expressions normalisées (de « descripteurs »).

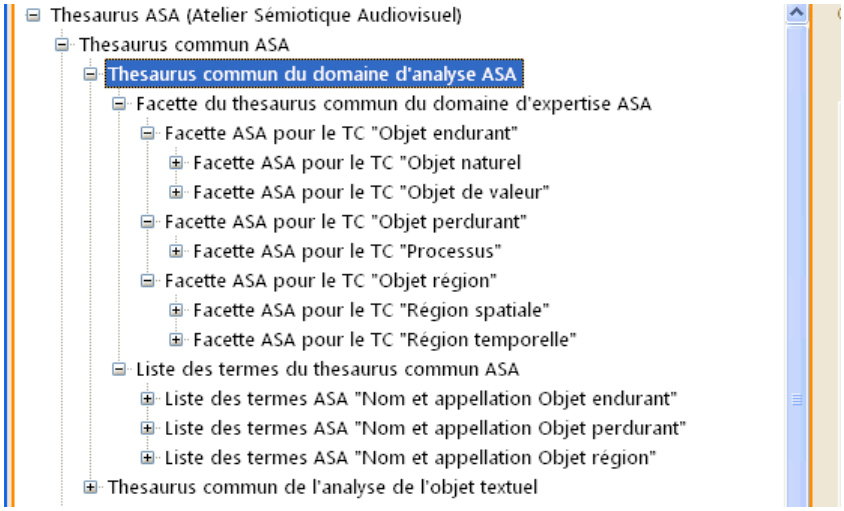

Figure 15.2. *Le thesaurus commun – facettes et listes des termes*

La figure 15.2 nous montre l'organisation du thesaurus commun sous forme de plusieurs *collections de facettes* et d'un ensemble de listes de termes ou d'expressions normalisées.

Comme déjà dit, une facette classe sémantiquement une liste d'expressions normalisées. La figure 15.3 nous en fournit un exemple concret. Elle nous montre une facette spécifique intitulée *Facette ASA pour le TC spécialisé « Civilisation ancienne d'Europe »*.

A cette facette est associée une liste d'expressions (de descripteurs) qui identifient différentes civilisations européennes anciennes : <Civilisation italique>, <Civilisation baltique>, <Civilisation étrusque>, etc. Lors de son travail de

description, l'analyste peut utiliser cette liste d'expressions normalisées sous forme d'un élément spécifique figurant dans la procédure de *description contrôlée** dont la fonction spécifique est de permettre à l'analyste de réaliser la tâche de l'identification et de l'explicitation de tous les *sujets** relatifs à une ou plusieurs civilisations européennes antiques.

Ceci dit, nous l'avons déjà signalé ailleurs (voir les chapitres 10, 11.4 et 14.4), une expression normalisée peut appartenir à *plusieurs facettes*. Cela signifie que l'expression normalisée appartenant à plus qu'une facette est considérée comme possédant *différents sens* dans l'*univers du discours** ASA.

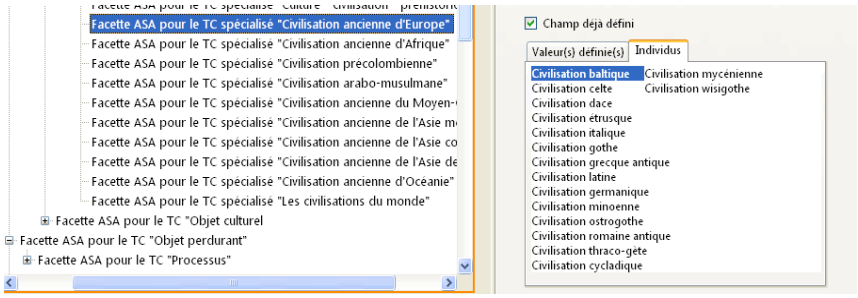

Figure 15.3. *Exemple d'une facette composée d'une liste d'expressions normalisées*

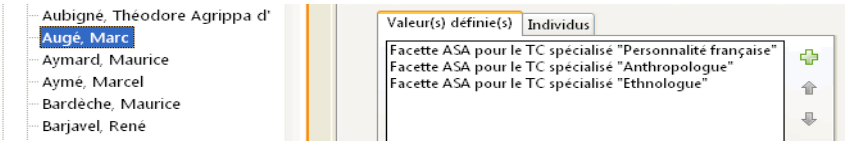

Figure 15.4. *Exemple d'une expression appartenant à plusieurs facettes*

La figure 15.4 nous montre l'exemple concret du nom d'une personne ; < Augé, Marc> qui fait partie d'une longue liste de noms de personne dont nous avons eu besoin pour la description des corpus audiovisuels analysés dans le cadre des ateliers d'expérimentation du projet ASA-SHS.

Le nom dans sa forme normalisée <Augé, Marc> possède, dans l'*univers du discours** ASA, trois acceptations, trois sens différents : le fait d'être une *personnalité française*, le fait d'être un *anthropologue* et le fait d'être un *ethnologue*. Il est fort probable que ce même nom possède encore toute une diversité d'autres sens – en dehors de l'univers du discours ASA. Mais dans celui-ci, il possède exactement ces trois sens.

Comme on peut le voir dans la figure 15.2, l'élément *Facette ASA pour le TC spécialisé « Civilisation ancienne d'Europe »* fait partie de toute une série de collections de facettes. Chaque facette représente, comme expliqué dans le chapitre 11.4, une dimension du contenu (du signifié), c'est-à-dire un axe sémantique spécifique d'un terme conceptuel ou d'une configuration de termes conceptuels ou encore d'un terme conceptuel instancié[74].

Dans la version actuelle du thesaurus commun, nous avons classé les différentes facettes dont nous avons eu besoin pour analyser les objets faisant partie de l'univers du discours ASA, d'une manière parfaitement similaire à la *structure taxinomique* des deux méta-lexiques des termes conceptuels désignant d'une part les objets d'analyse et d'autre part les activités spécifiques pour analyser les dits objets.

Ainsi distinguons-nous, comme le montre la figure 15.2, des collections de facettes relatives au terme conceptuel [Objet Endurant], des collections de facettes relatives au terme conceptuel [Objet perdurant] ou encore des collections de facettes relatives au terme conceptuel [Procédure d'analyse structurale de l'objet textuel]. Chaque collection peut être composée, en principe, de collections encore plus spécialisées de facettes.

Pour l'avenir, *deux grands enjeux* de recherche et de développement se dessinent. Le premier est celui de la mise à jour du thesaurus ASA, de son enrichissement, par les différentes ressources terminologiques et autres existantes mais qui lui sont extérieures tout en conditionnant ce processus d'enrichissement par rapport aux besoins et attentes des analystes travaillant pour telle ou telle archive spécifique.

Nous pensons ici d'abord aux ressources provenant d'un langage d'indexation tel que RAMEAU[75] de la Bibliothèque Nationale de France ou d'un thesaurus tel que MOTBIS[76] du CNDP-CRDP composé de toute une série de micro-thesaurus spécialisés qui correspondent, *grosso modo*, aux domaines taxinomiques organisant le méta-lexique des termes conceptuels ASA. Mais nous pensons également aux listes (ouvertes) des noms de lieux, de personnes, d'institutions, d'œuvres, etc. dont l'analyste d'une archive particulière pourrait avoir besoin et qui lui seraient fort utiles pour simplifier son travail d'analyste (la simplification consistant essentiellement dans le fait de « cocher » telle ou telle valeur pour un terme

---

[74] Un terme instancié est un terme conceptuel possédant une valeur spécifique. Rappelons que le terme conceptuel [Pays] est dit *terme générique* ; le terme [Pays : <Pérou>] est dit *terme spécialisé* ou, mieux, *instancié*. Ainsi l'élément *Facette PCIA « Provinces de Cuzco »* est une facette (une dimension du signifié) d'une configuration composée du terme générique [Province] et du terme spécialisé [Territoire : <Cuzco>].

[75] Voir http://rameau.bnf.fr/

[76] Voir http://www.cndp.fr/thesaurus-motbis/site/

conceptuel à renseigner au lieu d'en produire une description/indexation verbale libre).

Bien entendu, un tel processus d'enrichissement ne peut pas se faire exclusivement d'une manière « manuelle ». Il doit reposer sur des traitements au moins partiellement automatiques d'appariement des listes d'expressions et des facettes composant le thesaurus ASA avec les données métalinguistiques provenant de ressources externes au système ASA.

Un deuxième enjeu – tout aussi important que le premier – est celui de la *réutilisation des indexations libres* (c'est-à-dire des descriptions verbales ou autres réalisées via la *procédure de la description libre**) réalisées par la communauté d'analystes travaillant avec les outils et ressources ASA pour une archive audiovisuelle particulière.

Une telle réutilisation consisterait à fournir à l'analyste d'un texte ou d'un corpus de textes audiovisuels des propositions contrôlées d'expressions librement réalisées auparavant par d'autres analystes de la « communauté ASA ». Ces expressions librement réalisées pourraient composer des listes de valeurs prédéfinies – d'une manière parfaitement similaire aux expressions normalisées interprétant le sens d'un terme conceptuel (voir les exemples ci-dessus, figure 15.3 et 15.4). L'analyste aurait alors soit le choix de réutiliser des expressions forgées par d'autres membres de la communauté d'analyste (en le cochant dans la liste des expressions disponibles) soit d'« inventer » une nouvelle formule pour expliciter un terme conceptuel ou une configuration de termes conceptuels.

Chapitre 16

Les briques configurationnelles des modèles de description

16.1. Introduction

Dans les chapitres précédents, nous avons présenté le *méta-lexique des termes conceptuels** de l'*univers du discours** ASA ainsi que le thesaurus qui fournit tout un ensemble de valeurs déjà prédéfinies pour interpréter tel ou tel terme conceptuel, telle ou telle configuration de termes conceptuels.

Dans ce chapitre, nous allons discuter comment les termes conceptuels d'une part et le thesaurus de l'autre s'intègrent pour former un modèle de description ou encore une bibliothèque de modèles de description propre à une archive audiovisuelle donnée.

Dans la section 16.2, nous reviendrons encore une fois sur le concept du modèle de description – concept qui nous a accompagné tout au long de ce livre.

Les sections 16.3 et 16.4 sont consacrées à une présentation de ce qu'est une *séquence de description*. Nous verrons également que toute séquence de description trouve sa place dans une bibliothèque de séquences de description. Deux points importants que nous discuterons dans ces deux sections concernent d'abord celui de la définition d'une séquence de description à l'aide d'une *paire canonique de schéma de définition** et ensuite celui de la sélection et de l'insertion d'une séquence dans un *modèle de description**.

Enfin, la section 16.5 présentera encore le plan des schémas de définition d'un modèle de description – schémas de définition qui se divisent, comme nous le savons, en deux grandes catégories : les *schémas de définition des objets d'analyse** et les *schémas de définition des activités d'analyse**.

16.2. Analyse d'un texte audiovisuel et modèles de description

Comme déjà dit, un *modèle de description** d'un texte ou d'un corpus de textes audiovisuels est composé d'au moins une, mais en règle générale de plusieurs *séquences** locales. Chaque séquence, à son tour, est composée au moins d'un *schéma* définissant l'*objet d'analyse** (de la séquence de description) et d'un *schéma* définissant la *procédure d'analyse**.

Le schéma définissant l'objet d'analyse, lui, est spécifié par au moins un terme conceptuel désignant le type d'objets ou de domaines faisant partie de l'*univers du discours** ASA. Le schéma spécifiant la *procédure d'analyse** est constitué par au moins une *activité de description** dans l'univers du discours ASA selon laquelle l'*objet d'analyse** doit être indexé (au sens large du terme ; voir chapitres 1 et 14).

On voit bien qu'un modèle de description possède une organisation à la fois *hiérarchique*, *modulaire* et *configurationnelle* :

– *organisation hiérarchique* dans le sens qu'il est composé d'unités plus locales que sont notamment les *séquences* et les *schémas* ;
– *organisation modulaire* dans le sens que chaque séquence et chaque schéma forme une *unité structurale autonome* qui peut faire partie de différents modèles de description (en ce qui concerne les séquences) ou de séquences (en ce qui concerne les schémas de définition) ;
– *organisation configurationnelle* dans le sens qu'un terme conceptuel désignant soit un *type d'objets** à analyser (et relevant donc du méta-lexique des termes conceptuels débutant par le terme de base [Objet d'analyse] soit une *activité d'analyse** (et relevant donc du méta-lexique des termes conceptuels débutant par le terme de base [Procédure d'analyse]) ne peut intégrer un modèle de description qu'en tant qu'*entité positionnée par rapport à toutes les autres entités sélectionnées* par le modèle via ses schémas de définition (d'objets d'analyse et de procédures d'analyse) et ses séquences de description.

Cette organisation très particulière, sophistiquée et contraignante d'un modèle de description est en rapport direct avec notre conception de l'*analyse** d'un texte ou d'un corpus de textes audiovisuels telle que nous l'avons exposée au chapitre 4.

Nous avons, souvenons-nous, distingué entre quatre niveaux principaux d'analyse (voir figure 4.3 ; chapitre 4.4). Le premier niveau, le plus général, est celui du choix d'un certain *type d'analyse*. Dans ce livre, nous avons privilégié l'analyse du contenu audiovisuel qui est un type parmi d'autres. Le deuxième niveau est déterminé par le choix effectué au premier niveau (c'est-à-dire par le choix du type d'analyse). Ce deuxième niveau identifie les différentes *tâches d'analyse* à effectuer pour réaliser un type d'analyse choisi. Par exemple, comme le montre la figure 4.3, l'analyse du contenu audiovisuel connaît, entre autre, la tâche de la *description référentielle** ou encore la tâche de la *description de la mise en discours** d'un objet thématisé dans un texte ou un corpus de textes audiovisuels. Le troisième niveau identifie les *procédures d'analyse* appropriées pour effectuer une tâche d'analyse. Dans ce livre, nous avons donné une priorité à deux procédures très communes et pouvant être utilisées dans une grande diversité de tâches d'analyse. Il s'agit des procédures de *description libre** et de *description contrôlée**. Enfin, le quatrième niveau identifie les *activités d'analyse* spécifiques qui définissent une procédure d'analyse. Ces activités, nous le savons déjà, sont toutes définies dans le méta-lexique des termes conceptuels débutant par le terme de base [Procédure d'analyse].

Cette vision *à quatre niveaux* de l'activité de l'analyse s'exprime en effet directement dans l'organisation hiérarchique, modulaire et configurationnelle d'un modèle de description ASA.

Ainsi, un *type d'analyse* spécifique (= 1er niveau, voir figure 4.3, chapitre 4) se manifeste sous forme d'un ou de plusieurs modèles de description. Par exemple, l'analyse du contenu d'un texte audiovisuel qui est un type d'analyse spécifique, sollicite un ou plusieurs *modèles de description thématique** qui font partie d'une *bibliothèque de modèles de description** – bibliothèque à l'aide de laquelle on analyse, par exemple, les *objets** faisant partie de l'*univers du discours** d'une archive audiovisuelle.

Une *tâche d'analyse* spécifique (= 2ème niveau, voir figure 4.3, chapitre 4) se manifeste sous forme d'une ou de plusieurs séquences de description. Une séquence de description forme, comme nous le savons déjà, une *brique** spécifique qui compose un modèle de description. Ainsi, la tâche de la *description référentielle** qui vise l'identification et l'explicitation du domaine ou de l'objet de connaissance thématisé dans un texte audiovisuel, est supportée par une ou plusieurs séquences fonctionnellement spécialisées pour cette tâche (voir notre exemple au chapitre 3).

Une *procédure d'analyse* (= 3ème niveau, voir figure 4.3, chapitre 4) à son tour est supportée par un ou plusieurs *schémas de définition**. Comme déjà expliqué dans

le chapitre précédent, le schéma de définition d'une *activité de description** compose, avec le *schéma définissant l'objet d'analyse** (c'est-à-dire l'objet à décrire), une séquence de description. Par exemple, le schéma de la procédure de *description libre** peut composer non seulement une séquence fonctionnellement spécialisée en la description référentielle d'un domaine de connaissance mais aussi une séquence fonctionnellement spécialisée en la description de la mise en discours du domaine de connaissance thématisé, une séquence fonctionnellement spécialisée en la description de l'expression audiovisuelle ou verbale du domaine de connaissance thématisé, etc. (voir nos exemples au chapitre 3).

Enfin, une *activité d'analyse* (= 4ème niveau, figure 4.3) est définie dans le *méta-lexique des termes conceptuels** désignant toutes les activités d'analyse de l'univers du discours ASA. Une ou plusieurs activités de description composent un schéma de définition. Ainsi, l'activité [Réaliser une présentation synthétique] peut définir – ensemble avec d'autres activités de description – le schéma de *description libre**, le schéma de *description contrôlée**, le schéma de réalisation d'un commentaire méta-textuel, le schéma d'illustration iconographique, etc.

Un rôle particulièrement important revient ici à deux grands types de configurations (ou modules) qui non seulement composent tout modèle de description mais qui définissent également les rapports spécifiques qu'entretiennent entre eux les *termes conceptuels** sélectionnés par ou dans un modèle de description. Ces deux types de configuration sont :

– d'une part les *séquences de description* et ;
– d'autre part les *schémas* définissant les *objets d'analyse* ainsi que les *activités d'analyse*.

16.3. La bibliothèque de séquences composant le modèle de description thématique

Nous nous limitons ici aux séquences qui composent les *modèles de description thématique**, c'est-à-dire les modèles qui servent à la description du contenu d'un texte ou d'un corpus de textes audiovisuels. Comme nous le savons déjà, les modèles de description thématique forment un type particulier de *modèles de description** parmi d'autres dont, par exemple, les modèles de *description audiovisuelle** *stricto sensu* (c'est-à-dire des plans visuels et/ou sonores), les modèles de *description pragmatique** (d'adaptation d'un texte aux attentes et compétences d'un public visé) ou encore les modèles de (c'est-à-dire de l'identité formelle d'un texte audiovisuel).

Un modèle de description se forme à partir d'un ou de plusieurs champs d'analyse fonctionnellement spécialisés. Chaque champ fonctionnel d'un modèle de description est constitué par une ou plusieurs structures locales appelées *séquence*.

Une séquence sert à la description d'une partie ou d'un aspect spécifique de l'objet à décrire. Toute séquence composant un modèle de description est donc *thématiquement* et *fonctionnellement* spécifiée par rapport à un *champ d'analyse* spécifique. Nous en identifions notamment les cinq suivants :

– le *champ de description référentielle* de l'objet/des objets de connaissance thématisé(s) dans un texte/corpus de textes audiovisuels. C'est ici qu'on identifie et décrit, par exemple et pour revenir aux ateliers d'expérimentation du projet ASA-SHS, telle ou telle culture, telle ou telle œuvre littéraire, telle ou telle fouille archéologique, telle ou telle discipline scientifique, telle ou telle personnalité scientifique, etc. ;

– le champ de *contextualisation référentielle du ou des objets de connaissance thématisés* dans le texte/corpus de textes audiovisuels. Ce champ se différencie, en un champ de description appelée *Localisation spatiale et/ou géographique* du domaine de connaissance, en un champ de description appelée *Localisation temporelle et historique* du domaine de connaissance et parfois encore en un champ appelée *Contextualisation thématique* (c'est-à-dire *institutionnelle, sociale, épistémique*,...) du domaine de connaissance ;

– le champ d'analyse de la *mise en discours des objets de connaissance thématisés*. Ce champ d'analyse permet à l'analyste de fournir des renseignements sur les particularités du traitement énonciatif et discursif d'un sujet donné dans une vidéo : qui est le responsable (c'est-à-dire l'énonciateur) de la thématisation de l'objet ou domaine de connaissance ?, selon quel point de vue est-il traité ?, quelle est la hiérarchisation discursive de l'information (c'est-à-dire quelles informations sont les plus importantes, les plus centrales et quelles informations sont moins importantes, moins centrales) ?, quel est le niveau de spécialisation dans le traitement de l'objet thématisé ?, quelles informations sont supposées connues et quelles informations sont traitées comme des nouveautés ?, etc. ;

– le champ d'analyse de l'*expression verbale* (peu développé dans le cadre de notre recherche) et de l'*expression audiovisuelle* des objets de connaissance thématisés et mis en discours. Ce champ permet à l'analyste de fournir des renseignements sur les particularités significatives du langage d'une part *verbal* (mais aussi gestuel, postural, somatique) et d'autre part *audiovisuel* utilisés pour exprimer et communiquer le discours sur un objet ou un domaine de connaissance ;

– le champ d'analyse *méta-textuel du commentaire*. Ce champ donne la possibilité à l'analyste d'exprimer plus explicitement son point de vue soit sur son analyse soit sur la façon dont comment est traitée le domaine dans la vidéo source.

