



HAL
open science

Le tournant sémiotique dans les archives audiovisuelles.

Peter Stockinger, Steffen Lalande, Abdelkrim Beloued

► **To cite this version:**

Peter Stockinger, Steffen Lalande, Abdelkrim Beloued. Le tournant sémiotique dans les archives audiovisuelles. : Vision globale et éléments conceptuels de mise en oeuvre.. Les Cahiers du numérique, 2015, 11 (4). hal-01292376

HAL Id: hal-01292376

<https://hal.science/hal-01292376>

Submitted on 22 Mar 2016

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

LE TOURNANT SEMIOTIQUE DANS LES ARCHIVES AUDIOVISUELLES¹.

Vision globale et éléments conceptuels de mise en œuvre

PETER STOCKINGER, STEFFEN LALANDE, ABDELKRIM BELOUED

1. Les archives numériques

La notion « archives numériques » revêt en effet deux problématiques différentes mais qui sont indissociablement liées à :

- la problématique des archives au sens d'une banque de données (ouvertes) ;
- la problématique des archives au sens d'une ressource ou d'un ensemble de ressources (d'un centre de ressources).

Les archives au sens d'une *banque de données* (ouvertes, d'accès libre, ...) font partie d'un *projet d'archivage* spécifique, sont composées d'une collection (d'une série de collections) de données numériques (composant le fonds – fermé ou ouvert - d'une archive) et peuvent servir *potentiellement* à toute sorte

¹ Cet article est issu de recherches menées actuellement dans le cadre du projet Campus AAR financé par l'Agence Nationale de la Recherche (ANR) en France (<http://www.agence-nationale-recherche.fr/>). Le projet Campus AAR est consacré à la conception et au développement d'une plateforme technique et scientifique ouverte pour la création, l'analyse, la diffusion et l'éditorialisation d'archives audiovisuelles en SHS (sciences humaines et sociales). Composé d'un consortium de 4 partenaires (Equipe Sémiotique Cognitive et nouveaux Médias (ESCoM) de la Fondation Maison des Sciences de l'Homme (FMSH) ; Institut National de l'Audiovisuel (INA), société ARMADILLO et Centre de Communication Scientifique Directe (CCSD) du CNRS), le projet a débuté en janvier 2015 et se terminera début 2017). Pour mieux connaître le projet et suivre son évolution, merci de consulter son site officiel : <http://campusaar.hypotheses.org/>.

d'exploitations pratiques. Un bon exemple d'archive au sens d'une banque de données ouvertes est l'archive pluridisciplinaire ouverte HAL² du CCSD³ qui est un entrepôt pérenne de documents de la recherche d'accès libre. Mais à côté de HAL qui jouit d'un statut officiel de référentiel national, on trouve toute une diversité de banques de données institutionnelles ou personnelles qui ont souvent une vocation patrimoniale et/ou pratique (i.e. de servir de ressource à telle ou telle activité sociale).

Entendu dans le deuxième sens, les archives existantes ou à venir servent d'objet de recherche théorique et appliquée en vue, par exemple, d'améliorer les *accès aux données archivées*, d'avancer sur le *terrain de l'interconnexion* des données archivées, de mieux contrôler l'évolution technologique des *supports*, de rendre possible la *préservation* de l'information à travers différents formats, d'améliorer les *conditions de stockage* des données et des métadonnées, de mieux connaître les *procédés* et les « façons de faire » qui déterminent les *pratiques courantes d'archivage*, etc.

Les archives au sens d'un *pôle de ressources épistémiques* mobilisables dans des contextes et pour des usages spécifiques les plus variés renvoient à la problématique centrale de la *réutilisation « active »* d'une donnée ou d'un ensemble de données archivées ce qui implique le plus souvent une *transformation qualitative* des données concernées (Stockinger 2015).