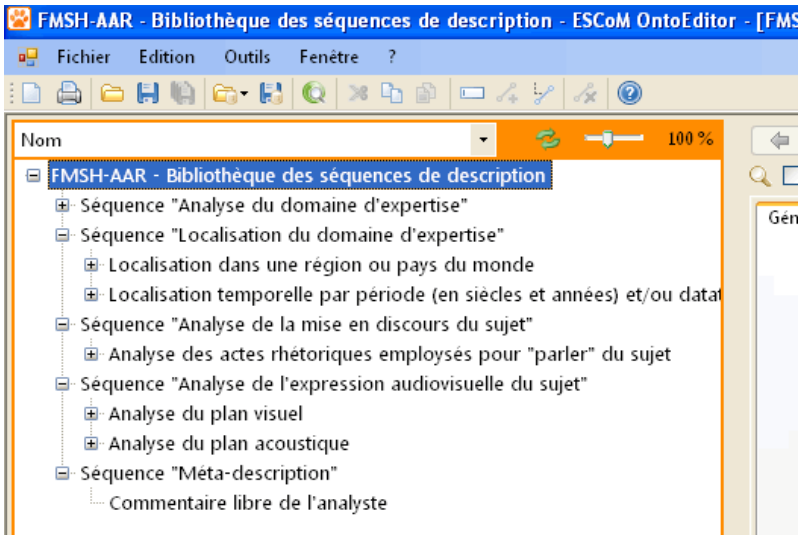

Figure 16.1. *Extrait de la bibliothèque des séquences composant les modèles de description du contenu audiovisuel de l'univers du discours des archives FMSH-AAR*

Retenons surtout le fait qu'à un *champ d'analyse* correspond toujours une *bibliothèque de séquences fonctionnellement appropriées*. La figure 16.1 nous montre l'extrait d'une telle *bibliothèque de séquences* que nous utilisons, dans notre cas, pour définir la *bibliothèque des modèles de description du contenu* des textes audiovisuels qui font partie du fonds des archives audiovisuelles FMSH-AAR consacrées à la production et diffusion d'événements scientifiques de la Fondation Maison des Sciences de l'Homme à Paris.

Comme on peut le voir dans la figure 16.1, la bibliothèque des séquences est composée de 5 grandes collections, à savoir :

– une collection de séquences intitulée *Analyse du domaine d'expertise* dont la figure 16.2 nous propose une vue plus détaillée. Cette collection réunit toutes les séquences fonctionnellement spécialisées en l'identification et l'explicitation (description, indexation, etc.) des objets et domaines de connaissance thématisables dans la production audiovisuelle des archives FMSH-AAR. Autrement dit, cette collection constitue les ressources nécessaires pour définir le *premier* champ fonctionnel d'un modèle de description thématique (du contenu audiovisuel), c'est-à-dire celui réservé à la tâche de la description référentielle (dans le cadre de l'analyse du contenu audiovisuel qui forme, répétons-le encore une fois, un *type particulier* d'analyses du texte ou d'un corpus de textes audiovisuels). Pour le moment, la collection en question de séquences spécifie des sujets référentiels tels que *Recherche scientifique en SHS*, *Visions, idéologies et représentations*, *Identités*

sociales ou encore *Mémoires collectives*. La collection des séquences représentant ces champs référentiels d'analyse spécialisés, peut être, à tout moment, complétée par de *nouvelles séquences*. L'*univers du discours** d'une archive n'étant pas figé, tout dépend ici de la *spécificité* du corpus ou des corpus en jeu et de la *politique de diffusion* et de *valorisation* du patrimoine scientifique (c'est-à-dire à la réponse à des questions du genre : « Quels sujets, quelles régions géopolitiques, quelles époques, etc. à mettre en avant pour quel public et quel contexte d'usages ? ») ;

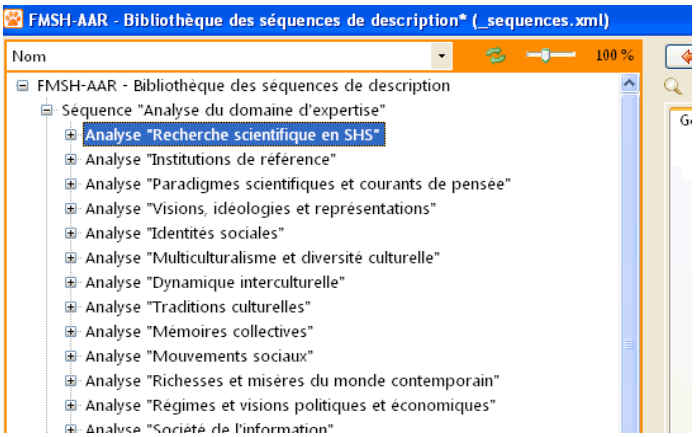

Figure 16.2. *Vue plus détaillée de la collection des séquences* « Analyse du domaine d'expertise » *(exemple : l'univers du discours des archives FMSH-AAR)*

– une collection de séquences intitulée *Localisation du domaine d'expertise* (voir figure 16.1). La figure 16.3 nous montre son organisation dans le cadre de la définition du *métalangage de description** de l'*univers du discours** du fonds audiovisuel composant les archives FMSH-AAR. Comme on peut le constater, cette collection est divisée en un premier ensemble de séquences utilisées pour la localisation spatiale (géographique ou géopolitique) d'un objet ou domaine de connaissance thématisé et en un deuxième ensemble de séquences utilisées pour la localisation temporelle (chronologique). Comme nous l'avons déjà dit ci-dessus, rien n'empêche de spécifier encore davantage cette collection afin de pouvoir encore mieux localiser un objet ou domaine de connaissance traité dans un texte audiovisuel ;

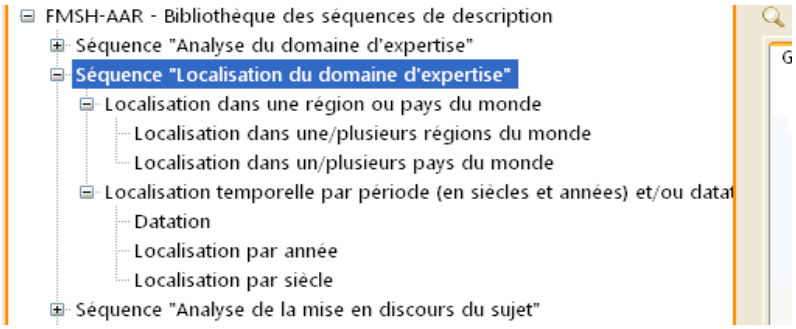

Figure 16.3. *Vue plus détaillée de la collection des séquences « Localisation du domaine d'expertise » (exemple : l'univers du discours des archives FMSH-AAR)*

– une collection de séquences intitulée *Analyse de la mise en discours du sujet* (figure 16.1). Peu développée dans le cadre de la mise en place du métalangage de description des textes audiovisuels faisant partie des archives FMSH-AAR, cette collection ne connaît qu'un seul ensemble de séquences que nous avons définies pour mieux identifier certains actes de discours utilisés pour « parler » d'un objet ou d'un domaine de connaissance (voir figure 16.1). Dans notre exemple, il s'agit surtout de permettre à l'analyste d'identifier les *actes de définition* et *d'exemplification* (de telle ou telle notion, de tel ou tel objet, de telle ou telle expression, etc.) dans un texte ou un corpus de textes audiovisuels. Les résultats d'une telle analyse peuvent être utilisés pour la mise en place, sur le portail des archives FMSH-AAR, d'interfaces spécialisées d'accès à des passages audiovisuels dans lesquels le public intéressé trouvera des définitions et des exemples concrets sur une problématique donnée ;

– une collection de séquences intitulée *Analyse de l'expression audiovisuelle du sujet* (figure 16.1). Comme le montre la figure 16.1, cette collection n'est développée que pour l'analyse de la mise en scène audiovisuelle d'un objet ou d'un domaine thématisé : point de vue visuel, cadrage et plans visuels, cadre acoustique, etc. ;

– enfin une collection de séquences réservées au métadiscours de l'analyste – collection qui ne se compose (dans notre exemple) que d'une seule séquence permettant à l'analyste d'ajouter un commentaire libre à sa description de tel ou tel sujet.

Toute bibliothèque de séquences de description est organisée d'une manière identique à l'exemple que nous venons de présenter brièvement ci-dessus. N'oublions pas cependant que nous avons développé ici seulement les collections de séquences dont nous avons besoin pour spécifier les modèles de description permettant de réaliser un type d'analyse bien spécifique. Il s'agit de la *description*

*du contenu** audiovisuel (voir la figure 4.3 au chapitre 4 qui identifie pour ce type d'analyse les principales tâches de description à réaliser). Pour que la bibliothèque des séquences soit complète, il faut ajouter encore les séquences et collections de séquences dont on peut avoir besoin pour réaliser les tâches des autres types principaux d'analyse que sont la *description paratextuelle**, la *description pragmatique** ou encore la *description audiovisuelle* stricto sensu*.

Toute séquence peut être réemployée dans différentes bibliothèques de modèles de description. Autrement dit, la *même séquence* peut servir à la définition du métalangage de description d'univers du discours de *différentes* archives audiovisuelles. Par exemple, la séquence *Analyse « Recherche Scientifique en SHS »* est une séquence que nous n'utilisons pas seulement pour décrire des corpus faisant partie des archives FMSH-AAR mais aussi pour décrire ceux d'autres archives pour lesquelles le domaine de la recherche scientifique en SHS possède une pertinence.

Cette précision renvoie directement à l'importance de bien distinguer entre les *différents niveaux d'intégration* qui caractérisent le *métalangage de description** de l'*univers du discours** d'une archive audiovisuelle. Ainsi, faut-il distinguer entre :

– le niveau des *termes conceptuels* et le niveau des *schémas d'indexation* pour sélectionner et renseigner un terme (ou une configuration de termes) conceptuels (voir chapitre 14) ;
– le niveau des *schémas de définition* se différenciant en des *schémas de définition des objets d'analyse* et des *schémas de définition des activités d'analyse* (voir ci-après) ;
– le niveau des *séquences d'analyse* et celui des *collections de séquences d'analyse* ;
– le niveau des *modèles de description* (thématique, audiovisuelle, paratextuelle, méta-textuelle, etc.) ;
– ainsi que, enfin, le niveau de la *collection de modèles de description* et de la *bibliothèque* stricto sensu* des *modèles* définissant le *métalangage de description** de l'*univers du discours** d'une archive audiovisuelle donnée.

16.4. Définition et insertion d'une séquence dans un modèle de description

Toute séquence forme un objet structural et fonctionnel qui possède son *organisation interne*. Voyons rapidement l'exemple de la séquence *Analyse « Recherche Scientifique en SHS »* (figure 16.4).

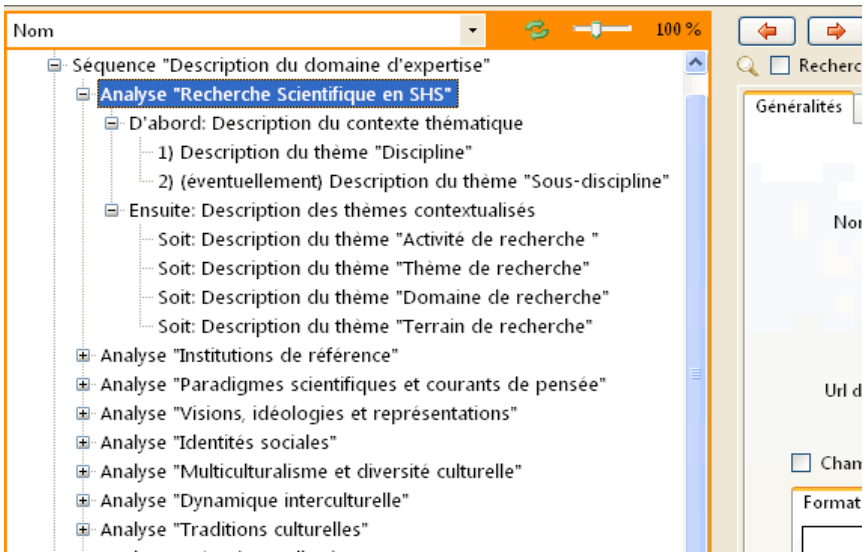

Figure 16.4. *Définition syntagmatique de la séquence* Analyse « Recherche Scientifique en SHS » *(exemple : l'univers du discours des archives FMSH-AAR)*

Comme nous le montre la figure 16.4, une séquence peut se déployer en plusieurs sous-séquences plus spécialisées qui se suivent les unes les autres selon un *ordre syntagmatique* explicite.

La relation syntagmatique, dans notre exemple, est celle de la *présupposition unilatérale* que nous avons déjà rencontrée à plusieurs reprises. Elle prescrit à l'analyste de renseigner d'abord le contexte de l'objet ou du domaine thématisé dans le texte audiovisuel. Dans notre cas, il s'agit de préciser la ou les disciplines scientifiques dont on parle dans un texte audiovisuel (voir dans la figure 16.4, la séquence intitulée *Description du thème « Discipline »*). Une fois cette partie du sujet renseignée, l'analyste peut, dans une deuxième étape, éventuellement préciser certains points de l'objet ou du domaine thématisé. Dans notre cas, il est proposé à l'analyste de fournir, si disponible, des informations sur les activités de recherche, le ou les domaines de recherche ou encore le ou les thèmes de recherche (voir dans la figure 16.4, les séquences intitulées *Description du thème « Activité de recherche »*, *Description du thème « Domaine de recherche » ; Description du thème « Thème de recherche »*).

Retenons de notre discussion rapide de l'exemple de la description référentielle du domaine de la *recherche en sciences humaines et sociales* qu'une séquence se présente, dans son organisation interne la plus simple, comme *une seule entité* composée, comme nous le verrons tout de suite ci-après, d'au moins deux schémas

de définition, c'est-à-dire un schéma de définition de l'*objet d'analyse** et un schéma de définition de l'*activité d'analyse**. Mais elle peut également se déployer en tout un réseau de sous-séquences qui se positionnent les unes par rapport aux autres selon une structure syntagmatique précise.

Enfin, une organisation syntagmatique de séquences telle que celle représentée par notre exemple (figure 16.4) peut être complétée par des nouvelles sous-séquences qui constituent soit des *alternatives* à une sous-séquence déjà identifiée et définie dans l'organisation syntagmatique en place soit encore des *compléments*, des *ajouts*.

Figure 16.5. *Définition d'une séquence par une paire canonique appropriée de schémas conceptuels de définition*

Prenons le cas de la sous-séquence *Description du thème « Activité de recherche »*. Celle-ci est définie et positionnée comme une séquence simple, comme une entité indivise dans la *configuration des séquences* à l'aide de laquelle nous pouvons analyser des sujets consacrés à la recherche en sciences humaines et sociales (figure 16.4). Rien n'empêche cependant de définir des séquences alternatives à celle-ci – séquences alternatives qui permettent d'adapter le travail de l'analyse à des spécificités d'un corpus ou d'une une archive audiovisuelle donnée. Lorsque la macro-séquence intitulée *Analyse « Recherche Scientifique en SHS »* sera sélectionnée pour figurer dans un modèle de description thématique, alors ne sera pas sélectionnée la sous-séquence *Description du thème « Activité de recherche »* telle qu'elle se présente dans la figure 16.4 mais, au contraire, l'une ou l'autre de ses alternatives. Ces remarques s'appliquent également au cas où il faut ajouter une ou plusieurs séquences spécialisées à une configuration existante de séquences.

Toute séquence – simple ou composée de deux ou plusieurs séquences plus spécialisées – est une *entité relationnelle* qui, d'une part, positionne les *schémas conceptuels de définition** les uns par rapport aux autres et qui, d'autre part, se positionne par rapport à d'autres séquences pour former un *modèle de description**.

Comme déjà dit à plusieurs reprises, toute séquence est définie par au moins deux, sinon plusieurs schémas conceptuels de définition. La figure 16.5 nous montre comment on définit, dans notre cas via l'interface de l'outil *OntoEditeur* de l'Atelier de Modélisation ASA, une séquence de description donnée en identifiant le pair canonique approprié des schémas de définition.

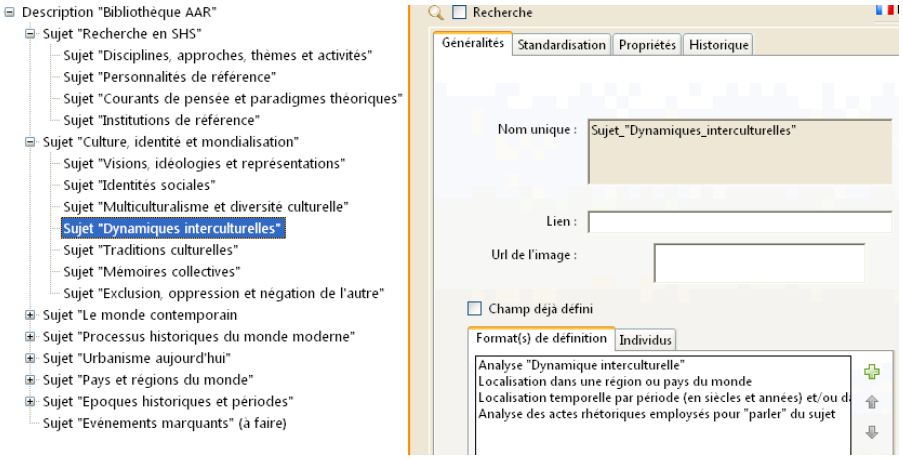

Figure 16.6. *Place de la séquence « Analyse « Dynamique interculturelle » dans un modèle de description thématique faisant partie de la bibliothèque des modèles de description des archives FMSH-AAR)*

Dans notre cas, la séquence *D'abord : Description du thème principal* (qui fait partie de la macro-séquence *Analyse « Dynamique interculturelle »*) est ainsi définie :

– par *un schéma conceptuel* spécifiant les *objets d'analyse appropriés* (voir figure 16.5, la partie droite de l'image où figure le lien vers le schéma *Choix du/des TC approprié(s)*) et d'autre ;

– par un *schéma conceptuel* spécifiant la *procédure de description* qui est ici celle de la *description libre** dans sa version simplifiée (c'est-à-dire l'analyste ne doit renseigner qu'*un seul* champ d'indexation qui est celui de l'expression minimale ; voir pour plus d'informations, le chapitre 14).

La figure 16.6 nous montre le processus de l'intégration d'une séquence (dans notre cas, d'une macro-séquence) dans un modèle de description thématique spécifique. Dans notre cas, la macro-séquence *Analyse « Dynamique interculturelle »* est intégrée ensemble avec trois autres macro-séquences pour former le modèle de description du contenu *Sujet « Dynamiques interculturelles » :*

– la macro-séquence en question sert ici à l'explicitation (description, interprétation, etc.) d'un fait considéré (par l'analyste, par l'auteur du texte analysé, etc.) comme représentant un cas concret de la dynamique interculturelle ;

– deux macro-séquences servent à localiser le fait analysé (il s'agit de la macro-séquence *Localisation dans une région ou pays du monde* et de la macro-séquence *Localisation temporelle par période et/ou date précise*) ;

– et, enfin, une macro-séquence sert à décrire la mise en discours du sujet (il s'agit de la macro-séquence *Analyse des actes rhétoriques employés pour parler du sujet*).

16.5. Présentation synthétique d'une bibliothèque de schémas de définition

Nous le savons déjà, une séquence est composée par au moins deux schémas de définition dont l'un doit appartenir à la catégorie des schémas *Objet d'analyse* et l'autre à la catégorie *Procédure d'analyse*.

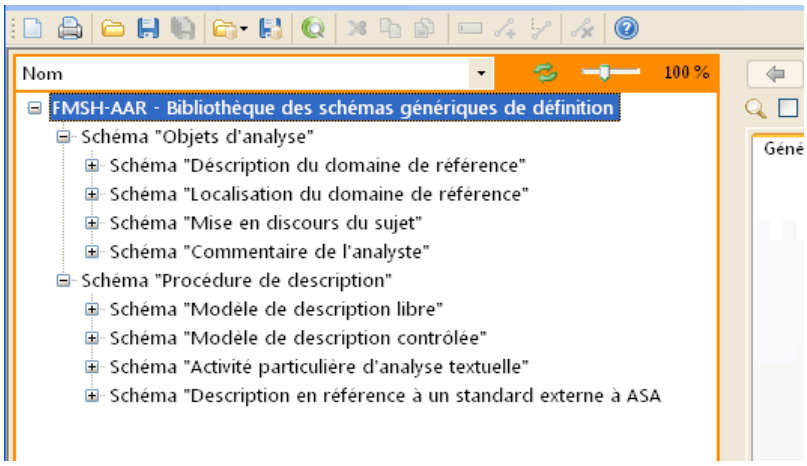

Figure 16.7. *Extrait de la bibliothèque des schémas de définition des archives audiovisuelles FMSH-AAR*

Chaque archive possède sa bibliothèque de schémas. Ainsi, la figure 16.7 nous montre l'extrait de la bibliothèque des schémas de définition pour les archives FMSH-AAR. Celle-ci est composée de deux grandes collections correspondant aux deux catégories principales des schémas déjà citées.

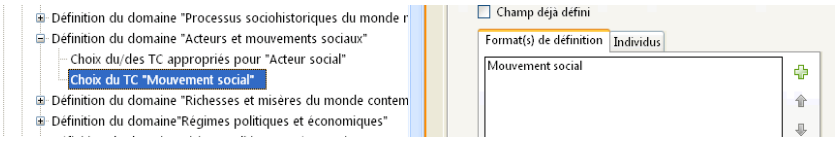

Figure 16.8. *Schéma de définition à structure interne simple*

L'organisation interne des schémas en *collections de schémas* plus spécialisées correspond, en ce qui concerne les schémas de définition des objets d'analyse, aux différents types d'*objets** pertinents à la description d'un texte ou d'un corpus de textes audiovisuels :

– *objets référentiels* (c'est-à-dire faisant partie de tel ou tel domaine empirique documenté par un corpus audiovisuel) ;
– *objets contextuels* (de localisation spatiale, temporelle, etc.) ;
– *objets discursifs* (c'est-à-dire de mise en discours) ;
– *objets verbaux et audiovisuels* (c'est-à-dire d'expression visuelle, acoustique, syncrétique, etc.) ;
– *objets réflexifs* (c'est-à-dire servant à produire un métadiscours dont l'analyse est l'objet).