Comme déjà mentionné ci-dessus au sujet de la place des archives dans le dispositif de la communication globale d'une organisation, les données archivées peuvent constituer des ressources potentielles pour des activités spécifiques telles qu'une campagne concrète de communication, le lancement d'un produit, l'organisation d'un cycle de formation pour le personnel, etc. Bien sûr, un usage courant d'une archive est celui de la simple consultation de données archivées pour s'en inspirer, par exemple, ou pour trouver des nouvelles idées, pour s'en servir sous forme d'une « bonne pratique », etc. Mais, on peut aussi envisager d'utiliser ou, plutôt, de réutiliser un corpus de données archivées dans le cadre d'une des activités que nous venons de citer. Aussi dans le cadre des archives numériques de la recherche (exemple qui nous intéresse davantage ici), la réutilisation de données archivées (sous forme d'enregistrements audiovisuels, de prépublications, de banques d'images, ...) est à l'ordre de jour dans le cadre de l'enseignement supérieur ou de la veille scientifique consacrée, par exemple, à un problème sociétal, une région géopolitique ou un mouvement social (pour des exemples variés, cf. Stockinger et al. 2011). Mais pour pouvoir utiliser d'une manière

² Le site HAL: <https://hal.archives-ouvertes.fr/>

³ CCSD = Centre de Communication Scientifique Directe: <http://www.ccsd.cnrs.fr/>

appropriée les données archivées pour ces genres de projet, souvent un processus d'interprétation et d'adaptation de ces dernières est nécessaire.

Pour rendre un peu plus concret cette notion de transformation active d'une donnée archivée/d'un corpus de données archivées en une ressource *sui generis*, voici quelques activités tout à fait récurrentes qui montrent toute la complexité à la fois *conceptuelle*, *technique* et *pratique* de la mise en place d'une archive au sens d'un « centre » ou « pool » de ressources.

- La reclassification (de données archivées et déjà classées dans une archive) selon un plan de classification propre à l'utilisateur ou un groupe d'utilisateurs ;
- La redescription et réindexation (de données archivées déjà indexées) à l'aide de modèles de description appropriés aux besoins de l'utilisateur ;
- La recontextualisation (de données archivées et déjà mises en relation avec d'autres données) ;
- Le redécoupage virtuel (de données audiovisuelles, visuelles, textuelles ... archivées) pour ne retenir que les passages intéressant l'utilisateur ;
- La réécriture partielle (i.e. la modification même de la textualité d'origine d'une donnée archivée) sous forme, par exemple, d'un montage virtuel de la donnée archivée avec différents éléments textuels (segments audiovisuels, ...)
- La rééditorialisation et la republication d'une ou d'un corpus de données archivées.

Ces différentes activités ayant comme objectif de transformer une donnée archivée en une *ressource sui generis*, montrent toute la pertinence de la *vision sémiotique* (du « semiotic turn », cf. Stockinger 2015) pour la recherche sur les archives numériques : d'une manière générale, le fonds d'une archive peut être considéré comme un ensemble (quantitativement plus ou moins important) de *données textuelles* au sens large du terme qui à la fois documentent, conservent et transmettent le *discours* d'un *acteur social* (d'une institution, d'un groupe social, d'une communauté, voire d'une personne) sur un domaine, un objet ou encore une époque.⁴

⁴ Cette vision fait référence aux travaux antérieurs de Michel Foucault sur la fonction des archives dans la constitution d'une *rationalité* (par exemple *médicale* ou *pénale*, en ce qui concerne les recherches de Foucault), i.e. sur la constitution d'un ensemble de savoirs et savoir-faire formant un ensemble de standards (de traditions, normes ou règles) auxquels se réfère un corps social et/ou professionnel (médecins, psychiatres, pénalistes, ...) pour exercer son métier.

2. sémiotique des archives

2.1. Le cadre théorique

Afin de rendre opérationnel cette vision générale, il faut avoir recours à un cadre théorique du texte et du discours dont les principaux éléments sont représentés par la figure 1.

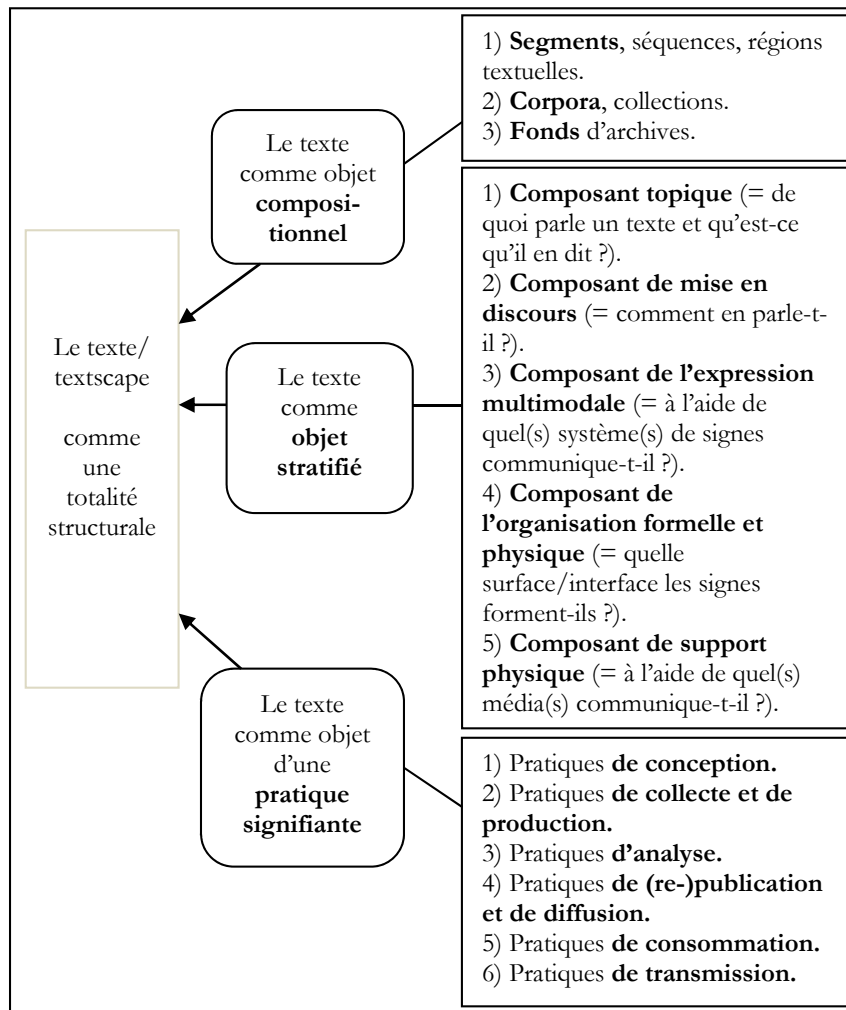


Figure 1. La vision sémiotique structurale sous-tendant le processus de modélisation conceptuelle des procédures d'analyse d'un objet/corpus d'objets média

Brièvement résumé, l'objet « texte » n'est pas du tout restreint à son avatar culturel du texte imprimé (du livre, du journal, ...) mais recouvre plutôt toute « entité signifiante » (selon A.J. Greimas et J. Courtés, 1979) pour un acteur, i.e. toute entité qui possède un certain sens pour un acteur (figurent ainsi parmi les « données textuelles » notamment les *artefacts*, les *objets naturels* qui possèdent un sens pour un acteur donné ainsi que les *gestes* qui expriment les activités de toutes sortes). Dans le cas des archives audiovisuelles, ces entités signifiantes sont restreintes à quelques grandes catégories : l'*écrit* et l'*imprimé*, le *visuel statique*, l'*audiovisuel* et le *sonore* (comprenant la parole, la musique et le bruit). Nous réunissons ces différentes catégories de « données textuelles » sous le terme générique de *média* ou encore d'*objet média*.

Cette vision sous-tend et motive l'ensemble des ressources métalinguistiques (terminologies, modèles de description) dont l'acronyme est *ASA* (*Atelier de Sémiotique Audiovisuelle*). Ces ressources métalinguistiques ont été d'abord conçues et testées dans le cadre d'un projet ANR du même nom⁵. Aujourd'hui, elles sont reprises dans le cadre du projet *Campus AAR*⁶ pour les conformer aux exigences formelles propres au sémantique web. L'objectif du projet Campus AAR est de fournir à la « communauté » académique des SHS (sciences humaines et sociales) un environnement intégré de production, analyse, éditorialisation et diffusion d'archives audiovisuelles. Cet environnement sera composé d'un outil appelé Studio Campus AAR qui comprendra un outil d'analyse de corpus d'objets média et un autre outil réservé à l'éditorialisation et la publication de corpus audiovisuels. Le « cœur » du Studio est composé de deux classes de modèles conceptuels :

- Le premier type recouvre les *modèles d'analyse* (ou scénarios d'analyse) qui représentent « l'univers du discours » d'une archive. L'analyste d'une archive dispose d'une série de formulaires dynamiques (intégré dans l'outil d'analyse) à l'aide desquels il décrit son corpus d'objets média.
- Le deuxième type recouvre les *modèles de publication* mis à la disposition de la personne/des personnes ayant en charge l'éditorialisation et la diffusion des objets média préalablement analysés.