Une bibliothèque de schémas de définition réunit tous les schémas qui interviennent dans la constitution des *modèles de description**. Mais, en règle générale, elle est pourvue de bien plus des schémas de définition qui sont potentiellement pertinents à l'élaboration de nouveaux modèles de description ou encore à celle de modèles de description « alternatifs » par rapport à ceux qui existent et qui font partie d'une *bibliothèque de modèles de description** définissant l'*univers du discours** d'une archive audiovisuelle.

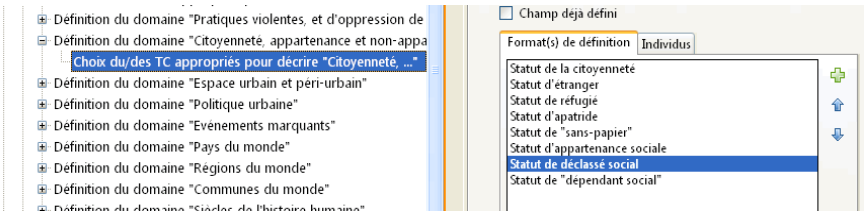

Figure 16.9. *Un schéma de définition réunissant une liste de termes conceptuels désignant l'objet ou le domaine d'analyse*

Voyons maintenant l'organisation interne d'un schéma de définition. Tout schéma de définition est composé, comme déjà dit à plusieurs reprises, d'un ou de plusieurs *termes conceptuels** qui proviennent soit du méta-lexique des termes conceptuels désignant les objets d'analyse de l'univers du discours ASA soit du méta-lexique des termes conceptuels désignant les activités d'analyse.

Le schéma de définition le plus simple se caractérise par la sélection d'un et d'un seul terme conceptuel. Cette sélection équivaut à l'affirmation d'une r*elation unaire d'identification* d'un terme conceptuel. La figure 16.8 nous en fournit un exemple. Le schéma définissant l'objet d'analyse *Mouvement social* s'appuie uniquement et exclusivement sur le terme conceptuel [Mouvement social].

Mais un schéma peut également sélectionner plusieurs termes conceptuels peu importe si ces derniers appartiennent au même domaine taxinomique du méta-lexique des termes conceptuels ou pas. Un exemple nous est fourni dans la figure 16.9. Ce schéma définit un objet d'analyse intitulé *Citoyenneté, appartenance et non-appartenance sociale*. Pour cela, ce schéma s'appuie sur un ensemble de termes conceptuels tels que [Statut de la citoyenneté], [Statut d'étranger], [Statut d'apatride], etc. parmi lesquels l'analyste pourra choisir le ou les termes les plus appropriés pour indexer un texte audiovisuel qui traite de la question de la citoyenneté et/ou de l'appartenance/non-appartenance sociale.

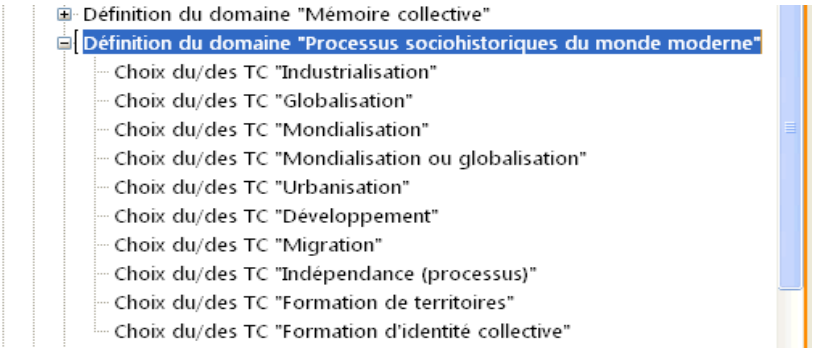

Figure 16.10. *Un schéma de définition réunissant une liste de termes conceptuels désignant l'objet ou le domaine d'analyse*

Un troisième cas de figure, enfin, est celui du « grand schéma de définition (d'un *objet d'analyse** ou d'une *procédure d'analyse**) » ou, comme nous préférons l'appeler, du *macro-schéma de définition* (d'une manière analogue à une macro-séquence). Un macro-schéma est composé de plusieurs schémas plus spécialisés qui se positionnent les uns par rapport aux autres. Chaque schéma plus spécialisé comporte au moins un terme conceptuel (mais, comme le montre l'exemple de la figure 16.9, un schéma peut également comporter plusieurs termes conceptuels). La figure 16.10 nous montre la décomposition du schéma de définition de l'objet d'analyse *Processus sociohistoriques du monde moderne* en toute une série de schémas plus spécialisés : *schéma Globalisation*, *schéma Urbanisation*, *schéma Migration*, etc.

Le macro-schéma *Processus sociohistoriques du monde moderne* peut être sélectionné tel quel dans une séquence de description référentielle. Dans ce cas, il propose à l'analyste toute une série d'options pour configurer la structure de sa topique (voir chapitre 5 pour plus d'explications). On peut également ne sélectionner que certains schémas composant un macro-schéma de définition tel que celui représenté par la figure 16.9.

Conclusion et perspectives

Arriver à terme de ce livre, essayons de résumer en quelques lignes les principaux points et enjeux de notre travail.

Face à la prolifération de données audiovisuelles numériques, une des questions centrales qui se posent est celle de savoir comment en faire des *ressources* (intellectuelles) *sui generis*. Autrement dit, ce n'est pas le fait qu'un document audiovisuel soit disponible sous forme numérique pour qu'il devienne déjà un bien (culturel) pour un public donné !

En effet, il est une chose de mettre en place une vidéothèque (ou une médiathèque) numérique pourvue d'un fonds filmique ou sonore qui documente, par exemple, les activités d'une institution scientifique ou d'enseignement, il en est une autre de transformer ce fonds en des ressources pour la recherche, l'enseignement, la valorisation de l'image de l'institution, etc. Ces deux activités n'ont rien à voir l'une avec l'autre et aussi longtemps que l'on considère que la « simple » numérisation, et mise en ligne, suffit pour régler la question de la constitution, communication et exploitation de patrimoines de connaissances, aussi longtemps on fera fausse route.

Dans ce livre nous avons essayé de traiter à la fois d'un point de vue théorique que pratique et technique la question de la *transformation* d'une « simple » donnée numérique en une ressource pour un public ou un utilisateur. Nous interprétons cette

transformation comme un processus d'*appropriation* d'une donnée numérique par le public ou l'utilisateur individuel intéressé. L'appropriation peut concerner soit la matérialité d'une donnée numérique existante (l'utilisateur modifie celle-ci en ayant recours, par exemple, à l'opération du montage), soit son contenu (*lato sensu*). Cette deuxième forme d'appropriation est appelée *analyse** au sens le plus large du terme : l'utilisateur, le public décrit, interprète, commente, amende, adapte, etc., le contenu de la donnée numérique ou d'une partie (d'une séquence, d'un segment) de celle-ci.

Dans ce livre, nous nous sommes donc concentrés sur la problématique de l'analyse d'un texte ou d'un *corpus* de textes audiovisuels. Nous avons tracé rapidement le cadre théorique qui nous guide, qui détermine notre approche de l'analyse du texte (audiovisuel), nous avons essayé également d'identifier différents types d'analyse comme par exemple, l'*analyse thématique**, c'est-à-dire du contenu au sens stricte du terme (du sens, du message, etc.) d'un texte audiovisuel ; l'*analyse audiovisuelle**, c'est-à-dire l'analyse de formes d'expression visuelle et/ou acoustique d'un texte audiovisuel ; ou encore l'*analyse pragmatique**, c'est-à-dire de l'adaptation d'un texte audiovisuel à un contexte d'usage particulier.

Ensuite, nous avons essayé de définir et de développer un ensemble de ressources dites métalinguistiques dont nous avons besoin pour spécifier et développer les *modèles de description** qui sont nécessaires à tel ou tel type d'analyse. En nous intéressant à des corpus audiovisuels spécifiques composant une archive audiovisuelle, nous avons enfin élaboré des bibliothèques de modèles de description spécifiques avec, comme hypothèse, que chaque bibliothèque de modèles décrit plus ou moins bien, plus ou moins partiellement (voire partialement, c'est-à-dire selon le point de vue et l'intérêt propre à un analyste) l'*univers du discours** d'une archive donnée.

L'organisation conceptuelle qui sous-tend la démarche proposée dans ce livre est en effet assez complexe et fait appel à toute une série de ressources métalinguistiques spécifiques : méta-lexiques de termes conceptuels désignant les objets d'analyse de l'univers du discours d'une archive ; méta-lexique désignant les activités d'analyse pour décrire tel ou tel objet ou telle ou telle configuration d'objets ; bibliothèque de schémas d'indexation ; bibliothèque de briques génériques pour construire des modèles de description ; bibliothèques de modèles de description ; thesaurus.

Mais la tradition de la recherche sur le texte, son analyse, son interprétation, sa « réécriture », montre bien qu'il s'agit ici d'une problématique complexe qui ne peut être ni réduite à une simple question sans intérêt et « archi-connue » qui est celle de l'indexation par mots clé ni encore déléguée à des hypothétiques avancées révolutionnaires de l'indexation automatique qui devrait régler quelque chose qui nous semble être totalement hors de sa portée, c'est-à-dire l'intérêt humain (la curiosité, etc.) qui précède toute analyse ainsi que la dimension herméneutique qui est propre à toute activité d'analyse textuelle.

Ceci étant, nous sommes également parfaitement conscients des limites de notre approche au moins telle qu'elle se présente dans ce livre. Les deux méta-lexiques qui constituent le « cœur » du système métalinguistique ASA ne sont pas encore stabilisés et leur emploi pour définir et développer des modèles de description permettant l'analyse d'univers de discours différents de ceux que nous avons traités dans notre recherche, reste problématique. La bibliothèque des séquences d'analyse de la mise en discours est encore relativement pauvre - tout comme celle des séquences réservées à la description de l'expression verbale ou audiovisuelle. Aussi les modèles de description dont nous disposons pour réaliser des analyses paratextuelles (de type Dublin Core) sont très figés et difficilement adaptables – tout comme ceux dont nous disposons actuellement pour procéder à une analyse pragmatique et à des adaptation/traductions de données audiovisuelles.

D'autres limites de notre approche concerne une prise en compte encore assez timide des approches et techniques « externes » susceptibles d'apporter de nouvelles solutions. Nous pensons, par exemple, à une intégration bien plus poussée des différents standards dans l'approche ASA, la mise en relations systématique avec d'autres ontologies et thesaurus, l'utilisation aussi des techniques de *speech to text* et de la *fouille du texte* pour à la fois contribuer à l'enrichissement de nos ressources métalinguistiques et à une sorte de pré-description/pré-indexation de textes audiovisuels à l'aide de modèles de description ASA.

Enfin, notre intérêt dans la conception et réalisation d'un système d'analyse de corpus audiovisuels a toujours était motivé par le souhait de le mettre à la disposition de tout acteur individuel ou collectif, notamment du monde « académique ». L'idée – pour parler ainsi – directrice qui nous guidait tout au long de ces douze années de recherche et de développement consacrées aux archives audiovisuelles, était celle de la constitution d'un *Campus d'archives audiovisuelles de la recherche et de l'enseignement* – une sorte de « You Tube académique », une coopérative mise à la disposition à toute partie prenante dans le processus de la constitution, l'analyse, la diffusion et le partage de patrimoines culturels ou scientifiques sous forme d'archives audiovisuelles personnelles ou collectives.

Notre souhait est, bien sûr, de pousser notre recherche dans ces directions citées et d'en faire profiter les acteurs intéressés. Mais, comme nous le savons, ce type de recherche ne peut pas se faire seul. Elle exige non seulement des ressources humaines, techniques et financières importantes mais aussi un cadre propice à la recherche scientifique, une « tranquillité d'esprit » – conditions qui, comme nous le craignons, sont devenues hélas utopiques, aujourd'hui…

Bibliographie

[ADA 90]ADAM J.-M., *Eléments de linguistique textuelle*, Bruxelles, Editions Mardaga, 1990.

[ADP 89] ADAM, J.-M., PETITJEAN A., *Le texte descriptif*, Paris, Editions Nathan, 1989.

[AST 02] ASTON G., « The learner as corpus-designer », *in : KETTEMANN B. et MARKO G. (eds.), Teaching and learning by doing corpus analysis*, Editions Rodopi B.V., p. 9-25, 2002.

[AUS 70] AUSTIN J.L., *Quand dire c'est faire*, Paris, Le Seuil, 1970.

[BAC 05] BACHIMONT B., « Construction de la « top ontology » de Ménélas. Entre physique et métaphysique », *document en ligne, 2005*. (http://www.spim.jussieu.fr/doc/ontologies/Bachimont-SticSante-08122005.pdf)

[BCU 07] BIBER D., CONNOR U., UPTON T.A., « Discourse on the Move », *Using corpus analysis to describe discourse structure*, John Benjamins B.V., 2007.

[BRA 07] BRAUN S., « Designing and exploiting small multimedia corpora for autonomous learning and teaching », *in: HIDALGO E, QUEREDA L. et SANTANA J. (eds.), Corpora in the Foreign Language Classroom*, Editions Rodopi B.V., p. 31-46, 2007.

[BRG 86] BRONDAL V., GREIMAS A.J., « Omnis et totus. Comment définir les indéfinis ? », in : *Actes Sémiotiques*, Documents 72, 1986.

[BRI 97] BRINKER K., *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methode*, Erich Schmidt Verlag, 1997.

[CHE 11a] CHEMOUNY, M., *Atelier d'analyse de corpus audiovisuels, in : P. Stockinger (éd.),* Les archives audiovisuelles, Hermes Science Publishing, Paris, p. 71-89, 2011.

[CHE 11b] CHEMOUNY, M., *L'analyse d'un corpus audiovisuel sur Mille et Une Nuits ; in : P. Stockinger (éd.),* Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, Paris, p. 25-43, 2011.

[CHM 02] CHARAUDEAU P., MAINGUENAU D., *Dictionnaire d'analyse du discours*, Paris, Editions du Seuil, 2002.

[CRA 08] CRAVEN L. (ÉD.), *What are Archives. Cultural and Theoretical Perspectives : A Reader*, Ashgate, 2008.

[DED 11] DE PABLO E., DESLIS J., *L'analyse de l'expression audiovisuelle ; in : P. Stockinger (éd.),* Les archives audiovisuelles, Hermes Science Publishing, Paris, p. 89-109, 2011.

[DEP 11a] DE PABLO E., *Atelier de segmentation des ressources audiovisuelles, in : P. Stockinger (éd.),* Les archives audiovisuelles, Hermes Science Publishing, Paris, p. 55-71, 2011.

[DEP 11b] DE PABLO E., *L'analyse d'un corpus audiovisuel consacré à la fabrication traditionnelle du pain ; in : P. Stockinger (éd.),* Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, Paris, p. 43-61, 2011.

[DES 11a] DESLIS, J., *La communication scientifique via Facebook et Twitter ; in : P. Stockinger (éd.),* Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, 2011, p. 157 – 183

[DES 11b] DESLIS J., *Usages de plateformes de diffusion de contenus numériques ; in : P. Stockinger (éd.),* Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, Paris, p. 183-203, 2011.

[DES 11c] DESLIS J., *Usages des agrégateurs de contenus et des réseaux communautaires ; in : P. Stockinger (éd.),* Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, Paris, p. 203-221, 2011.

[DES 11d] DESLIS J., *Maquette d'une publication audiovisuelle sous forme d'un portail web ; in : P. Stockinger (éd.),* Les archives audiovisuelles, Hermes Science Publishing, Paris, p. 167-193, 2011.

[DPS 11] DE PABLO E., STOCKINGER P., *Le republication d'une ressource audiovisuelle ; in : P. Stockinger (éd.),* Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, Paris, p. 61-83, 2011.

[ECO 79] ECO U., *Lector in fabula,* Milano, Bompiano, 1979.

[EI 95] EICHLER E., HILTY G., LOFFLER H., STEGER H., ZGUSTA L. (ÉDS.), *Namenforschung. Name Studies. Les Noms Propres. Ein internationales Handbuch zur Onomastik 1 et 2*, de Gruyter, 1995-1996.

[ESE 10] EUROPEANA SEMANTIC ELEMENTS SPECIFICATION. EUROPEANA V1.0, EUROPEAN UNION 2010.

[GEE 86] GEERTZ C., *Savoir local, savoir global. Les lieux du savoir,* PUF, PARIS, 1986.

[GEN 87] GENETTE G., *Seuils,* Le Seuil, Paris, 1987.

[GIN 86] GINZBURG C., *Mythes, emblèmes, traces. Morphologie et histoire,* Paris, Editions Verdier, 1986.

[GRE 66] GREIMAS A.J., *Sémantique structurale. Recherche de méthode,* Paris, Larousse, 1966.

[GRE 70] GREIMAS A.J., *Du Sens,* Le Seuil, Paris, 1970.

[GRE 76] GREIMAS A.J., *Maupassant, la sémiotique du texte : exercice pratique,* Le Seuil, Paris, 1976.

[GRE 79] GREIMAS A.J., Courtes J., *Sémiotique. Dictionnaire raisonné de la théorie du langage,* Hachette, Paris, 1979.

[GRE 83] GREIMAS A.J., *Du Sens II. Essais Sémiotiques,* Le Seuil, Paris, 1983.

[GRU 98] GRUNIG B.N., GRUNIG R., *La fuite du sens,* Paris, Hatier, 1998.

[GUE 11] GUERINET R., *Le développement informatique du modèlede publication « Portail web » ; in : P. Stockinger (éd.),* Les archives audiovisuelles, Hermes Science Publishing, Paris, p. 249 – 269, 2011.

[HAB 81] HABERMAS J., *Theorie des kommunikativen Handelns, vol 1 et 2,* Suhrkamp Verlag, 1981.

[HQS 07] HIDALGO E., QUEREDA L., SANTANA J., *Corpora in the Foreign Language Classroom,* Editions Rodopi B.V., 2007.

[ISA 04] ISAAC A.,TRONCY R. (eds.), *Designing and Using an Audio-Viusal Description Core Ontology*, Institut National de l'Audiovisuel (INA), en ligne 2004 (http://citeseerx.ist.psu.edu/viewdoc/download ?doi=10.1.1.97.9559&rep=rep1&type=pdf).

[JAC 99] JAWORSKI A., COUPLAND N., *The Discourse Reader,* London, Routledge & Keagan, 1999.

[KEM 02] KETTEMANN B., MARKO G. (eds.), *Teaching and learning by doing corpus analysis*, Amsterdam, Editions Rodopi B.V., 2002.

[LEG 11] LEGRAND-GALARZA V., *Une archive sur le patrimoine immatérieldes populations andines au Pérou et en Bolivie ; in : P. Stockinger (éd.)*, Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, Paris, p. 83-123, 2011.

[LEL 11] LEGRAND-GALARZA V., LEMAITRE F., *Tracer les usages d'une vidéo : les promesses de la technologie VDI ; in : P. Stockinger (éd.)*, Nouveaux usages des archives audiovisuelles numériques, Hermes Science Publishing, Paris, p. 212-254, 2011.

[LEM 11a] LEMAITRE F., *L'environnement numérique ASA in : P. Stockinger (éd.)*, Les Archives Audiovisuelles, Hermes Science Publishing, Paris, p. 195-226, 2011.

[LEM 11b] LEMAITRE F., *Le Studio ASA in : P. Stockinger (éd.)*, Les Archives Audiovisuelles, Hermes Science Publishing, Paris, p. 227-246, 2011.

[LEV 58] LEVI-STRAUSS C., *Anthropologie structurale I*, Paris, Plon, 1958.

[LEV 62] LEVI-STRAUSS C., *La pensée sauvage*, Paris, Plon, 1962.

[LOT 72] LOTMAN J.M., *Die Struktur literarischer Texte*, München, Wilhelm Fink Verlag, 1973.

[MAC 02] MAINGUENEAU P., CHARAUDEAU P., *Dictionnaire d'Analyse du discours*, Editions du Le Seuil, Paris, 2002.

[MAI 96] MAINGUENEAU P., *Les termes clés de l'analyse du discours*, Editions Le Seuil, Paris, 1996.

[MCK 06] MC KAY S., *Researching second language classrooms*, Hilldale, Lawrence Erlbaum Ass., 2006.

[NAN 05] NANARD M., NANARD J., *SAPHIR – Un cadre de référence pour spécification des hypermédias par intention*, Centre National de la Recherche (CNRS) – LIRMM ; document en ligne, 2005.
(http://hal-lirmm.ccsd.cnrs.fr/docs/00/10/65/05/PDF/D612.PDF)

[PBR 07] PEREZ BASANTA C., RODRIGUEZ MARTIN M.E., *The application of data-driven learning using film transcripts for teaching conversational skills*, in : HIDALGO, E, QUEREDA, L. et SANTANA, J. (eds.), Corpora in the Foreign Language Classroom, Editions Rodopi B.V., p. 141-160, 2007.