⁵ Il s'agit du projet ASA-SHS de 2009 à 2012); cf. le blog du projet : <http://asashs.hypotheses.org/>

⁶ Le projet Campus AAR, coordonné par l'auteur de cet article, comprend un consortium de 4 partenaires français (Fondation Maison des Sciences de l'Homme, Institut National de l'Audiovisuel, CNRS – CCSd et société Armadillo) et a débuté en janvier 2014 pour une durée de 36 mois. Pour plus d'informations, cf. le blog scientifique du projet Campus AAR : <http://campusaar.hypotheses.org/>

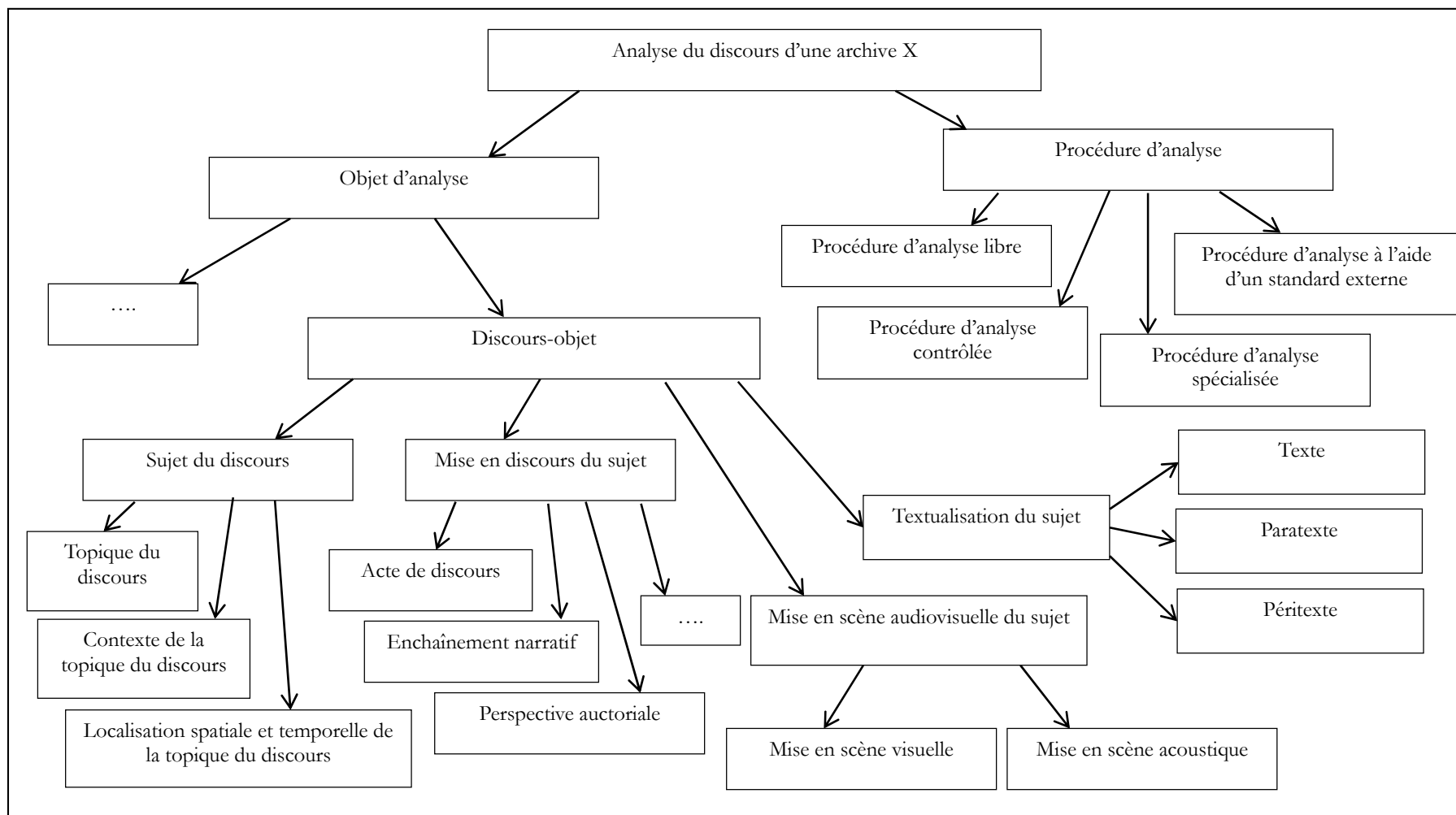


Figure 2. Vision générale de l'analyse du discours d'une archive X

Cet article ne traitera que des modèles d'analyse, i.e. du métalangage élaboré et utilisé par l'analyste pour décrire, classer, indexer, ... un objet ou un corpus d'objets média. Dans ce chapitre (chapitre 2), nous détaillerons la structure et les choix conceptuels qui sous-tendent l'élaboration de modèles d'analyse pour classer, décrire et indexer de corpus d'objets média qui font partie du fonds d'une archive. Dans le troisième chapitre, nous présenterons quelques exemples de modèles d'analyse ainsi que leur place dans le métalangage ASA. Lalande et Beloued montreront dans leur contribution (cf. leur article dans ce numéro) comment articuler et rendre opérationnels cette vision et ces modèles dans un cadre du sémantique web.

Conception et structure de ce métalangage se fonde sur la vision sémiotique du texte (figure 1). La figure 2 fournit une représentation de cette vision générale adaptée à l'univers du discours d'une archive et de son explicitation (de son analyse).

2.2. Le discours-objet

Dans la figure 2 nous voyons deux grandes branches – une branche appelée [Objet d'analyse] et une autre branche appelée [Procédure d'analyse]. A chaque objet d'analyse correspond une ou une famille de procédures d'analyse. La branche [Objet d'analyse] dans la figure 2 n'identifie qu'une seule sous-branche, celle qui nous intéresse ici, à savoir la sous-branche [Discours-objet].

[Discours-objet] veut dire que l'objet d'analyse d'un média (audiovisuel, visuel statique, sonore, ...) est le *discours audiovisuel* (i.e. le discours utilisant les modalités d'expression visuelle et/ou acoustique pour communiquer, faire partager un message, une information). En ce qui concerne donc l'objet d'analyse [Discours-objet], il est appréhendé, comme le montre la figure 2, à l'aide des quatre critères suivants :

- Le [Sujet du discours] : le *sujet du discours* correspond, grosso modo, à ce dont est question dans une ressource audiovisuelle.
- La [Mise en discours du sujet] : la *mise en discours du sujet* recouvre les différentes « stratégies » dont dispose l'auteur d'une ressource audiovisuelle pour traiter son sujet.
- La [Mise en scène audiovisuelle du sujet] : la *mise en scène audiovisuelle du sujet* recouvre les formes d'expression visuelle et acoustique (dont verbale) pour communiquer le sujet.
- La [Textualisation du sujet] : la *textualisation du sujet* recouvre les possibilités de manifestation physique d'un sujet et sa communication dans un média (vidéo, photo, ...), dans une partie d'un média (exemple : un segment

vidéo, une région photo, un chapitre de brochure, ...), dans plusieurs médias, etc.

Ces quatre critères font partie de toute une tradition de recherche en sciences du langage et, plus particulièrement, en sémiotique (du discours). Dans le cadre du projet Campus AAR, un ensemble de choix méthodologiques ont dû être faits afin de pouvoir utiliser concrètement ces connaissances.

Ces choix entraînent inévitablement des simplifications par rapport à l'état des connaissances scientifiques relatives à l'organisation et au fonctionnement du discours mais *ils permettent de dépasser de loin les pratiques habituelles* de description et d'indexation de ressources audiovisuelles dans le cadre d'un projet d'archive.