[RAS 87] RASTIER F., *Sémantique interprétative*, Paris, PUF, 1987.

[RIC 83] RICOEUR P., *Temps et récit (3 vol.)*, Paris, Editions du Seuil, 1983.

[SAC 11] SAKUNDABHAI P., CHEMOUNY M., *Les usages d'une ressource audiovisuelle*, in : P. Stockinger (éd.), Les archives audiovisuelles, Hermes Science Publishing, p. 151-167, 2011.

[SAR 07] SARFATI G.-E., *Eléments d'analyse du discours*, Paris Armand Colin, 2007.

[SCA 77] SCHANK R., ABELSON R., *Scripts, plans, goals and understanding : an inquiry into human knowledge structures*, Hillsdale, Lawrence Erlbaum Ass. 1977.

[SCH 03] SCHUTZ A., Theorie die Lebenswelt. Die pragmatische Schichtung der Lebenswelt (éd. par M. Endreß et I. Srubar), UVK – Universitätsverlag Konstanz, 2003.

[SDL 11] STOCKINGER P., DE PABLO E., LEMAITRE F., *Contexte et enjeux ; in : P. Stockinger (éd.)*, Les Archives Audiovisuelles, Hermes Science Publishing, Paris, p. 21-54, 2011.

[SDL 12] STOCKINGER P., DE PABLO E., LEMAITRE F., *The production and management of digital archives*, Londres/New York, ISTE/Wiley & Sons, (sous presse), 2012.

[SEA 72] SEARLE J., *Les actes de langage*, Paris, Hermann, 1972.

[SEA 72] SEARLE J., *La construction de la réalité sociale*, Paris, Gallimard, 1995.

[SOW 84] SOWA J., *Conceptual Structures : Information Processing in Mind and Machine*, Boston, Addison-Wesley, 1984.

[STO 83] STOCKINGER P., *Semiotik*, Stuttgart Heinz Verlag, 1983.

[STO 85] STOCKINGER P., « Prolégomènes à une théorie de l'action », *Actes Sémiotiques – Documents VII/ 62*, 1985.

[STO 87] STOCKINGER P., De la structure conceptuelle. Essai d'une description sémiolinguistique des situations statives et dynamiques dans le langage, Thèse d'Etat à l'Université de Paris III – La Sorbonne, 1987.

[STO 88] STOCKINGER P., « Frame representations of political knowledge and reasoning. Folk representations of law and historical representations of nation », *Support en ligne d'une série de conférences à l'Université de Helsinki*, 1988. (http://www.semionet.fr/ressources_enligne/p_stockinger/1988/semio_politique.pdf)

[STO 89] STOCKINGER P., « Schème canonique et standards conventionnels », *Semiotic Inquiry*, 9/1-3, p. 119-135, 1989.

[STO 93] STOCKINGER P., « Conceptual analysis, knowledge management and conceptual graph theory », *Lenguas Modernas*, n° 20, 1993.

[STO 94] STOCKINGER P., « Textsemantische Probleme in der Konzeption und Entwicklung von wissensbasierten Hypertexten. Ein Arbeitspapier », *Institut National des Langues et Civilisations Orientales (INALCO), document en ligne*. (http://www.semionet.fr/ressources_enligne/p_stockinger/1994/HyperDeut.pdf)

[STO 99] STOCKINGER P., *Les nouveaux produits d'information. Conception et sémiotique du document*, Hermes Science Publications, Paris, 1999.

[STO 01] STOCKINGER P., *Traitement et contrôle de l'information*, Hermes Science Publications, Paris, 2001.

[STO 03a] STOCKINGER P., *Le document audiovisuel,* Hermes Science Publishing, Paris, 2003.

[STO 03b] STOCKINGER P., « Global culture, media culture and semiotics », *Support en ligne d'une conférence donnée dans le cadre de l'International Summer School of Semiotics,* Imatra, Finlande, 2003.
(http://www.semionet.fr/ressources_enligne/conferences/2003/globalmedia.pdf)

[STO 03c] STOCKINGER P., (ED.), *Portails et collaboratoires pour la recherche et l'enseignement*, Hermes Science Publishing, Paris, 2003.

[STO 05a] STOCKINGER P., *Les sites web*, Hermes Science Publishing, Paris, 2005.

[STO 05b] STOCKINGER P., Semiotics of culture. Lectures on lifeworld and culture, Cours dans le cadre du « European Master in Intercultural Communication » (EMICC), Anglia Polytechnic University, Cambridge (UK), 2005.
(http://www.semionet.fr/enseignement/annes/05_06/cambridge_06.html)

[STO 07] STOCKINGER P., *Des archives audiovisuelles monolingues sur un site multilingue, in : J. Nowicki, M. Oustinoff (éds.),* Traductions et Mondialisation, CNRS Editions (Hermès. Cognition – Communication – Politique, 49), p. 69-77, 2007.

[STO 10a] STOCKINGER P., *L'adaptation et la republication de ressources audiovisuelles numériques, in : M. Oustinoff, J. Nowicki, J. Machado da Silva (éds.),* Traductions et Mondialisation, vol 2, CNRS Editions (Hermès. Cognition – Communication – Politique, 56), p. 63-71, 2010.

[STO 10b] STOCKINGER P., Semiotics of audiovisuel discourse. A general framework for the analysis of digital audiovisual documents, cours en ligne, Università della Svizzera Italiana (USI) – Executive Master in Intercultural Communication (MIC), 2010.

[STO 11a] STOCKINGER P. (ED.), *Les archives audiovisuelles*, Hermes Science Publishing, Paris, 2011.

[STO 11b] STOCKINGER P. (ED.), *Nouveaux usages des archives audiovisuelles numériques*, Hermes Science Publishing, Paris, 2011.

[STO 11d] STOCKINGER P., Un cadre méthodologique pour analyser un système de références culturelles (cours en ligne), INALCO (Institut National des Langues et Civilisations Orientales), 2011.
(http://www.semionet.fr/ressources_enligne/Enseignement/11_12/mag_1_ICL4a02a/Cours_Notion_Culture.ppt)

[STO 11e] STOCKINGER P., Introduction a la Veille d'Informations (cours en ligne), INALCO (Institut National des Langues et Civilisations Orientales), 2011.
(http://www.semionet.fr/ressources_enligne/Enseignement/11_12/cilm_3_ICL3a02b/cours/Cours_Veille_Termino.pptx)

[TFC 03] TSINARAKI C., FATOUROU E., CHRISTODOULAKIS S., *An Ontology-Driven Framework for the Management of Semantic Metadata describing Audiovisual Information*, MUSIC-TUC (Technical University of Crete), 2003.

http://citeseerx.ist.psu.edu/viewdoc/download ?doi=10.1.1.104.3374&rep=rep1&type=pdf

[TAY 98] TAYLOR C., *Les sources du moi. La formation de l'identité moderne*, Paris, Editions du Seuil, 1998.

[TOR 99] TOROP P., « Cultural semiotics and culture », *Sign Systems Studies*, vol. 27, p. 9-23, 1999.

[WIT 03] WITTGENSTEIN, L., *Philosophische Untersuchungen*, Frankfurt/Main, Bibliothek Suhrkamp, 2003.

[WON 03] MASOLO C., BORGO S., GANGEMI A., GUARINO N., OLTRAMANI A., WonderWeb Deliverable D18 : Ontology Library. ISTC-CNR, 2003.
(online : http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf)

[WRI 63] VON WRIGHT G.H., *The Variety of Goodness*, London, Routledge & Keagan Paul, 1963.

[VAH 10] VAUCELLE A., HUDRISIER H., « Langages structurés et lien social », *in : tic&société*, en ligne, vol. 4, n° 1, 2010.
http://ticetsociete.revues.org/790

Index

A

activité d'analyse, 73, 76, 264, 298
activité de description, 83
activité de l'analyse, 28
analyse, 48, 73
analyse audiovisuelle, 12, 46
analyse conceptuelle, 95
analyse de l'expression audiovisuelle, 76
analyse de la contextualisation référentielle, 76
analyse du contenu, 11, 46
analyse du discours, 71, 78, 129
analyse du discours tenu, 76
analyse du domaine de référence, 76
analyse paratextuelle, 11, 45
analyse pragmatique, 12, 46, 51
analyste, 265

B

Bachimont, B., 224
base canonique du vocabulaire des termes conceptuels, 219
bibliothèque ASA de terminologies propres à un standard, 290
bibliothèque de modèles de description, 51, 52, 56, 181
bibliothèque de schémas d'indexation, 69, 280
bibliothèque de schémas de définition, 192, 307, 308
bibliothèque de séquences de description, 300, 302
bibliothèque des modèles de description, 206
bibliothèque des séquences, 89
bibliothèque des séquences de description, 195

C

cadre axiologique, 117
cadre discursif, 126
champs d'analyse, 62
Chemouny, M., 130, 132
classème, 233, 272
classification, 241
collections de facettes, 291
commentaire métatextuel, 77
concept designer, 265
configuration, 192
configuration thématique, 12

contextualisation historico-temporelle, 111, 117
contextualisation référentielle, 97
contextualisation spatiale, 12, 98
contextualisation temporelle, 12
contextualisation thématique, 12, 97, 117
corpus audiovisuel, 20
corpus d'analyse, 23
corpus de publication, 24
corpus de terrain, 21
corpus de traitement, 22
corpus patrimonial, 25

D

de Pablo, E., 41, 131, 170
description audiovisuelle, 182
description contrôlée, 65, 108, 168
description d'un texte audiovisuel, 151
description du sujet, 144
description libre, 65, 151, 152, 155
description paratextuelle, 75, 182
description pragmatique, 183
description référentielle, 65, 152, 304
description thématique, 182
Deslis, J., 24
dimension sémantique, 233
discours-objet, 208
domaine d'expertise, 44
domaine référentiel, 174
domaine taxinomique, 235
Dublin Core, 49, 75, 182

E

énonciateur, 124
énonciation, 124
Europeana, 72
expression contrôlée, 272

F

facette, 272
facette (du thesaurus), 189
folksonomie, 287
Foucault, M., 10

G

Ginzburg, C., 47
graphe conceptuel, 87
Greimas, A.J., 12, 15, 138, 230, 259, 284
Grunig, B.-N., 125
Grunig, R., 125
Guérinet, R., 10

H

Habermas, J., 210

I

indexation, 30
indexation thématique de base, 78
ingénieur de connaissances, 265

L

Legrand-Galarza, V., 25, 273
Lemaitre, F., 10, 203
Lévi-Strauss, C., 175, 221
linguistic turn, 10
localisation géographique, 98
localisation géographique physique, 104
localisation géopolitique, 99, 102
localisation temporelle, 71, 108
LOMFR, 72, 274, 275

M

médiation linguistique, 277
méta-description, 12, 182
méta-description*, 12
méta-iconographie, 270, 283
métalangage de description, 29, 52, 180, 222, 289, 303
métalangage de description ASA, 27
méta-lexique ASA, 84, 224
méta-lexique des objets d'analyse ASA, 289
méta-lexique des termes conceptuels, 67, 90, 208, 288, 308
méta-lexique des termes conceptuels ASA, 14
micro-thesaurus, 13, 115, 171, 174, 189, 241, 272

mise en discours, 12, 123, 124, 129
mise en discours d'une topique, 123
modèle de description, 19, 27, 39, 51, 53, 67, 87, 88, 183, 265, 296, 298
modèle de description audiovisuelle, 298
modèle de description paratextuelle, 298
modèle de description pragmatique, 298
modèle de description thématique, 56, 67, 298
MOTBIS, 293
mot-clé, 163

O

objet acoustique, 187
objet audiovisuel, 308
objet contextuel, 308
objet d'analyse, 67, 69, 85, 192
objet de connaissance, 139
Objet de culture axiologique, 255
Objet de culture d'art, 257
Objet de culture de croyance, 255
Objet de culture épistémique, 255
Objet de valeur, 220, 231, 234
Objet discours, 248, 251
objet discursif, 187, 308
Objet Endurant, 220, 230
Objet état (statif), 221
Objet langue, 248, 253
Objet naturel, 220, 231
Objet Perdurant, 220, 221, 258
Objet processus, 221
objet référentiel, 11, 86, 187, 308
objet réflexif, 187
Objet Région, 220, 221, 261
Objet région spatiale, 221
Objet région temporelle, 221
Objet texte, 248, 252
objet verbal, 308
objet visuel, 187
ontologie descriptive, 69
ontologie DOLCE, 219
ontologie formelle, 219
ontologie géopolitique, 103
ontologie GOLD, 253
Open Archive Initiative, 49
outil d'analyse, 68, 266

P

paysage textuel, 31
procédure contrôlée, 13
procédure d'analyse, 67, 69, 73, 76, 297
Procédure d'analyse, 266
procédure de description, 13, 65, 192
procédure de description contrôlée, 65, 108, 265, 272, 287
procédure de description libre, 13, 65, 194, 265, 272

R

RAMEAU, 49, 293
Rastier, F., 234
ressources métalinguistiques, 180
ressources métalinguistiques ASA, 272, 288
rhème, 126

S

schéma d'indexation, 278
schéma de définition, 13, 64, 67, 90, 185, 192, 206, 298, 307
schémas d'indexation par facette, 285
schémas de définition, 265
Schütz, A., 90, 209
Searle, J., 230
sémiotique, 32
séquence de description, 13, 61, 67, 90, 195, 206, 265, 297, 299
séquence de description référentielle, 69
structure de l'expression verbale et audiovisuelle, 12
structure de la mise en discours, 12
structure thématique, 84, 123
structure topique, 12, 86, 87, 91, 137, 139, 140, 199
structure topique potentielle, 147
sujet, 265
Système, 234
système métalinguistique ASA, 185

T

tâche d'analyse, 73, 297
taxon, 234

terme conceptuel, 87, 90, 138, 186, 209, 233
terminologie, 272
texte audiovisuel, 32, 267
textscape, 31
thème d'un discours audiovisuel, 265
thème discursif, 126
thesaurus, 14, 71, 167, 171, 188
thesaurus ASA, 90, 207, 287
thesaurus commun, 291
Thesaurus commun ASA, 290
Thesaurus propre à un groupe d'utilisateurs ASA, 290
topique, 86
topos, 216
type d'analyse, 50, 73, 75, 297

U

univers du discours, 10, 52, 57, 303

V

valeur discursive d'une information, 122

W

Web Dewey, 49
Wikipédia, 108

Glossaire des acronymes

Note : les expressions en gras suivies d'un astérisque (exemple « **projet ASA-SHS*** ») constituent une entrée dans le glossaire.

A

AAR© (vidéothèque -)

1/ Acronyme pour *« Archives Audiovisuelles de la Recherche »*.
2/ « Les AAR » sont la vidéothèque générale du Programme AAR. Cette vidéothèque diffuse environ 5500 heures de vidéos en ligne en une quinzaine de langues. Les auteurs et réalisateurs de ce fonds forment une « communauté internationale » de quelques 2500 chercheurs, enseignants, intellectuels, artistes, hommes/femmes politiques, journalistes, professionnels, etc. travaillant dans quelques 85 pays du monde.
3/ Le fonds des AAR documente des grandes parties de la recherche en sciences humaines et sociales. Il englobe également des enregistrements de manifestations culturelles et pédagogiques (expositions, concerts, cours, etc.), des documentations audiovisuelles de recherches de terrain (sur l'émigration européenne, les victimes des dictatures latino-américaines, la fabrication du pain, etc.) ainsi que de reportages et documentaires.
4/ Le fonds audiovisuel est composé d'une part de productions réalisées par l'ESCoM depuis 2001 jusqu'en 2009 (et, dans un rythme bien moindre, depuis 2009 jusqu'à aujourd'hui) ainsi que d'apports audiovisuels appartenant à d'autres institutions françaises et étrangères et/ou à des particuliers actifs dans la recherche.

5/ *(Note P. Stockinger 2018 :)* La vidéothèque AAR dans sa version n'est aujourd'hui plus accessible après son arrêt officiel en 2016. Une partie du fonds collecté se trouve aujourd'hui diffusé sur le portail HAL Campus AAR : https://hal.campus-aar.fr/

AAR© (Programme -)

1/ Le Programme AAR est un programme de recherche et de développement de l'**ESCoM**, mis en place en 2001 dans le cadre du projet **OPALES*** portant sur l'indexation sémantique des ressources audiovisuelles.
2/ Le programme AAR est consacré aux bibliothèques numériques audiovisuelles et à leurs usages concrets notamment dans le cadre de la recherche elle-même, de l'éducation (formelle ou informelle) ainsi que de la valorisation de patrimoines scientifiques et culturels.
3/ Le programme AAR développe, plus particulièrement, les ressources et outils nécessaires pour le traitement (description, indexation, etc.) et la publication enligne de corpus audiovisuels.
4/ Une documentation très détaillée sur le Programme AAR peut être consultée en ligne sur le site du portail HAL Campus AAR : https://hal.campus-aar.fr/

ADA (Programme -)

1/ Acronyme pour *« Atelier des Arkéonautes »*.
2/ Le Programme ADA constitue un des **ateliers d'expérimentation*** du **projet ASA-SHS*** consacré à la constitution d'un corpus audiovisuel documentant la recherche en archéologie, la mise en place d'un **métalangage de description*** pour ce domaine de connaissances, l'analyse du corpus à l'aide du **Studio ASA*** et la publication du corpus analysé – également via le **Studio ASA*** – sous forme d'un portail proposant des accès diversifié au corpus en question : accès par **sujets de connaissance***, accès par **thesaurus***, accès par **collections de documents filmiques***, etc.
3/ Un point particulièrement important traité dans le cadre du programme ADA est l'expérimentation des nouvelles logiques de diffusion et de valorisation de patrimoines culturels et scientifiques via les réseaux sociaux et le web 2.
4/ Le programme ADA prend forme à travers une série de réalisations concrètes : portail web comportant la vidéothèque ADA qui est composé d'un corpus d'environ 135 heures de documents audiovisuels consacrés au domaine de l'archéologie ; chaînes vidéo expérimentales sur You Tube, Daily Motion et Viméo ; réseau Twitter et Facebook ; « agrégations de contenus via des applications telles que Scoop.it, Pearltrees ou Netvibes ; etc.

ALIA (Programme -)

1/ Acronyme pour *« Atelier Littéraire d'Ici et d'Ailleurs »*.
2/ Le Programme ALIA constitue un des **ateliers d'expérimentation*** du **projet ASA-SHS*** consacré à la constitution d'un corpus audiovisuel consacré à la littérature française et du monde, son analyse à l'aide du **Studio ASA*** et la publication du corpus analysé – également via le **Studio ASA*** – sous forme d'un portail proposant des accès diversifié au dit corpus : accès par **sujets de connaissance***, accès par **thesaurus***, accès par **collections de documents filmiques***, etc.
3/ Un point particulièrement important traité dans le cadre du programme ALIA est le travail de l'**analyste*** du corpus. Loin de le réduire à la tâche d'une « simple » indexation de type bibliothécaire ou documentaliste, l'analyste est ici considéré comme un spécialiste du domaine littéraire à qui revient la tâche fort complexe d'expliciter, de « montrer », à travers son travail, la valeur ajoutée (l'intérêt, la pertinence, l'utilité, etc.) d'un texte audiovisuel ou d'une partie d'un texte audiovisuel pour un public donné.
3/ Le programme ALIA se réalise sous forme d'un portail web et d'une vidéothèque enligne comportant environ 150 heures de vidéos en grande partie finement analysées et indexées ainsi qu'une série de publication spécialisées et adaptées à des usages éducatifs particuliers.

ARC (Programme -)

1/ Acronyme pour *« Archives Rencontre des Cultures »*.
2/ Le Programme ARC constitue un des **ateliers d'expérimentation*** du **projet ASA-SHS*** consacré à la constitution d'un corpus audiovisuel documentant la diversité culturelle et le dialogue entre cultures, la mise en place d'un **métalangage de description*** pour ce domaine de connaissances, l'analyse du corpus à l'aide du **Studio ASA*** et la publication du corpus analysé – également via le **Studio ASA*** – sous forme d'un portail proposant des accès diversifié au corpus en question : accès par **sujets de connaissance***, accès par **thesaurus***, accès par **collections de documents filmiques***, etc.
3/ Un point particulièrement important traité dans le cadre du programme ARC est l'exploitation d'archives audiovisuelles dans le cadre de l'enseignement universitaire et de la valorisation d'un patrimoine universitaire (produit, réalisé aussi bien par les chercheurs et enseignants que les étudiants). ARC servira en effet d'environnement de travail et de portail audiovisuel à la *Filière Communication et Formation Interculturelles* (CFI) de l'INALCO (*Institut National des Langues et Civilisations Orientales*) et à ses formations en communication interculturelle au niveau de la licence (L2 et L3) et du master (M1 et M2). Les exploitations se réaliseront en cinq directions principales : i) cours utilisant des ressources audiovisuelles appropriées et disponibles sur le portail ARC ; ii) cours d'analyse de textes audiovisuels ; iii) cours de création et de publication de textes audiovisuels ;

iv) cours utilisant les nouveaux médias sociaux pour collecter, analyser, diffuser et valoriser de contenus audiovisuels ; v) projets de création de patrimoines audiovisuels à thème relevant des langues et civilisations enseignées à l'INALCO (exemple : patrimoines des littératures africaines ; patrimoines andins en relation avec le **programme PCIA***, etc.)
4/ Le programme ARC prend forme à travers une série de réalisations concrètes : portail web comportant la vidéothèque ARC qui est composé d'un corpus d'environ 140 heures de documents audiovisuels ; réseaux Twitter et Facebook ; « agrégation de contenus netvibes, scoop.it, etc. » ; etc.

ASA-SHS (projet -)

1/ ASA-SHS est l'acronyme pour *« Atelier de Sémiotique Audiovisuelle pour la description de corpus audiovisuels en sciences humaines et sociales »*.
2/ ASA-SHS est un projet de recherche de l'ESCoM/FMSH financé par l'ANR (Agence Nationale de la Recherche) en France. Son numéro de référence est : ANR-08-BLAN-0102-01. ASA-SHS a débuté en janvier 2009 et se termine officiellement fin décembre 2011. Les résultats de ce projet pourront, cependant, être réutilisés dans de nouveaux projets de recherche
3/ L'objectif principal du projet ASA-SHS est le développement de **ressources métalinguistiques*** (c'est-à-dire un **métalangage***) pour la description de corpus audiovisuels afin d'en faire des ressources à proprement parler, c'est-à-dire des « outils », des « instruments » de connaissances dans des contextes et pour des usages spécifiques (notamment dans la recherche elle-même, l'éducation, la valorisation de patrimoines culturels et scientifiques, etc.).
4/ L'élaboration des ressources métalinguistiques pour la description de corpus audiovisuels s'appuie sur la sémiotique du document audiovisuel [STO 03] ainsi que sur l'analyse du discours. Les **ressources métalinguistiques ASA*** sont composées notamment 1/ de *méta-lexiques de termes conceptuels* (ou concepts) nécessaires à la description d'un document audiovisuel, 2/ de *modèles de description* spécialisés à l'analyse de tel ou tel aspect structural d'un document audiovisuel (tel que le plan thématique, le plan visuel ou sonore, le plan discursif, etc.), 3/ d'un important *thesaurus* de termes ou de « descripteurs » prédéfinis.
5/ Les ressources métalinguistiques ASA sont utilisées dans un environnement de travail sophistiqué appelé **Studio ASA*** composé d'un atelier de segmentation de corpus audiovisuels, d'un atelier de description de corpus audiovisuels, d'un atelier de publication des corpus audiovisuels décrits et d'un atelier de la gestion des ressources métalinguistiques elles-mêmes.
6/ Les ressources métalinguistiques ASA ainsi que le Studio ASA sont développées, testés et validés grâce à plusieurs **ateliers d'expérimentation***, consistant en des travaux intensifs de description, de publication et de valorisation de corpus audiovisuels.

ASA© (Studio -)

1/ Le Studio ASA est un environnement de travail permettant la segmentation, l'analyse (description, annotation, indexation adaptation linguistique, etc.), la publication enligne de corpus audiovisuels ainsi que la gestion (définition, adaptation, etc.) des **ressources métalinguistiques ASA***.

2/ Le studio ASA comprend **quatre** grandes parties appelées « **Ateliers** » : i) l'**atelier de segmentation*** de textes audiovisuels (de vidéos, pour le moment) ; ii) l'**atelier de description*** d'un texte audiovisuel et/ou de certaines de ses parties (segments) ; iii) l'**atelier de publication*** d'un texte audiovisuel ou d'un corpus de textes audiovisuels ; iv) l'**atelier de modélisation*** des ressources métalinguistiques (concepts, configurations de concepts, thesaurus, etc.).

3/ Chaque Atelier se présente sous forme i) d'un ***environnement de travail**** *spécialisé* et ii) d'une de *documentation technique, d'aides, d'exemples concrets*, etc. Ainsi, le Studio ASA, comprend-il un environnement de travail pour la segmentation d'une vidéo, un environnement de travail de description/d'analyse à proprement parler d'une vidéo, un environnement de travail de publication d'une vidéo ou d'un corpus de vidéo et, enfin, un environnement de travail de scénarisation, c'est-à-dire de définition et de développement d'un métalangage nécessaire pour analyser et publier des corpus audiovisuels.

4/ Les environnements de travail existants sont développés d'une manière inégale : l'environnement le plus élaboré est celui de l'analyse de corpus audiovisuels. Cependant, tous les environnements sont opérationnels. De même, la documentation technique devant accompagner chaque environnement n'est souvent réalisée que d'une manière rudimentaire.

5/ Chaque environnement de travail, à son tour, est pourvu d'un ou de plusieurs logiciels, applications informatiques, etc. Ainsi, **Interview*** de l'INA est actuellement le logiciel utilisé dans l'**Atelier de Segmentation*** ; **Semiosphere*** est un ensemble d'applications développées par l'ESCoM et sert de « brique technologique » à l'**Atelier de Publication*** ; l'environnement de travail de l'**Atelier de Description*** est composé d'un ensemble d'applications de gestion de bibliothèque de formulaires (« modèles ») de saisie d'informations et de gestion des données générées par l'analyste ; enfin, l'**Atelier de Modélisation*** utilise actuellement un éditeur xml qui s'appelle **OntoEditeur*** (également développé par l'ESCoM) pour la définition des **ressources métalinguistiques***.

6/ Le Studio ASA a été développé par l'équipe ESCoM dans le cadre du projet **ASA-SHS*** (2009 – 2011) financé par l'ANR dans le cadre du « Programme Blanc 2008 ».

Atelier de Segmentation©

1/ L'Atelier de Segmentation ASA est un environnement de travail spécialisé pour le découpage d'un document audiovisuel en des segments (« passages » ou « séquences ») qui intéressent l'**analyste***. Il fait partie du **Studio ASA*.**

2/ L'environnement de travail de l'Atelier de Segmentation ASA est aujourd'hui composé d'un outil logiciel permettant d'effectuer le découpage d'un document audiovisuel. Il s'agit de l'outil Interview développé à l'origine par INA-Recherche et adapté ensuite aux besoins techniques de découpage de documents audiovisuels dans le cadre du **Programme AAR***. L'outil *Interview* pourra être remplacé, dans les mois à venir, par des outils plus performants permettant un découpage « multicouche » ainsi que la segmentation d'images statiques. L'environnement de travail de l'Atelier de Segmentation est également doté d'un ensemble de documents techniques et explicatifs à l'adresse de personnes ou institutions souhaitant l'utiliser.

Atelier de Description ©

1/ L'Atelier de Description d'un texte audiovisuel (ou, plus court, *Atelier de description ASA*) est l'environnement de travail permettant à l'analyste de procéder à une analyse d'un texte audiovisuel dans son intégralité ou seulement d'une de ses parties. Il fait partie du **Studio ASA*.**

2/ L'Atelier de Description est composé de trois parties fonctionnelles majeures : i) une première partie réservée à la ***méta-description**** elle-même (c'est-à-dire à la présentation de l'analyse et de ses objectifs) ; ii) une partie réservée à la *description de l'objet audiovisuel dans son ensemble* ; iii) une partie réservée à la *description des **segments**** *spécifiques du texte audiovisuel* préalablement identifiés par l'analyste.

3/ Les parties (ii) et (iii) de l'analyse du texte audiovisuel dans son ensemble et de tel ou tel segment spécifique d'un texte audiovisuel se décomposent de nouveau en une série de parties fonctionnellement plus spécialisées permettant à l'analyste de procéder à une description systématique de son objet.

4/ Le travail d'analyse d'un document audiovisuel se réalise sous forme d'une série de **formulaires interactifs***. Chaque formulaire interactif représente un **modèle de description***. Un modèle de description, lui, est formé à partir des **ressources métalinguistiques ASA***. Ainsi, distingue-t-on une bibliothèque de formulaires interactifs réservés à la méta-description (c'est-à-dire à l'explicitation d'une analyse, de son contenu, de ses objectifs, etc.), une bibliothèque de formulaires interactifs réservés à l'analyse d'une vidéo, d'un document audiovisuel dans son intégralité et une bibliothèque de formulaires interactifs réservés à l'analyse d'un segment particulier faisant partie d'un document audiovisuel.

5/ Les formulaires interactifs peuvent être adaptés aux besoins spécifiques de tel ou tel analyste, ou encore de tel ou tel groupe d'analystes étant donné que chaque formulaire représente un **modèle de description*** élaboré à partir les **ressources métalinguistiques ASA***. Mais l'adaptation d'un formulaire interactif exige un travail parfois important au niveau de l'adaptation, voir même de la ré-définition des **modèles de description*** concernés ou encore des ressources **métalinguistiques ASA*** nécessaires pour élaborer un **modèle de description***. Ce travail d'adaptation ou de redéfinition de modèles de description se fait via l'**Atelier de Scénarisation ASA***.

Atelier de Publication ©

1/ L'Atelier de Publication ASA fait partie du Studio ASA. Il est réservé, comme son intitulé l'indique déjà, à la publication de corpus audiovisuels traités et décrits préalablement via l'Atelier de Description ASA.
2/ L'Atelier de Publication permet la publication de corpus audiovisuels sous forme d'un site portail (de type **ADA***, **ALIA*** ou **ARC***) mais aussi sous forme de dossiers spécialisés (**dossiers thématiques***, **dossiers bilingues***, **dossiers pédagogiques***, etc.) ou encore de **vidéo-glossaires***, de **parcours narratifs***, etc.
3/ Dans le cadre d'un programme de *« mise en ligne » de corpus audiovisuels* – mise en ligne qui suit une politique de publication préétablie – le processus de publication se fait d'une manière automatique. C'est le cas, par exemple, du site **portail AAR*** mais aussi des sites portail ayant servi de champs d'expérimentation au **projet ASA-SHS*** : **ARC***, **ADA***, **ALIA***, **PCIQ***, **PCIA***.
4/ Cependant, l'Atelier de Publication ASA ouvre la voie à la publication personnalisée, voire à la **republication*** de corpus audiovisuels déjà publiés sur la base de **modèles de publication*** représenté de nouveau par des **formulaires interactifs***. Il s'agit ici de la possibilité offerte au **rédacteur-auteur*** de pouvoir réutiliser soit des corpus audiovisuels déjà publiés (par exemple sous forme d'un **dossier thématique*** ou **bilingue***) soit des corpus décrits et indexés (via l'**Atelier de Description***) afin de « créer » (publier ou republier) des ressources audiovisuelles numériques selon ses propres besoins ou intérêts et/ou en tenant compte de son public et de ses attentes. Cette option a été expérimentée dans les projets **LOGOS*** et **SAPHIR*** (pour voir des résultats concrets, on peut consulter les sites **DLC*** et **PCM***). Par contre, elle ne constitue pas un objet de recherche pour le projet **ASA-SHS***. L'Atelier de Publication ASA, dans sa version actuelle, ne présente ainsi que des fonctions très élémentaires d'une publication/ republication personnalisée de corpus audiovisuels déjà décrits et indexés, voire publiés.
5/ L'Atelier de Publication ASA se présente, aujourd'hui, sous forme d'un **environnement de travail*** composé d'un ensemble de services web développés par ESCoM et réunis sous l'appellation **Semiosphère***.

Atelier de Modélisation©

1/ L'Atelier de Modélisation ASA est l'un des quatre ateliers qui composent le Studio ASA.
2/ Il sert au cogniticien ou « concept designer » de définir, de développer, de tester et de valider les modèles de description dont on se sert dans l'**Atelier de Description*** sous forme de formulaires interactifs pour analyser des corpus audiovisuels.
3/ L'Atelier de Modélisation se présente sous forme d'un logiciel développé par l'**ESCoM***. Ce logiciel du nom de **OntoEditeur*** permet au cogniticien : i) de développer et/ou de gérer les ressources métalinguistiques nécessaires pour la

réalisation de modèles de description ; ii) de réaliser les modèles de description eux-mêmes en utilisant les ressources métalinguistiques disponibles.

C

CONVERGENCE (projet -)

1/ Convergence est un projet de R&D européen qui a débuté en juin 2010 et qui durera jusqu'en février 2013. Il est coordonné par le CNIT (Consorzio Interuniversitario per le Telecomunicazioni) à Rome et financé dans le cadre du 7ème PCRD ; n° : FP7-257123). Le but du projet CONVERGENCE est d'enrichir Internet avec un nouveau modèle de service de *publication-souscription*, centré sur le *contenu* et basé sur un container commun pour tout type de données numériques, incluant les personnes et les objets du monde réel (RWOs). Ce container commun, appelé *Versatile Digital Item* (VDI, *Element Numérique Versatile* en français), est un paquet structuré de contenu numérique et de méta-informations, identifié de manière unique (comme peut l'être l'url d'un site web) et qui étend le concept de *Digital Item* défini par le MPEG-21.
2/ L'intérêt de l'ESCoM et du programme AAR dans ce projet concerne le fait que tous les usages d'une vidéo enligne peuvent être tracés via la technologie VDI. Cela ouvre la voie vers une circulation de contenus numériques et une appropriation de ceux-ci qui respectent les droits de leurs auteurs et propriétaires. Le champ d'expérimentation choisi est celui de la production, de la diffusion et du partage de contenus culturellement sensibles faisant partie du patrimoine culturel immatériel des communautés *quechuaphones andines* (voir le **programme PCIA***).
3/ L'url du site officiel du projet Convergence : http://www.ict-convergence.eu/

E

ESCoM©

1/ ESCoM (Equipe Sémiotique Cognitive et Nouveaux Médias) est un programme de recherche de la Fondation Maison des Sciences de l'Homme (FMSH) à Paris. Créée en 1991, l'ESCoM est spécialisée dans les domaines de la sémiotique du texte et du discours appliquée aux problématiques de la production, de l'analyse et de la publication-diffusion de patrimoines numériques de connaissances. Un deuxième axe de recherche de l'ESCoM est celui de l'analyse culturelle et de la communication interculturelle (axe de recherche développé en concertation avec la Filière Communication et Formation Interculturelles (CFI) de l'Institut National des Langues et Civilisations Orientales (INALCO) à Paris.
2/ En 2001, dans le cadre du projet de R&D français **OPALES***, l'ESCoM a créé le **programme AAR*** qui constitue, son chantier d'expérimentation le plus important.

3/ Depuis 1994, l'ESCoM est régulièrement présente dans des projets de R&D européens et français. Le **projet ASA-SHS*** en est un des plus importants étant donné qu'il a permis à l'équipe d'achever en grande partie plus que 10 années de recherche et de développement consacrés à la description, indexation et publication de corpus audiovisuels en contexte numérique.

ESCoM Suite 2011©

1/ L'ESCoM Suite 2011 est déjà la troisième édition d'un ensemble d'outils et de logiciels développés en grande partie par l'ESCoM et servant à la production, publication et gestion de données et d'archives audiovisuelles.
2/ L'ESCoM Suite 2011 est composée des outils suivants : *OntoEditeur* (un éditeur xml pour la création et gestion de ressources métalinguistiques nécessaires à la description de corpus audiovisuels) ; *Interview* (logiciel de segmentation et d'analyse de vidéo, co-développé avec INA Recherche) ; *GAAV* (« Gestion d'Archives AudioVisuelles », outil de publications et de gestion de publications audiovisuelles), *ESCoM Playlist Maker* (outil permettant de créer de playlists à partir d'une description/indexation réalisée avec Interview et en vue de la publication d'une telle description/indexation), *ESCoM ffCoder* (outil d'encodage de fichiers vidéo en vue de leurs publication dans un format de « streaming »).
3/ Les outils composant l'ESCoM Suite sont intégrés dans les différents ateliers composant le **Studio ASA*** (**Atelier de Segmentation***, **Atelier de Description***, **Atelier de Modélisation***, **Atelier de Publication***, etc.). L'ESCoM Suite fera ainsi place au Studio ASA.

O

OntoEditeur

1/ OntoEditeur est un éditeur xml, développé par l'**ESCoM*** pour la réalisation et la gestion des **ressources métalinguistiques*** (hiérarchie de termes conceptuels, schémas et modèles de description, etc.) nécessaires à l'analyse de corpus audiovisuels.

P

PCIA (portail -)

1/ PCIA est l'acronyme pour « Patrimoine culturel immatériel quechua » qui est un portail servant de champ d'expérimentation au **projet ASA-SHS***.

S

SAPHIR (projet -)

1/ SAPHIR (« Système d'Assistance à la Publication Hypermédia ») est un projet de recherche français. Financé par l'ANR et coordonné par INA Recherche, il a débuté en 2006 et s'est terminé fin 2010.
2/ En étroite concertation avec INA Recherche, l'**ESCoM*** a concentré ses efforts sur l'analyse sémiotique de corpus audiovisuels, la définition d'un **métalangage de description*** adapté à traiter de corpus provenant du **programme AAR*** ainsi que d'un *métalangage de publication* en privilégiant le genre du vidéo-lexique. L'ESCoM a également développé, dans le cadre de ce projet, une version améliorée d'**Interview***, outil de segmentation et de description « de base » de vidéos, tout en l'intégrant dans le **processus de travail*** propre au **programme AAR***.
3/ Les résultats du projet SAPHIR ont constitués l'input majeur pour les activités de recherche et de développement du projet **ASA-SHS***.

Studio ASA©

1/ Le Studio ASA est un **environnement travail*** numérique développé principalement dans le cadre du et grâce au projet **ASA-SHS*** financé par l'ANR.
2/ Le Studio ASA est composé de plusieurs Ateliers : i) **Atelier de Segmentation*** d'une donnée audiovisuelle ; ii) **Atelier de Description*** d'une donnée audiovisuelle ; iii) **Atelier de Publication*** d'une donnée audiovisuelle ; iv) **Atelier de Modélisation*** destiné à la mise en place et gestion des ressources métalinguistiques et de modèles nécessaires pour décrire et/ou publier de données audiovisuelles. La liste des ateliers n'est pas close – d'autres ateliers pourront s'y ajouter, notamment un atelier consacré plus particulièrement au traitement technique et auctorial (c'est-à-dire au montage) d'une donnée audiovisuelle.
3/ Le Studio ASA inclut actuellement les outils et logiciels de l'**ESCoM Suite***. Mais des outils et logiciels pouvant être développés ailleurs pourront y être intégrer, au fur et à mesure.

Glossaire des termes spécialisés

Note : les expressions en gras suivies d'un astérisque (exemple « **projet ASA-SHS*** ») constituent une entrée dans le glossaire.

A

Activité d'analyse

1/ Type d'actions concrètes qui sert à désigner, dénommer, indexer, annoter, etc., un **objet d'analyse***.
2/ Une activité d'analyse compose, seule ou ensemble avec d'autres activités d'analyse, une procédure de description (dont, plus spécifiquement, la **procédure de description libre*** et la **procédure description contrôlée***).
3/ Toute activité d'analyse est identifiée par son terme conceptuel correspondant dans le **méta-lexique*** dont le terme de base est [Procédure d'analyse].

Analyse (d'un corpus audiovisuel)

1/ Terme qui désigne l'une des deux classes principales de tâches et activités du travail sur un corpus audiovisuel, l'autre classe étant celle du **traitement*** d'un corpus audiovisuel.
2/ L'analyse est un travail d'explicitation de la ***valeur*** *potentielle* d'un texte audiovisuel pour un public donné (l'**analyste** pouvant être ce public). Une **valeur** potentielle peut être un nouvel apport en informations à propos d'un sujet particulier, des images manquantes ou inédites, des explications permettant de mieux comprendre un fait, etc.).

3/ Nous distinguons entre quatre niveaux d'analyse :
i) le niveau le plus général est celui du **type d'analyse/de description** (parmi les types d'analyse le plus récurrents nous comptons notamment : la **méta-description***, la **description paratextuelle***, la **description audiovisuelle*** stricto sensu, la **description thématique***, la **description pragmatique*** et aussi une série d'analyses plus spécialisées);
ii) vient ensuite le niveau de la **tâche d'analyse*** (c'est-à-dire tout type d'analyse comporte toujours une ou plusieurs tâches de description). Nous avons travaillé plus particulièrement sur les tâches de description composant le type d'analyse **description thématique (du contenu)*** d'un texte ou d'un corpus de textes audiovisuels. Parmi les tâches les plus récurrentes, nous comptons, par exemple, la tâche de la **description référentielle***, la tâche de la **description de la mise en discours***, la tâche de la **description** de l'**expression** (ou de la **mise en scène**) **audiovisuelle*** et la tâche du commentaire (à distinguer du type d'analyse « méta-description ») ;
iii) chaque tâche d'analyse est composée d'une ou de plusieurs **procédures d'analyse***. Nous distinguons notamment deux procédures : la **procédure de description libre*** (c'est l'analyste qui saisit les informations nécessaires pour expliciter, décrire, commenter, etc., son objet d'analyse) ; la **procédure de description contrôlée*** (l'analyste se sert d'un ou de plusieurs **thesaurus***/**micro-thesaurus*** pour expliciter son objet d'analyse) ;
iv) enfin, une procédure de description est composée d'une ou de plusieurs **activités d'analyse*** (dont par exemple, l'activité de la désignation d'un objet de connaissance, de sa dénomination, de sa description textuelle, de sa localisation géographique et/ou chronologique, et ainsi de suite). Chaque activité de description est, à son tour, définie par un ou plusieurs **schémas d'indexation***.
4/ Une analyse (concrète) d'un texte ou d'un corpus de textes audiovisuels s'inscrit forcément en une **politique *lato sensu* d'analyse**.

Analyste (Rôle d'-)

1/ Ce terme réfère au travail intellectuel consacré à un corpus audiovisuel (ou, plus généralement, textuel) : a) *identification* d'un corpus de travail (ou **corpus d'analyse*)** ; b) *localisation* des passages ou « moments » pertinents (à un objectif posé) et *segmentation* (c'est-à-dire « extraction » des passages et moments pertinents), c) *description contrôlée* ou *libre* des passages et *indexation* à proprement parler, d) et – enfin si pertinent - *traduction-adaptation* des passages préalablement décrits.
2/ Le rôle de l'analyste est un parmi d'autres rôles qui caractérisent le travail dans le cadre d'archives (ou de bibliothèques) numériques. D'autres rôles importants sont, par exemple : le rôle de l'éditeur et le **rôle du cogniticien*** (ou du « concept designer »).
3/ Le rôle de l'analyste se manifeste, concrètement, sous des formes assez diverses : simple activité « quotidienne » de classification d'archives personnelles ; travail du

bibliothécaire ou du documentaliste classifiant et indexant un fonds selon des critères prédéfinis ; ou encore, travail d'un spécialiste du domaine (d'un « expert ») documenté par un fonds audiovisuel.
4/ Le rôle de l'analyste présuppose une série de compétences (c'est-à-dire de connaissances et de savoir-faire) parmi lesquelles nous retenons notamment les trois suivantes : a) *compétences en analyse textuelle* (c'est l'aptitude de produire une expertise dite sémiotique) ; b) *compétence du domaine d'analyse* (c'est l'aptitude de produire une expertise dite référentielle) ; ainsi que c) connaissances et savoir-faire en matière des technologies et outils nécessaires pour réaliser et promouvoir une analyse (c'est l'aptitude de conduire une veille technologique permanente, d'identifier des techniques et outils appropriés au travail d'analyse ainsi que d'utiliser effectivement ces techniques et outils pour réaliser et diffuser le travail d'analyse).
4/ Etant donné l'importance de plus en plus évidente de l'identification, la description, la publication, la diffusion et l'appropriation d'informations pertinentes dans une société largement déterminée par une économie des connaissances, le rôle de l'analyste évolue aujourd'hui très rapidement en des nouveaux métiers et professions.

Archive(s)

1/ Ce terme est utilisé ici dans un sens très large conformément à son usage au sens d'archives numériques.
2/ Une archive est composée d'un fonds textuel qui forme un *espace (évolutif) de connaissances* (potentielles) et donc aussi de valeurs, de croyances, de savoir-faire, d'idéologies de « pensées de groupe » etc. Autrement dit, une archive peut être considérée comme une forme de *matérialisation textuelle d'une culture*, d'un cadre de référence (au sens de Charles Taylor [TAY 98]) pour une personne, un groupe social, une institution (cf. [STO 11d]. Ainsi une archive remplit différentes fonctions essentielles dans la constitution, maintien et l'évolution d'une identité individuelle ou collective et de sa transmission. Par ailleurs, une archive constitue une des ressources les plus importantes pour l'analyse ou l'expertise culturelle (géopolitique, linguistique, sociale, économique, etc.).
3/ En adoptant une approche textuelle et discursive (popularisée par l'expression *linguistic turn* [CRA 08] et à l'origine duquel on trouve des savants tels que Michel Foucault, Michel de Certeau ou encore Jacques Derrida), on peut dire que pour sa ou ses (différentes) « communautés d'usagers » une archive déploie un **univers du discours*** (plus ou moins facilement identifiable, plus ou moins homogène ou hétéroclite, plus ou moins général, …) dans lequel se trouve potentiellement thématisé(s) et traité(s) le ou les cadres de référence (voir supra) pour la ou les communautés en question. Autrement, vouloir rendre explicite l'univers du discours d'une archive, veut dire adopter un certain point de vue (c'est-à-dire celui de l'analyste œuvrant au nom de « sa » communauté d'usagers), l'expliciter à l'aide des **modèles de description*** qui représentent l'intérêt et les attentes de l'analyste et l'implémenter sous forme de formulaires interactifs de description qui font partie

d'un outil informatique d'analyse ou de description de ressources audiovisuelles d'une archive. C'est, en quelque mots, l'approche telle qu'elle a été définie et expérimentée dans le cadre du **projet ASA-SHS***.
4/ Techniquement parlant, une archive se compose typiquement des éléments suivants : i) fonds textuel (écrit, visuel, acoustique, audiovisuel, …) ; ii) bibliothèque numérique (= structure du fonds textuel et sa gestion) ; iii) portail web (espace de publication) pourvu d'interfaces d'accès au fonds et de prestations diverses dont, notamment, les publications et services d'exploitation.

B

Bibliothèque de modèles de description

1/ L'**univers du discours*** d'une archive audiovisuelle, c'est-à-dire ses sujets, thèmes, discours, stratégies d'expression audiovisuelle etc., peut être explicité par le recours à une bibliothèque de **modèles de description***.
2/ Une bibliothèque de modèles de description est composée, en règle générale, d'un ensemble de collections spécialisées de tels modèles. Par exemple, la bibliothèque des modèles de description des archives ARC (Archives *A la Rencontre des Cultures*) est, actuellement, composée de collections de modèles permettant à l'analyste de décrire des textes audiovisuels dans lesquels est question de civilisations, de formations culturelles, du patrimoine culturel matériel et immatériel, de la dynamique interculturelle, ou encore, d'une manière plus précise, des langues du monde, des musiques du monde, des cuisines du monde, etc.
3/ Une bibliothèque de modèles de description est évolutive. Selon les objectifs et les intérêts des parties prenantes d'une archive, de nouvelles collections peuvent se former, des collections déjà existantes peuvent être adaptées ou figées (c'est-à-dire laisser en l'état dans lequel elles se trouvent à un certain moment donné). Tout dépend ici de la politique (scientifique, culturelle, etc.) adoptée pour faire vivre une archive et pour permettre son exploitation par des « usagers » les plus divers.

Brique(s) de modèle(s)

1/ Ce terme désigne l'ensemble de configurations conceptuelles (de termes conceptuels) génériques et/ou (partiellement) référencées qui composent les **modèles de description***.
2/ Les principales classes de briques de modèles sont : i) les **schémas de définition de l'objet d'analyse*** ; ii) les **schémas de définition des procédures d'analyse*** ; iii) les **schémas d'indexation*** ***stricto sensu*** définissant une **activité d'analyse***; les schémas référentiels (pas traités dans ce livre); les **séquences de description***.

C

Concept (voir : terme conceptuel)

Configuration audiovisuelle

1/ La configuration audiovisuelle définit la forme d'expression ou de mise en scène d'un sujet thématisé et traité dans un texte audiovisuel. D'abord, elle recouvre les plans visuels et acoustiques ainsi que la linéarisation et la synchronisation des deux plans. Ensuite, elle réunit les différentes techniques de mise en scène (audiovisuelle) : techniques visuelles (telles que le cadrage, le point de vue, les mouvements de la caméra, les choix des plans visuels, l'éclairage, etc.) et techniques acoustiques (telles que le choix d'une source sonore, le cadrage sonore, etc.). L'usage systématique de certaines techniques pour exprimer un certain type de sujets contribue à ce qu'on appelle parfois un langage visuel (audiovisuel) propre à tel ou tel type de communication audiovisuelle.
2/ Techniquement parlant, la configuration audiovisuelle définit une petite bibliothèque de **séquences*** et de **modèles de description*** dont se sert l'analyste soit pour procéder à la description audiovisuelle stricto sensu (c'est-à-dire à la description des plans visuels, acoustiques ou audiovisuels) soit pour procéder à la description de la mise en scène ou encore de l'expression audiovisuelle d'un sujet, d'un thème de discours.
3/ Contrairement aux séquences de description **topique*** et de **modèles de description*** thématique, les séquences et modèles de description audiovisuelle sont relativement indépendant d'un domaine (objet) de connaissances à un autre, d'un corpus de textes audiovisuels à un autre ou encore d'une archive à une autre.

Configuration discursive (de mise en discours)

1/ La configuration discursive définit les stratégies de mise en discours d'un objet thématisé pour en faire un thème de discours (de l'auteur du discours) à proprement parler. Ainsi s'intéresse-t-on, en prenant en considération le discours sur un objet thématisé, le point de vue de l'auteur (de l'énonciateur), ce qu'il présuppose comme *connu* ou, au contraire, comme *nouveau* dans un contexte de communication donné, ses *références* pour étayer son point de vue sur le sujet, le niveau de la *spécialisation* dans le traitement discursif de l'objet thématisé, etc.
2/ Nous distinguons entre deux types d'analyse : d'une part l'analyse de la *mise en discours d'un sujet*, d'une **topique*** (type d'analyse extensivement traitée dans ce livre ; voir notamment le chapitre 7) et d'autre part la *description discursive du texte audiovisuel* (pas traitée dans ce livre). Dans le premier cas, comme nous le savons

déjà, on s'intéresse à la façon comment un auteur traite un sujet spécifique, comme il traite, par exemple, le sujet « culture technique de la civilisation *chavin* » : *il s'intéresse à quoi ; qu'est-ce qu'il présuppose comme connu chez le destinataire de son discours ?; les informations produites par lui, viennent-elles toutes de lui ou réfère-t-il à d'autres sources ?;* etc. En revanche, l'analyse discursive du texte audiovisuel s'intéresse à l'identification et la description du type et/ou du genre de discours qui caractérise un texte. Typiquement, il s'agit ici de questions telles que : *tel texte audiovisuel (en entier ou tel ou tel passage spécifique) est-ce une narration, une description, une explication ? tel texte audiovisuel texte audiovisuel est-ce un discours didactique, un discours publicitaire, un discours scientifique ?* etc.
3/ Techniquement parlant, la *description de la mise en discours d'un sujet, d'une topique* (voir chapitre 7) se fait à l'aide d'un petit ensemble (d'une petite collection) de séquences spécialisées de description. D'une manière similaire à la collection des séquences spécialisées dans la description audiovisuelle, aussi cette collection de séquences spécialisées, est relativement indépendante d'un domaine/objet de connaissance à un autre, d'un corpus audiovisuel à un autre, voire d'une archive à une autre. Autrement dit, elle peut être utilisée quasiment telle qu'elle pour l'analyse de la mise en discours de topique traitant des domaines de connaissance les plus variés.
4/ Enfin, en ce qui concerne l'analyse discursive du texte audiovisuel (deuxième type d'analyse introduit dans le paragraphe 2 ci-dessus), elle n'est pas beaucoup développée, pour l'instant et se présente sous forme d'un modèle de description très simple et générale mis à la disposition à l'analyste intéressé dans l'**Atelier de Description du Studio*** ASA.

Configuration thématique

1/ La configuration thématique définit la structure du contenu audiovisuel et sert à la mise en place des modèles de description thématique (ou du contenu) accessibles à l'analyste *via* une série de formulaires interactifs de travail dans l'**Atelier de Description du Studio*** ASA.
2/ La configuration thématique se compose de plusieurs types fonctionnellement spécialisés de configurations parmi lesquelles nous comptons notamment la **configuration topique***, la **configuration discursive***, la **configuration de la mise en scène audiovisuelle*** (et verbale – type de configurations non traité ici).

Configuration topique

1/ Techniquement parlant, la configuration topique explicite et définit une vision (voire une théorie pas obligatoirement scientifique) d'un domaine de connaissances qui fait partie – dans le cadre de notre recherche - de l'**univers du discours*** d'une **archive audiovisuelle***.
2/ Une configuration topique est composée de **termes conceptuels*** (génériques ou référencés) qui se positionnent les uns par rapport aux autres selon des relations

spécifiques (relations telles que celles de la spécialisation taxinomique, de l'appartenance d'un objet d'analyse à un objet formant un tout, de la caractérisation d'un objet d'analyse par une propriété ou une qualité, de la localisation temporelle ou spatiale, etc.). Visuellement parlant, une configuration topique s'exprime comme un réseau ou un graphe dont les nœuds représentent les **termes conceptuels*** et les arrêtes les relations (orientées).
3/ La configuration topique forme une **séquence de description*** composée d'au moins de deux **schémas de définition*** – un schéma définissant l'**objet d'analyse***, et un schéma définissant la **procédure d'analyse***. Mais une configuration topique peut être également composée par une *sélection* de schémas d'objets d'analyse. C'est, par exemple, le cas très répandu de la configuration topique qui d'abord identifie l'objet référentiel de connaissance (par exemple sous forme de la sélection et du positionnement des deux termes conceptuels [Civilisation] et [Culture] pour former un schéma de type [Formation culturelle d'une Civilisation]) et qui, ensuite, le localise spatialement et temporellement (par exemple sous forme d'une sélection du terme conceptuel [Région géographique] et [Epoque] pour former un deuxième schéma d'objet d'analyse qui se concatène avec le premier pour devenir un schéma conceptuel plus complexe – et moins général – qui servira à l'analyste de décrire les **sujets*** relatifs aux formations culturelles d'une civilisation à telle ou telle époque et dans telle ou telle région géographique.
4/ Les configurations topiques forment l'input principal du premier type fonctionnel des **séquences d'analyse*** dans un **modèle de description***, c'est-à-dire des séquences qui sont réservées à la description des **sujets*** thématisés dans un **texte audiovisuel*** ou dans une des parties qu'intéressent l'analyste.

Corpus audiovisuel

1/ Ensemble de textes visuels, sonores, filmiques, etc. de type/genre, de taille, de finition éditoriale etc., différents qui possède une fonction particulière selon l'étape du **processus de travail*** (de production-publication audiovisuelle) dans laquelle il est constitué et utilisé.
2/ Nous distinguons entre le **corpus de terrain***, le **corpus de traitement*** et le **corpus traité***, le **corpus d'analyse*** et le **corpus analysé***, le **corpus de publication*** et le **corpus publié***.
3/ La constitution d'un corpus audiovisuel doit suivre des procédures aussi explicites que possibles qui sont définies sous forme de guides composant la documentation technique qui accompagne le **processus de travail*** (de production-publication audiovisuelle). En effet, selon sa place et sa fonction dans le processus de travail de production-publication audiovisuelle, un corpus audiovisuel doit répondre à un ensemble d'attentes et de besoins et peut donc être évalué par rapport à des critères tels que couverture empirique, cohérence/pondération interne, qualité épistémique, etc.

Corpus d'analyse

1/ Le corpus d'analyse réunit l'ensemble de données audiovisuelles que l'analyste a sélectionné pour les soumettre à une analyse, c'est-à-dire à un ensemble d'activités visant à expliciter la **valeur*** potentielle du corpus (ou de telle ou telle partie du corpus) pour un public donné. Ce travail constitue une des étapes principales du **processus de travail*** dans le cadre du **programme AAR*** et s'effectue *via* l'**Atelier de Description ASA***.

2/ Le corpus d'analyse peut être composé de données provenant de sources différentes : i) de données provenant d'un seul **corpus de terrain*** (cas le plus standard) ou de plusieurs corpus de terrain (réalisés par la même personne/la même équipe ou par des acteurs différents) ; ii) de données d'un ou de plusieurs **corpus** (physiquement) **traités*** ; iii) de données provenant d'autres sources (c'est-à-dire d'autres **fonds audiovisuels***, etc.) ; iv) de **documents audiovisuels*** faisant partie d'un corpus ou de plusieurs **corpus** déjà **publiés*** (sur un site tel que celui des AAR) et qui seront soumises à des nouvelles analyses, des ré-descriptions, des réinterprétations, etc.

3/ Le corpus d'analyse formant l'input de la phase de l'**analyse*** dans le processus de travail de production-publication audiovisuelle, doit être distingué, fonctionnellement, du **corpus analysé** qui constitue le résultat d'une analyse (description, indexation, etc.).

4/ Il est également opportun de distinguer le corpus d'analyse du **corpus de traitement***, c'est-à-dire du corpus de données audiovisuelles sélectionnées pour être soumis à des modifications *techniques* ou auctoriales qui modifient l' « apparence » et la linéarité d'une donnée audiovisuelle.

Corpus analysé

1/ Ensemble de données audiovisuelles analysées (décrites, indexées, commentées, traduites/adaptées, etc.) documentant un ou plusieurs **terrains*** (un événement, une manifestation, un patrimoine, etc.).

2/ Le corpus analysé est le résultat de l'analyse de données audiovisuelles. L'analyse de données audiovisuelles est une des étapes principales du **processus de travail*** dans le cadre du **programme AAR*** et s'effectue *via* l'**Atelier de Description ASA***.

Corpus de publication

1/ Ensemble de données audiovisuelles composant l'input de l'étape de la publication à proprement parler (une des étapes principales du **processus de travail*** dans le cadre du programme AAR) et s'effectuant *via* l'**Atelier de Description ASA***.

2/ La pertinence du corpus de publication doit être évaluée par rapport aux objectifs d'une publication concrète. En tout cas, il peut être composé de données de nature

différente et provenant de sources différentes : i) uniquement d'un seul **corpus analysé*** (cas le plus récurrent dans le cadre des activités « standard » du programme AAR*) ; ii) de plusieurs corpus analysés* ; iii) d'un ou de plusieurs **corpus** de données **traitées*** (mais pas forcement analysées) ; iv) d'un ou de plusieurs **corpus de terrain*** (mais ni forcements traités, ni analysées) ; v) de **corpus** déjà **publiés***, etc.

Corpus publié

1/ L'ensemble de données audiovisuelles accessibles au public en général ou à telle ou telle catégorie spécifique du public sous forme d'un **genre de publication*** choisi (dans le cadre du programme AAR sous forme, par exemple, d'un site d'événement, d'un dossier thématique, d'un vidéo-livre interactif, d'un dossier bilingue, etc.).
2/ Il y a une distinction fonctionnelle à faire entre le **corpus de publication** et le corpus publié. Le corpus de publication réunit l'ensemble des données audiovisuelles servant d'input au processus de publication *via* l'atelier de publication ASA. Mais l'éditeur/auteur est libre de choisir, à l'intérieur du corpus audiovisuel de publication tel ou tel élément qu'il souhaite réellement publier au détriment d'autres éléments publiables mais pas retenu par l'éditeur/auteur (en revanche, dans un nouvel processus de publication le même éditeur/auteur ou une autre personne assumant ce rôle peut revenir sur les éléments non-publiés mais publiables pour en faire une nouvelle publication, etc.). Le corpus publié ne réunit que les éléments effectivement publiés enligne sous forme de tel ou tel **genre de publication***.

Corpus de terrain

1/ L'ensemble des données collectées ou produites à l'issu d'un **terrain*** consacré à la captation d'informations pour documenter une manifestation scientifique ou culturelle ou encore une recherche « de terrain » à proprement parler (comme, par exemple, une fouille archéologique, une enquête sociologique, le recueil d'expressions orales dans le cadre d'une recherche ethno-littéraire, etc.).
2/ Dans le cadre du programme AAR, la collecte des données audiovisuelles et la constitution du corpus de terrain est régie par un ensemble de principes explicités dans un document en ligne intitulé « collecte et conservation des données audiovisuelles » (document pouvant être consulté sur le portail AAR : http://www.archivesaudiovisuelles.fr/FR/about4.asp).

Corpus de traitement

1/ Ensemble de données sélectionnées pour être soumis à un traitement technique et/ou auctorial. Le *traitement technique* recouvre les activités du *derushage*, du nettoyage physique des fichiers contenant les données réunies dans un corpus de

terrain, l'amélioration de la qualité visuelle ou sonore des fichiers, etc. Le *traitement auctorial* est surtout concerné par le montage de données audiovisuelles (selon un scénario de montage, la création de transitions et d'effets spéciaux, la postsynchronisation son/image, l'ajout de commentaires en voix *off*, et ainsi de suite.
2/ Les données composant un corpus de traitement peuvent provenir d'un ou de plusieurs **corpus de terrain*** (cas le plus récurrent) mais elles peuvent aussi provenir d'autres **fonds audiovisuels***, voire de données déjà traitées et/ou publiées.
3/ Le corpus de traitement constitue l'un des deux inputs principaux de la phase du traitement/de l'analyse dans le **processus de travail*** de production-publication audiovisuelle (l'autre input étant formé par le **corpus d'analyse***).

Corpus traité

1/ Ensemble de données audiovisuelles issues de la phase du traitement (technique ou auctoriale) dans le **processus de travail*** de production-publication audiovisuelle – données audiovisuelles documentant un ou plusieurs **terrains*** (un événement, une manifestation, un patrimoine, etc.).

D

Description audiovisuelle (Type de -)

1/ La description audiovisuelle est un type particulier de description d'un corpus audiovisuel qui s'intéresse prioritairement au plan visuel, sonore et audiovisuel d'un texte audiovisuel au détriment de son contenu. Elle peut s'intéresser, par exemple, aux différents angles de prise de vue d'une scène pro-filmique, aux différents plans de cadrage et de mouvements de caméra ; elle peut s'intéresser aussi au bruitage, aux *soundscapes* typiques de tel ou tel type de scènes (comme, par exemple, de certains lieux ou accompagnant telle ou telle pratique sociale, etc.). On voit bien qu'une description audiovisuelle systématique débouche à ce qu'on pourra appeler une bibliothèque de *motifs visuels* et/ou *acoustiques* et/ou *audiovisuels*, c'est-à-dire des sortes de stéréotypes récurrents qui caractérisent l'écriture d'un corpus audiovisuel.
2/ Comme toute autre activité de description, aussi la description audiovisuelle se réalise *via* des formulaires interactifs mis à la disposition de l'analyste dans l'**Atelier de Description*** ASA.
3/ Ce type d'analyse doit être distinguée de la **description de la mise en scène (ou de l'expression) audiovisuelle** d'un **sujet** qui est une tâche spécifique composant le type d'analyse thématique (du contenu) d'un texte audiovisuel.

Description contrôlée (Procédure de -)

1 / Contrairement à la **procédure de la description libre***, la description contrôlée est une procédure qui s'appuie exclusivement sur un **thesaurus*** de termes prédéfinis (de « descripteurs ») pour désigner le **domaine de connaissances*** traité dans une vidéo, un document audiovisuel.
3/ La procédure de la description contrôlée se base sur plusieurs activités de description (chaque activité étant définie dans le **méta-lexique hiérarchique des termes conceptuels*** de l'analyse sémiotique – méta-lexique qui constitue une des **ressources métalinguistiques ASA*** nécessaires à l'élaboration des **modèles de description***). La seule activité obligatoire, pour cette procédure, est celle de la sélection d'au moins un terme prédéfini dans une liste de tels termes pour identifier et, parfois, classifier l'objet de la description. Comme toute autre activité de description, aussi celle-ci se réalise *via* des formulaires interactifs mis à la disposition de l'analyste dans l'**Atelier de Description*** ASA.

Description libre (Procédure de -)

1/ La description libre est une des procédures de description les plus importantes et forme, ensemble, avec la **description contrôlée*** (via un **thesaurus***) la procédure la plus centrale dans le cadre du **projet ASA-SHS***.
2/ « Description libre » veut bien dire que c'est à l'analyste de saisir la valeur appropriée pour qu'un **terme conceptuel*** puisse représenter adéquatement un domaine de connaissances thématisé sous forme d'un **sujet*** dans un corpus audiovisuel. Dans la **description contrôlée***, ces valeurs sont prédéfinies sous forme de termes ou de descripteurs faisant partie d'un **thesaurus***.
3/ La procédure de la description libre se base sur plusieurs activités de description (chaque activité étant définie dans le **méta-lexique hiérarchique des termes conceptuels*** de l'analyse sémiotique – méta-lexique qui constitue une des **ressources métalinguistiques ASA*** nécessaires à l'élaboration des **modèles de description***). La seule activité obligatoire, pour cette procédure, est celle de la saisie d'une expression (linguistique) minimale pour désigner, dénommer d'une manière appropriée le domaine de connaissance traité dans une vidéo, un document audiovisuel. Comme toute autre activité, aussi celle-ci se réalise *via* des formulaires interactifs mis à la disposition de l'analyste dans l'**Atelier de Description*** ASA.

Description de mise en discours (Tâche de -)

1/ La description de la mise en discours fait partie de la **description thématique***. Elle s'intéresse plus particulièrement au cadrage discursif d'une **topique***, c'est-à-dire au fait comment l'auteur (l'énonciateur) traite une topique choisie et le *transforme en un thème propre à son discours* : son point de vue, la mobilisation ou non d'autres « voix » sur le sujet, le niveau choisi de précision, etc.
2) Elle forme une tâche spécifique du type d'analyse appelé **analyse thématique*** ou encore analyse du contenu audiovisuel d'un texte ou d'un corpus de textes

audiovisuels. D'autres tâches spécifiques sont : la **description référentielle***, la **description de l'expression audiovisuelle*** ou encore le commentaire de l'analyste.
3) La (tâche) de la description de mise en discours doit être distingué de l'analyse du discours (filmique) à proprement parler. Cette dernière s'intéresse à l'élucidation des actes de discours et de la rhétorique propre à un texte ou un corpus de textes audiovisuels. La description de la mise en discours, en revanche, s'intéresse à l'explicitation des spécificités d'une **topique*** en tant que thème du discours de son auteur.

Description de mise en scène (d'expression) audiovisuelle (Tâche de -)

1/ La description de la mise en discours fait partie de la **description thématique***. Elle s'intéresse plus particulièrement à l'explicitation des « stratégies » de mise en scène audiovisuelle d'une **topique*** sélectionnée et transformée en un thème de discours dans un texte ou un corpus de textes audiovisuels.
2/ Elle forme une tâche spécifique du type d'analyse appelé **analyse thématique*** ou encore analyse du contenu audiovisuel d'un texte ou d'un corpus de textes audiovisuels. D'autres tâches spécifiques sont : la **description référentielle***, la **description de la mise en discours*** d'une topique ou encore le commentaire de l'analyste.
3/ La description de la mise en scène audiovisuelle doit être distinguée de la **description audiovisuelle*** *stricto sensu*. Cette dernière s'intéresse prioritairement au plan visuel, sonore et audiovisuel d'un texte audiovisuel au détriment de son contenu.

Description paratextuelle (Type de -)

1/ La description paratextuelle est un type de description spécifique qui explicite l'identité formelle d'un objet d'analyse (d'une vidéo, d'un segment d'une vidéo d'une image, etc.), c'est-à-dire son intitulé, son/ses auteurs, son genre, éventuellement le lieu et la date de publication, etc. Elle précise également les questions de droits (copyright, droits d'exploitation, etc.) ainsi qu'éventuellement l'appartenance de l'objet analysé à tel ou tel « grand sujet » (tout en déléguant à la description thématique l'explicitation à proprement parler du sujet).
2/ La description paratextuelle offre, en gros, une identification/présentation d'un objet audiovisuel qui correspond aux 15 critères de description définissant le standard Dublin Core.
3/ Comme toute autre activité de description, aussi la description paratextuelle se réalise *via* des formulaires interactifs mis à la disposition de l'analyste dans l'**Atelier de Description*** ASA.

Description pragmatique (Type de -)

1/ La description pragmatique est un type de description spécifique qui s'intéresse à expliciter les usages potentiels d'un texte audiovisuel formant son objet et de l'adapter, dans la mesure du possible (et sans **traitement*** à proprement parler de l'objet dans sa « matérialité »), au profil, aux attentes d'un public donné. L'adaptation peut prendre la forme d'une traduction (littérale, synthétique, libre, …) et/ou la forme d'un « versionnement » intellectuel, culturel (comme c'est le cas, par exemple, dans le cadre de la vulgarisation d'un contenu particulièrement spécialisé afin de le rendre compréhensible à un public plus large).
2/ Comme toute autre activité de description, aussi la description pragmatique se réalise *via* des formulaires interactifs mis à la disposition de l'analyste dans l'**Atelier de Description*** ASA.

Description référentielle (Tâche de -)

1/ La description référentielle s'intéresse à l'identification et à la désignation des *objets* ou *domaines de connaissances* ainsi qu'aux *contextes* (*spatiotemporels* et/ou *thématiques*) qui sont sélectionnés dans un texte audiovisuel et qui se constituent sous forme de **topiques***.
2/ Elle forme une tâche spécifique du type d'analyse appelé **analyse thématique*** ou encore analyse du contenu audiovisuel d'un texte ou d'un corpus de textes audiovisuels.

Description thématique (Type de -)

1/ La description thématique est un type de description spécifique qui s'intéresse à l'explicitation des **sujets***, c'est-à-dire du contenu ou encore du message d'un objet audiovisuel, c'est-à-dire d'un texte ou d'un corpus de textes audiovisuels.
2/ La description thématique recouvre plusieurs types de description plus spécialisés : **description référentielle***, **description de mise en discours***, **description** de mise en scène (ou de l'expression) **audiovisuelle***, **commentaire**.
2/ La description thématique s'effectue *via* une bibliothèque de **modèles de description*** du contenu d'objets audiovisuels qui sont supposés de documenter tel ou tel aspect du domaine de connaissances d'une vidéothèque, d'un portail, ou, plutôt, d'un programme de production-publication audiovisuelle.
3/ C'est, bien sûr, la description thématique qui varie le plus entre les différents domaines de connaissances, contrairement aux autres types de description (**paratextuelle***, **audiovisuelle***, **pragmatique***, péritextuelle, etc.) qui sont relativement indépendants des spécificités référentielles de tel ou tel domaine particulier. Ainsi, la mise en œuvre d'une description thématique systématique présuppose un travail très important en termes de modélisation du domaine de connaissances à prendre en considération et, puis, de la définition et mise en œuvre d'un métalangage de description qui soit approprié aux spécificités du domaine en question.

2/ Comme toute autre activité de description, aussi la description pragmatique se réalise *via* des formulaires interactifs mis à la disposition de l'analyste dans l'**Atelier de Description*** ASA.

Document audiovisuel

1/ Terme qui désigne un **objet audiovisuel*** en principe *traité, analysé et publié*, c'est-à-dire un objet audiovisuel à qui est conféré, à travers un ensemble d'activités faisant partie du **processus de travail*** de production-publication audiovisuelle, le *statut* et la *fonction* de (justement et au moins dans les yeux de ses auteurs) *documenter quelque chose*, de renseigner, d'apporter une information, une connaissance, de pouvoir satisfaire une curiosité, etc. (pour plus d'explications voir nos livres [STO 99, STO 03]).
2/ Un document audiovisuel ne nous est accessible que sous forme d'un **genre de publication*** particulier. Dans le cadre des activités de publication du programme AAR*, il nous est accessible, par exemple, sous forme d'un **vidéo-livre interactif***, d'un documentaire, d'un reportage, d'un **dossier thématique***, d'un **dossier pédagogique*,** etc.
3/ Le document audiovisuel est un des résultats « tangibles », « visibles » du **corpus publié*** (dans le cadre du programme AAR, du corpus publié en ligne).
4/ Le document audiovisuel numérique n'est pas obligatoirement un texte audiovisuel statique, figé, définitif. Au contraire, notamment grâce à la **republication***, le document audiovisuel peut évoluer dans le temps, s'enrichir, changer de forme, etc.

Domaine de connaissance/d'expertise

1/ Le domaine de connaissance est l'univers référentiel couvert par le **métalangage de description*** et, plus particulièrement, la bibliothèque des **modèles de description*** d'un **fonds audiovisuel*** donné.
2/ L'expression « domaine de connaissance/d'expertise » veut bien indiquer qu'il faut distinguer entre le domaine tel qu'il est *représenté cognitivement* par le **modélisateur*** (= *domaine de connaissance au sens d'une expertise plus ou moins fiable, plus moins experte, plus ou moins admise collectivement, etc.*) et le *domaine du réel* supposé correspondre au domaine de connaissance sous forme d'une expertise !
3/ Dans le cadre du projet ASA-SHS, plusieurs domaines de connaissance ont été définis et explicités sous forme de métalangages de description spécifiques (aussi appelés **ontologies de domaine***). Il s'agit des domaines de connaissance des principaux ateliers d'expérimentation du projet ASA-SHS que sont les ateliers ARC, ADA, ALIA, PCIA et FMSH-AAR.

Dossier bi/multilingue

1/ Le dossier bi/multilingue est un genre particulier de publication spécifié et développé pour rendre compte expressément de la problématique comment diffuser un enregistrement audiovisuel monolingue sur un marché de connaissance qui est, lui, intrinsèquement multilingue.
2/ Le dossier bi/multilingue propose des versions (plus ou moins fidèles ou, au contraire, raccourcies, approximatives, adaptées, simplifiées, etc.) d'un enregistrement audiovisuel (réalisé dans une langue donnée) en une ou plusieurs langue cible.
3/ Ce genre publication a été expérimenté par l'**ESCoM*** pour le **programme AAR*** dans le cadre de plusieurs projets de recherche français et européens (voir notamment **SAPHIR***) Aujourd'hui, il existe toute une série de dossiers bilingues (de type français/espagnol ; français/anglais ; français/russe ; français/arabe ; etc.) qui peuvent être visionnés soit sur le site **portail AAR***, soit sur les portails thématiquement circonscrits tels que **DLC***, PCM* ou **SCC***.

Dossier pédagogique

1/ Le dossier pédagogique est un genre particulier de publication qui utilise/réutilise des enregistrements audiovisuels (traités ou non, analysés ou non) pour en faire des ressources pédagogiques à proprement parler soit pour l'enseignement formel, soit pour un apprentissage non-formel ou informel.
2/ En règle générale, le dossier pédagogique est organisé autour d'une série de chapitres où chaque chapitre est supposé représenter une étape dans l'acquisition, l'appropriation d'un certain type de connaissances ou de savoir-faire. Chaque chapitre est pourvu d'un matériel audiovisuel et d'un ensemble d'autres informations utiles à l'apprenant ou à l'enseignant. Le dossier lui-même comporte des guides à destination de l'enseignant et de l'apprenant ainsi que de propositions de tests et de validation des connaissances.
2/ Ce genre publication a été expérimenté par l'**ESCoM*** pour le **programme AAR*** dans le cadre du projet français **SAPHIR***. Aujourd'hui, il existe toute une collection de dossiers pédagogiques couvrant une variété de sujets et qui peuvent être visionnés sur le site **portail AAR***.

Dossier thématique

1/ Le dossier thématique est un genre particulier de publication qui réunit, comme son nom l'indique déjà, autour d'un sujet, d'un thème des contributions audiovisuelles différentes. Selon les objectifs affichés de communication, le dossier thématique peut prendre différentes allures : dossiers qui fait le point sur une question, dossier qui ouvre un débat, dossier de controverses, dossier de sensibilisation, etc.
2/ Ce genre publication a été expérimenté par l'**ESCoM*** pour le **programme AAR*** dans le cadre de plusieurs projets de recherche français et européens (voir notamment **SAPHIR*)**. Aujourd'hui, il existe toute une collection de dossiers

thématiques couvrant une variété de sujets et qui peuvent être visionnés sur le site **portail AAR***.

F

Facette (de sens)

1/ Terme qui désigne une dimension sémantique ou, plutôt, un *classème* (au sens de la sémantique structurale de Greimas [GRE 66])
2/ Une facette définit le rang des valeurs possibles d'un **terme conceptuel*** ou d'une configuration de termes conceptuels.

Fonds audiovisuel

1/ Le terme « fonds audiovisuel » désigne l'ensemble des données audiovisuelles dont dispose une archive audiovisuelle et dont se « nourrissent » (au moins en partie) les différents types fonctionnellement distincts de **corpus audiovisuels*** qu'on peut rencontrer dans le processus de travail de la production-publication audiovisuelle.

Formulaire interactif (de travail)

1/ Le formulaire interactif est l'interface de travail de l'**analyste*** (ou, selon le cas, de l'éditeur-auteur*, du **modélisateur***, …) lui permettant d'utiliser, dans son travail d'**analyse*** (ou de publication), les différents modèles composant le **métalangage*** (de description, de publication, etc.) pour traiter le **domaine de connaissance/d'expertise*** d'une archive audiovisuelle donnée.
2/ Ainsi, l'**Atelier de Description ASA*** est composé d'une série de formulaires interactifs permettant à un **analyste*** de procéder à une **méta-description***, à une **description audiovisuelle***, à une **description thématique***, etc. L'**Atelier de Publication ASA***, lui, est aussi composé de formulaires interactifs permettant à un **éditeur-auteur*** de « téléverser » (*upload*) de données audiovisuelles, de choisir de modèles de publication, d'importer de corpus analysés pour les publier, d'adapter un modèle de publication, etc.

G

Genre de publication

1/ Un genre de publication est un *modèle culturellement et historiquement situé* qui donne à un texte une certaine forme, une certaine « gestalt » reconnaissable par ceux/celles qui possèdent une compétence (de lecture, de compréhension) appropriée. Généralement parlant, un genre de publication peut être identifié par le recours simultané à une série de critères tels que le contenu, l'organisation narrative mais aussi formelle et physique, la mise en scène audiovisuelle, etc.
2/ Le cadre du **programme AAR***, utilise plusieurs genres de publication dont, notamment, le **vidéo-livre interactif***, le **dossier thématique***, le dossier bilingue, le vidéo-lexique. Chaque genre de publication est explicitement décrit. Il est converti en un modèle à l'aide duquel l'éditeur-analyste, *via* un formulaire interactif, peut envisager la **publication*** ou la **republication*** d'un corpus audiovisuel donné.

M

Métalangage de description

1/ Ensemble structuré de modèles de description utilisés sous forme de **formulaires interactifs*** dans l'**Atelier de Description*** ASA par l'analyste lors de son travail sur un corpus audiovisuel, un texte audiovisuel pris isolément ou encore tel ou tel passage (segment) spécifique d'un texte audiovisuel.
2/ La théorie sémiotique du texte audiovisuel sert de cadre de référence à l'élaboration d'un **métalangage de description*** ASA. Ainsi, conformément à ce cadre théorique, le **métalangage de description*** distingue plusieurs types fonctionnels de modèles de description dont notamment : (i) une classe de modèles de description réservée à la *production de la méta-description* elle-même (explicitation du contenu, des objectifs, des auteurs, du public visé, etc. d'une analyse concrète) ; (ii) une classe de modèles de description réservée à l'explicitation des *données paratextuelles* des objets audiovisuels soumis à l'analyse intitulé de l'objet, auteur(s), genre, langue, propriété intellectuelle, etc. ; (iii) une classe importante de modèles consacrée à l'analyse du *contenu* lui-même véhiculé par un corpus audiovisuel ; (iv) une classe de modèles consacrée plus particulièrement à la *mise en scène audiovisuelle du contenu* véhiculé par un corpus audiovisuel (modèles servant à l'analyse des plans visuels et acoustique) ; (v) une classe de modèles réservée à l'*adaptation contextuelle* et *linguistique* d'un corpus audiovisuel. Le métalangage de description ASA est, en d'autres termes, une **ontologie générique***, appelée **ontologie ASA***.
3/ Dans le cadre du projet ASA-SHS et de ses différents champs d'expérimentation, ont été élaborés des métalangages de description pour six domaines de connaissances/d'expertises correspondant aux principaux **ateliers d'expérimentation*** du projet. Ces six domaines ont en commun tous les modèles de description appartenant au type (i), au type (ii), au type (iv) et au type (v). Seuls les modèles de description de type (iii) varient systématiquement entre les six

ateliers. Chacun de ces ateliers possèdent donc ses propres modèles de description du contenu audiovisuel adapté à leur domaine de connaissance/d'expertise. Ces métalangages constituent ce qu'on appelle des **ontologies de domaine*** formées à partir d'une **ontologie générique*** qui est l'**ontologie ASA***.

Méta-description

1/ La méta-description est un type de description spécifique qui permet à l'analyste de « prendre la parole ». Ainsi, peut-il, par exemple, expliciter mieux le contenu et l'objectif de son analyse, la valeur ou l'intérêt des objets analysés, etc.
2/ La méta-description se présente d'une part sous forme d'un type d'analyse au même titre que l'**analyse thématique*** (du contenu audiovisuel) ou l'**analyse pragmatique***, et d'autre part, sous forme d'une tâche spécifique de la **description thématique***. Dans le premier cas, la méta-description a comme tâche d'expliciter contenu, problèmes adressés et objectifs d'une analyse (du contenu, audiovisuelle ou autre ; cf. [CHE 11a]) ; dans le deuxième cas, elle sert à l'analyste de commenter (mettre en valeur, critiquer, etc.) soit une topique analysée, soit l'analyse elle-même de la topique.

Méta-lexique ASA

1/ Le méta-lexique ASA forme une des ressources métalinguistiques indispensables dont nous avons besoin pour élaborer des métalangages de description qui sont spécifiques à l'univers du discours de telle ou telle archive – métalangages de description sous forme d'une **bibliothèque de modèles de description***. Il s'agit d'un vocabulaire hiérarchique – d'une ontologie – de **termes conceptuels*** (de « concepts ») qui servent d'input à la définition des différentes **briques*** dont se composent un modèle de description.
2/ Le méta-lexique ASA se compose en réalité de deux méta-lexiques parfaitement complémentaires : a) le méta-lexique qui identifie, désigne et classifie dans une structure taxinomique tous les *objets d'analyse de l'**univers du discours**** *ASA* ; b) le méta-lexique qui identifie, désigne et classifie dans une structure taxinomique toutes les ***activités d'analyse**** dont nous nous servons pour décrire le contenu d'un texte audiovisuel, sa mise en discours, son expression audiovisuelle, etc.

Micro-thesaurus ASA

1/ Un micro-thesaurus est constitué d'une **facette*** interprétée par une liste (hiérarchique) d'expressions normalisées (de descripteurs).
2/ Le micro-thesaurus est utilisé par la procédure de la **description contrôlée*** où il remplace la saisie libre par l'analyste de l'expression minimale pour désigner d'une manière appropriée un terme ou une configuration de termes conceptuels.

Modèle de collecte

1/ Un modèle de collecte (de données audiovisuelles, etc.) explicite les critères à prendre en considération lors de la constitution de **corpus de terrain***.
2/ Parmi les critères les plus importants, on compte, par exemple, les *critères de couverture empirique* (c'est-à-dire du domaine référentiel de connaissances), les *critères textuels* et *sémiotiques à proprement parler* (c'est-à-dire relatifs aux types et genres de données à collecter), les *critères paratextuels* (c'est-à-dire relatifs à l'identité formelle des données à collecter).
3/ L'élaboration des modèles de collecte repose sur un **métalangage de description*** du domaine de connaissances/d'expertise.
4/ Des modèles de collecte ont été définis pour le **programme AAR*** dans le cadre de l'organisation et de la conduite d'entretiens de recherche (c'est-à-dire consacrés à la recherche d'un invité/d'une invitée) avec une durée pouvant aller jusqu'à plusieurs heures de conversation. L'organisation, la conduite et le tournage de ces entretiens constituaient jusqu'en 2009 un terrain prioritaire du **programme AAR*** qui a été donc explicitement structuré par un modèle de collecte se matérialisant, pour le **producteur de terrain*** (rôle incluant la ou les personnes menant l'entretien), sous forme : i) de *procédures à suivre* ; ii) d'*outils* de *collecte* et de *gestion des données* collectées (dont la documentation à caractère juridique accompagnant un entretien) ; iii) du *script de préparation* de l'entretien ; iv) du *script de conduite* de l'entretien ; v) du *script d'analyse de l'entretien* servant d'input direct au **traitement*** et à l'**analyse*** à proprement parler du corpus documentant un entretien concret.

Modèle de description

1/ Un modèle de description est une hypothèse portant sur la « meilleure façon » de décrire un objet concret. Il fait partie du **métalangage de description*** d'un domaine de connaissances/d'expertise.
2/ Un modèle de description se compose d'un ensemble de ressources métalinguistiques dont notamment de termes conceptuels* organisés en schémas* et séquences*.
3/ Dans le cadre du projet ASA-SHS, plusieurs bibliothèques de modèles de description ont été élaborés et testés, une bibliothèque de modèles correspondant à un domaine de connaissance/d'expertise.
4/ A côté de ces bibliothèques de modèles spécifiques à un **domaine de connaissance*** (tel que celui de l'**atelier ALIA***), a été identifié, défini et réalisé une *bibliothèque commune* de modèles de description qui est *indépendante* d'un domaine de connaissances et qui peut donc être utilisée pour l'analyse de n'importe quel corpus audiovisuel. Il s'agit de modèles guidant la **description paratextuelle*** d'un texte audiovisuel, sa **description audiovisuelle*** stricto sensu ainsi que sa **description pragmatique***.

Modèle de publication

1/ Un modèle de publication explicite un **genre de publication*** particulier à l'aide duquel un **corpus de publication*** peut être publié et diffusé en ligne. Des genres de publication particuliers sont, par exemple, le **vidéo-livre interactif***, le **dossier thématique***, le **dossier bilingue***, etc.
2/ D'une manière similaire aux modèles de description, aussi les modèles de publication font partie du métalangage de description ASA où ils forment une bibliothèque à part.
3/ Dans le cadre du projet ASA-SHS, l'accent a été mis plutôt sur le développement des modèles de description et moins sur celui des modèles de publication. Ainsi, les modèles de publication se présentent actuellement sous une forme encore assez rudimentaire et figée.

Modélisateur (rôle de- ; aussi : « Concept designer »)

1/ Avec le rôle de l'**analyste*** et du **rédacteur-auteur***, un des trois rôles principaux identifiés dans le cadre du **projet ASA-SHS***.
2/ Le modélisateur (aussi appelé « concept-designer ») est le rôle à qui revient la délicate – et fort difficile tâche – de la définition et réalisation des **modèles de description*** de corpus audiovisuels documentant un **domaine de connaissance/d'expertise***. Ce travail présuppose (parfois) l'intervention dans les **ressources métalinguistiques*** ASA sous forme d'ajouts locaux ou de création de **ressources « utilisateurs »** (contrairement aux **ressources communes**) si le développement de nouveaux **modèles de description*** l'exige.
3/ Le modélisateur est supposé connaître parfaitement bien les ressources métalinguistiques sans lesquels aucun **modèle de description*** ou de **publication*** et aucun **formulaire interactif*** dans l'**Atelier de Description*** et de **Publication*** ne pourrait être réalisés.
4/ Mais le modélisateur est supposé être familier également avec les techniques de l'analyse conceptuelle, de la description/modélisation des connaissances et des approches et disciplines telles que les sciences cognitives, l'intelligence artificielle, la sémiotique et, d'une manière plus générale, les sciences du langage. Il doit, enfin, savoir dialoguer d'une part avec les personnes et équipes ayant en charge le développement technique des modèles de description et de publication, et d'autre part avec toutes les parties prenantes dans un projet de collecte, de traitement, d'analyse et de publication de corpus audiovisuels documentant un patrimoine scientifique ou culturel.

O

Objet (d'analyse)

1/ L'objet d'analyse est l'objet, l'entité sur quoi porte une **activité d'analyse*** (une description, un commentaire, une interprétation, etc.).
2/ Dans le cadre du projet ASA-SHS, nous distinguons entre différentes classes d'objets d'analyse : i) les objets d'analyse composés des référents d'un **domaine de connaissance/d'expertise*** propre à une vidéothèque/un portail ; ii) les objets servant à la localisation spatiale et temporelle des référents ; iii) les objets de nature discursive et énonciative servant à donner une vision particulière aux référents thématisés dans un texte audiovisuel ; iv) les objets de mise en scène ou de l'expression audiovisuelle des référents thématisés ; v) et les objets servant à réaliser un métadiscours (un commentaire, une appréciation, etc.) soit sur l'acte de l'analyse soit sur l'objet de l'analyse (le texte audiovisuel).

Objet audiovisuel (voir : Texte audiovisuel)

Ontologie générique ASA (voir : Métalangage ASA)

Ontologie de domaines ASA

1/ Une ontologie de domaine ASA est un métalangage de description élaboré pour analyser des corpus audiovisuels qui documentent un domaine de connaissances/d'expertise particulier.
2/ Une ontologie de domaine repose sur l'ontologie générique ASA de laquelle elle emprunte des termes conceptuels pertinents, des modèles de description ainsi que des parties du thesaurus commun tout en y ajoutant, si nécessaire, ses propres ressources métalinguistiques. Les ajouts métalinguistiques propres à un domaine composent une branche spéciale dans le méta-lexique des termes conceptuels ASA, des modèles de description ASA ainsi que dans le thesaurus commun ASA.
2/ Dans le cadre du projet ASA-SHS, des ontologies de domaine ont été définies pour les ateliers ARC, ADA, ALIA, PCIQ et FMSH/AAR.

P

Producteur de terrain

1/ Un des rôles principaux dans le **processus de travail*** caractérisant la constitution, le traitement/l'analyse et la publication/ diffusion de patrimoines de connaissances. Il identifie la ou les personnes ayant en charge la préparation et la réalisation d'un **terrain*** (« terrain » *lato sensu* incluant également les tournages d'événements circonscrits et pas seulement les « terrains » au sens de « terrains de recherche ») ainsi que la constitution du **corpus*** documentant le terrain.

2/ Le rôle du producteur de terrain se différencie en des rôles plus spécialisés : rôles liés à la conception/définition d'un terrain, rôles liés à la conduite d'un terrain, rôles techniques (il s'agit, par exemple des cadreurs, des preneurs de son, etc. qui assurent l'enregistrement audiovisuel d'un terrain).

Processus de travail (de production-publication audiovisuelle)

1/ Dans le cadre du **programme AAR***, le processus de travail permettant la constitution, publication et diffusion d'un patrimoine scientifique ou culturel est décomposée en cinq étapes principales : i) l'étape des activités préparatrices en amont de l'enregistrement d'un terrain (*lato sensu*) ; ii) l'étape de l'enregistrement d'un terrain (*lato sensu*) et de la collecte de toutes les données documentant celui-ci ; iii) l'étape du traitement et de l'analyse du corpus audiovisuel ; iv) l'étape de la publication audiovisuelle ; v) l'étape clôturant le processus de travail.
2/ Chaque étape est composée d'un ensemble de tâches et d'activités précises instrumentées et décrites dans la documentation technique qui accompagne le processus de travail.

Procédure d'analyse

1/ Une procédure d'analyse est une tâche composée d'une ou de plusieurs activités de description (chacune étant définie dans le **métalangage***, l'**ontologie générique*** ASA).
2/ Dans le cadre du projet ASA-SHS, ont été définies deux procédures d'analyse (de description) de base que sont la **description contrôlée*** et la **description libre***. Une troisième procédure est la procédure composite s'appuyant sur l'une et l'autre des deux procédures de base.

R

Rédacteur-auteur (rôle de-)

1/ A côté des rôles du producteur de terrain, de l'**analyste*** et du **modélisateur***, le rédacteur-auteur représente un troisième rôle identifié, problématisé et instrumenté dans le cadre du **projet ASA-SHS***.
2/ Le rédacteur-auteur intervient dans l'étape de la publication du **processus de travail*** de production-publication audiovisuelle. L'analyste et le rédacteur-auteur peuvent être « joués » par la même personne, le même groupe mais ce n'est pas, bien sûr, toujours le cas. De même, le rédacteur-auteur lui-même étant un rôle, celui-ci peut être joué par une seule personne à un moment donné, mais il peut être joué

par un collectif de personne, par la même personne ou différentes personnes distribuées dans l'espace et le temps.
3/ Concrètement parlant, le rédacteur-auteur se sert de l'**Atelier de Publication*** du **Studio ASA*** pour : i) se constituer son corpus de publication ; ii) sélectionner et, dans les limites techniques actuelles que présente l'**Atelier de Publication*** ASA, adapter le genre de publication à ses besoins ; iii) préparer son corpus pour la publication (sélection des éléments effectivement à publier, vérifier les métadonnées, ajouts de « nouvelles pages », etc.) ; iv) publier son corpus.

Relation conceptuelle

1/ Une relation conceptuelle représente un type spécifique de rapports s'établissant entre des objets (d'analyse)* et représentés par des **termes conceptuels*** ou, **concepts***.
2/ Les relations conceptuelles font partie du **métalangage ASA*** qui en distingue différents types. La distinction la plus importante étant celle entre i) les relations qui définissent les *rapports entre objets d'un domaine de connaissance/d'expertise* et ii) les relations qui définissent *les rapports entre **objets d'analyse*** et **procédures d'analyse****.
2/ Une relation conceptuelle sert à définir des configurations entre **termes conceptuels*** formant le **métalangage ASA***. Parmi les types les plus importants de configurations qui constituent le métalangage ASA, nous comptons les **schémas de définition***, les **séquences de description***, les schémas référentiels, les **schémas d'indexation*** à proprement parler et, enfin, les **modèles de description***.

Ressource audiovisuelle

1/ Terme désignant tout texte audiovisuel qui représente pour un public donné une **valeur*** cognitive et pratique comme, par exemple, de satisfaire une curiosité ou un besoin d'information.

Ressources métalinguistiques

1/ Terme qui désigne tout élément faisant partie du **métalangage*** ASA et qui aide le modélisateur dans sa tâche de mettre en place un métalangage ou une **ontologie de domaine***, c'est-à-dire de **modèles de description*** appropriés à un **domaine de connaissance/d'expertise***.

S

Schéma (conceptuel) de définition

1/ Le schéma de définition est une *micro-configuration* de **termes conceptuels*** (composée à la limite d'un seul terme conceptuel) qui constitue, avec d'autres schémas conceptuels une **séquence*** d'un modèle de description. Le schéma conceptuel permet donc de créer de **relations*** (autre que simplement taxinomiques) entre un, deux ou plusieurs **termes conceptuels***. Il fait partie du métalangage de description ASA.
2/ Les **ressources métalinguistiques ASA*** réunissent notamment trois classes complémentaires de schémas conceptuels : a) la classe des schémas définissant les **objets d'analyse*** qui font partie de l'**univers du discours*** d'une archive ; b) la classe des schémas définissant les procédures (**libres***, **contrôlées***, mixtes, standards ou simplifiées, etc.) de description d'un objet d'analyse, voire d'une **topique***, de sa contextualisation, de sa mise en discours, de son expression audiovisuelle, etc. ; c) la classe des schémas qui fixent « à priori » la valeur référentielle d'un **terme conceptuel*** (par exemple, si le domaine référentiel de connaissance est limité à la littérature française du Moyen Âge, le terme conceptuel [PERIODE] est *à priori* fixé par l'expression <Moyen-Âge> interprétée ensuite par des valeurs numériques appropriées pour représenter les bornes temporelles de cette époque).

Schéma d'indexation

1/ Le schéma d'indexation précise ce que l'analyste doit faire en choisissant une activité précise qui fait partie d'une **procédure d'analyse***. Une activité et, *à fortiori*, une procédure d'analyse peut être composées par plusieurs schémas d'indexation. Il se présente sous forme d'un **formulaire interactif*** composé de champs, de tableaux et autres éléments que l'analyste est supposé de renseigner.
2/ Les schémas d'indexation font partie du **métalangage ASA*** où ils forment une bibliothèque réunissant des collections spécialisées de schémas d'indexation *linguistique, textuelle, audiovisuelle, à l'aide d'un thesaurus, en référence à un standard* (tel que LOMFR ou Dublin Core, etc.), etc.

Séquence

1/ Un **modèle de description*** est composé de plusieurs séquences. Chaque séquence sert à la description/à l'analyse d'un type d'objets* déterminé.
2/ Dans le cadre du projet ASA-SHS, nous distinguons entre cinq types principaux de séquences : (i) les séquences qui servent à la description référentielle dce l'objet ou du domaine de connaissance thématisé dans un texte ou un corpus de textes audiovisuels ; (ii) les séquences qui servent à la description de la mise en discours de l'objet/du domaine thématisé ; iii) les séquences qui servent à la description de l'expression verbale et/ou de la mise en scène audiovisuelle du domaine thématisé ; (v) les séquences, enfin, qui servent à expliciter davantage le point de vue de l'analyste ainsi que le contenu et les objectifs de l'analyse.

3/ Une séquence est définie par deux types de **schémas de définition*** : (i) les schémas de définition de l'**objet d'analyse*** » (c'est-à-dire l'objet sur lequel porte l'analyse dans la séquence) et le schéma de définition de la **procédure d'analyse*** (c'est-à-dire les méthodes selon lesquelles un objet donné est analysé).

Sujet

1/ Un sujet c'est, intuitivement parlant, ce *dont « parle » un texte audiovisuel* ou encore ce qu'il « montre ».
2/ Dans le cadre de notre recherche portant sur le fonds du programme des **Archives Audiovisuelles de la Recherche*** (AAR), nous distinguons entre 1) les *sujets exposés oralement* et dont l'auteur (c'est-à-dire l'énonciateur) est *montré* visuellement ; et 2) les sujets *exposés visuellement* (ou « audio-visuellement »). Le premier type de sujets est représenté par les enregistrements d'entretiens, de conférences, d'allocutions, mais aussi de lectures, de déclamations, etc. - donc par les enregistrements audiovisuels d'un discours oral. Le deuxième type de sujets est représenté soit par des captations audiovisuelles de situations dites pro-filmiques soit par l'enregistrement de mises en scène.
3/ Plus techniquement parlant, nous définissons le sujet comme la thématisation d'une **configuration topique*** dans le texte, la configuration topique étant, rappelons-le, le modèle explicitant et définissant une vision d'un domaine de connaissance ou d'expertises qui fait partie de l'**univers du discours*** d'une **archive audiovisuelle***.

T

Terme conceptuel

1/ Terme conceptuel (parfois aussi appelé simplement « concept ») est une expression métalinguistique qui désigne un certain type d'**objets*** (dans le cadre du projet ASA-SHS, nous distinguons 5 types d'objets particuliers).
2/ Les termes conceptuels font partie du **métalangage de description*** ASA où ils sont organisés sous forme d'une hiérarchie de concepts. Cette hiérarchie de concepts distingue notamment entre i) les termes conceptuels représentant les **objets d'analyse*** du domaine d'expertise ASA et ii) les termes conceptuels représentant les activités et **procédures d'analyse*** ASA.
3/ Les termes conceptuels représentant les objets d'analyse (= première classe des termes conceptuels) forment la partie « domaine d'analyse » des **modèles de description*** et les termes conceptuels représentant les procédures d'analyse (= deuxième classe des termes conceptuels) forment la partie de l'analyse à proprement parler des **modèles de description***.

Terrain

1/ Le terme « terrain » désigne l'étape dans le **processus de travail*** (de production-publication audiovisuelle) qui est réservée à la collecte des données audiovisuelles documentant soit une manifestation scientifique (colloque, séminaire de recherche, etc.), culturelle (concert, exposition, etc.) ou autre (politique, sociale, etc.), voire une recherche « de terrain » à proprement parler (consacrée, par exemple, à la documentation d'un patrimoine culturel, d'une pratique sociale, etc.). Le résultat « tangible » de cette étape est le **corpus de terrain***.

2/ Dans le cadre d'un programme de production-diffusion d'un patrimoine de connaissances (tel que le **programme AAR***), l'étape du terrain constitue un des moments cruciaux (un « jalon ») du processus de travail (de production-publication audiovisuelle).

Thème (voir : Configuration thématique)

Topique ou Topos (voir : Configuration topique)

Thesaurus

1/ Le thesaurus est une des ressources principales pour la **description contrôlée***, dans le cadre du travail d'**analyse*** d'un **corpus audiovisuel***.

2/ Dans le cadre de la mise en place des différents **ateliers d'expérimentation*** du **projet ASA-SHS***, a été élaboré d'abord un thesaurus restreint et fort simple (pays du monde, périodes temporelles, langues du monde, auteurs de la littérature française, régions et départements français, etc.). Au fur et à mesure, ce premier thesaurus a été complété par des nouvelles *facettes* : une facette interprète une dimension du sens d'un **terme conceptuel*** (ou d'un **schéma*** de termes conceptuels) et est interprétée par une liste hiérarchique de termes prédéfinis (de « descripteurs »).

3/ Le thesaurus ASA fait partie des ressources du **métalangage de description*** ASA (au même titre que les indexations générées par les analystes *via* la procédure de **description libre***). On y distingue, plus particulièrement, entre un *thesaurus commun* (sous-entendu : à tout l'univers du discours ASA) et des *thesaurus propres* à l'univers du discours d'une archives particulière (par exemple, le domaine **PCIA*** possède, comme le domaine **FMSH-AAR***, ses propres facettes de thesaurus).

U

Univers du discours

1/ En adoptant une approche textuelle et discursive (popularisée par l'expression *linguistic turn* [CRA 08] et à l'origine duquel on trouve des savants tels que Michel Foucault, Michel de Certeau ou encore Jacques Derrida), une archive collecte, conserve, diffuse, transmets, etc. les discours d'une personne, d'un groupe social, d'une institution etc., sous forme de textes de toute sorte (dans notre cas, sous forme de textes audiovisuels numériques). Ces discours peuvent être limités sur un domaine bien précis mais ils peuvent également porter sur « n'importe quoi » pourvu qu'ils possèdent une pertinence, une valeur pour l'acteur détenteur de l'archive. Toujours est-il qu'ils forment ensemble un « tout » que nous appelons univers du discours (d'une archive) : tout nouveau texte s'inscrit dans cet univers tout en l'enrichissant et en le modifiant plus ou moins insensiblement ou, au contraire, d'une manière « dramatique ». Ce « tout » nommé univers du discours présente ce que certains appelle le *patrimoine*, d'autres la *mémoire*, d'autres encore la *tradition* ou l'*épistémè*, le *cadre de référence culturelle* d'une personne ou d'un groupe social.
2/ Plus techniquement parlant, l'explicitation (toujours partielle et partiale) de l'univers du discours d'une archive se fait sous forme de **modèles de description*** à l'aide desquels l'analyste se propose de décrire, d'étudier les textes (dans notre cas, *audiovisuels*) de l'archive.

V

Valeur (d'un texte audiovisuel)

1/ Terme qui désigne la capacité d'un texte audiovisuel de satisfaire un manque (un besoin, un désir, etc.) d'informations ou de connaissances auprès d'un public.
2/ L'**analyse*** d'un texte audiovisuel consiste en l'explicitation de cette valeur (dite potentielle) du texte pour un public donné (l'analyste pouvant être lui-même le public). Le cas échéant, sans ou avec un travail de traitement physique approprié, l'analyse peut conformer, c'est-à-dire adapter le profil, l'identité auctoriale du texte audiovisuel aux attentes et besoins (désirs, simples curiosités, etc.) d'un public visé.

Vidéo-lexique

1/ Le vidéo-lexique est un genre particulier de publication d'un corpus audiovisuel qui ressemble beaucoup au produit traditionnel du dictionnaire thématique : les chapitres thématiquement circonscrits d'un tel dictionnaire comportent – par ordre alphabétique – une liste d'expressions vedette qui sont définies et exemplifiées dans des articles qui leurs sont consacrés. De même, un vidéo-lexique est composé de plusieurs « chapitres » thématiquement circonscrits ; chaque chapitres contient un

ensemble d'expressions vedette et chaque expression vedette fait l'objet du discours dans de petits segments audiovisuels.
2/ Ce genre publication a été expérimenté par l'**ESCoM*** pour le **programme AAR*** dans le cadre de plusieurs projets de recherche français et européens (voir notamment **SAPHIR***). Aujourd'hui, des prototypes de vidéo-lexiques existent sur les langues et les cultures du monde. Ils sont diffusés sur le site portail AAR*****.

Vidéo-livre interactif

1/ Le vidéo-livre interactif est un genre particulier de publication d'un corpus audiovisuel. Sa structure possède des analogies avec le « livre » au sens classique du terme. Il est composé, de chapitres offrant à un lecteur intéressé de naviguer à travers un enregistrement audiovisuel (pouvant durer plusieurs heures) par « feuilletage » (c'est-à-dire comme s'il feuilletait un livre).
2/ Le vidéo-livre interactif est un des modèles de publication « standard » du **programme AAR***.