Le sujet du discours

Le « sujet du discours », comme le montre la figure 2, se compose lui-même de trois facettes ou parties :

- La [Topique du discours], i.e. ce dont on parle dans un discours (c'est le domaine ou l'objet du discours).
- Le [Contexte de la topique], i.e. la prise en compte éventuelle du cadre (historique, social, institutionnel, ...) permettant une meilleure compréhension de l'objet traité dans le discours.
- La [Localisation spatiale] et la [Localisation temporelle] de la topique du discours, i.e. la prise en compte éventuelle des lieux (géographiques, géopolitiques, ...) et des moments ou périodes (historiques) pertinents pour localiser l'objet ou le domaine dont est question dans le discours véhiculé par un média.

Dans le cadre d'une analyse concrète d'un objet média, *seule celle de la topique du discours est obligatoire* ! L'explicitation du contexte et l'identification des coordonnées géographiques et temporelles ajoutent, bien sûr, à la richesse d'une analyse mais elles restent facultatives.

La mise en discours du sujet

La [Mise en discours du sujet] constitue le deuxième critère cité ci-dessus pour procéder à l'analyse de l'objet « discours (audiovisuel) ». L'auteur (l'énonciateur) d'un discours à toutes les latitudes de traiter « à sa façon » le sujet de son discours, i.e. ce dont il « parle ». La recherche a identifié toute une série de « stratégies » dont peut se servir l'auteur (l'énonciateur) pour l'imprimer de son point de vue, de sa culture, de ses intérêts, ... le discours qu'il produit

sur un domaine, un objet particulier. Dans le cadre du projet Campus AAR, nous avons retenu plus particulièrement *trois stratégies* tout en étant conscient qu'il s'agit ici d'un petit nombre de stratégies discursives parmi une diversité bien plus importante. Ce choix est motivé par la spécificité des corpus traités dans le cadre du projet Campus AAR (i.e. en grande partie des ressources audiovisuelles académiques) :

- [Acte de discours] : le sujet d'un discours peut être traité et développé sous des formes rhétoriques les plus diverses : définition, exposé, narration, argumentation, débat (avec un contradicteur thématique ou simplement présumé), etc.
- [Enchaînement syntagmatique (ou narratif)] : le développement d'un sujet peut se faire en différentes étapes et selon des « logiques » spécifiques. Il existe ainsi un enchaînement du type « du plus simple au plus complexe » ; un autre enchaînement très typique suit un simple ordre chronologique ; un troisième enchaînement met en avant les choses les plus importantes à retenir suivies des détails plus ou moins importants.
- [Perspective auctoriale] : le sujet d'un discours peut être développé exclusivement du point de vue de l'auteur (énonciateur) du discours. Mais l'auteur peut aussi faire appel à d'autres points de vue, à des points de vue bien identifiés ou uniquement présumés, à des points de vue personnels (i.e. mis en avant par une personne seulement), de groupe (i.e. mis en avant par une collectivité) ou par « tout le monde ».

Dans le cadre d'une analyse-indexation concrète d'un objet média, l'analyste est libre de renseigner ces trois facettes, *la seule facette obligatoirement à renseigner étant celle de la topique du discours.*

La mise en scène audiovisuelle du sujet

Le troisième critère de l'analyse de l'objet « discours (audiovisuel) » est celui de la [Mise en scène audiovisuelle du sujet]. Il s'agit ici de renseigner éventuellement certaines caractéristiques significatives dans l'expression du sujet à l'aide de la parole, de la musique et de l'image (animée). Ce critère est peut-être moins important dans l'analyse du *discours filmé* (par exemple, dans l'analyse d'un entretien filmé, d'une conférence filmée) mais il devient central dans l'analyse d'événements, d'objets, de personnes, ... filmés comme c'est le cas, typiquement, des enregistrements audiovisuels de la recherche sur terrain, des enregistrements de patrimoines culturels (monuments, œuvres, instruments, ...) ou encore des reportages et documentaires. Ce critère recouvre les deux facettes suivantes :

- [Mise en scène acoustique] : Il s’agit de décrire manuellement certaines propriétés de la parole (d’une personne interviewée, d’un orateur, ...), du bruit ambiant ou encore, si pertinent, de la musique).
- [Mise en scène visuelle] : Il s’agit ici de décrire éventuellement (i.e. si pertinent) les formes plastiques de la mise en scène des représentations iconiques de l’objet du discours (cadrage, point de vue visuel, mouvement de caméra, etc.)

La textualisation du sujet

Le quatrième critère de l’analyse de l’objet « discours (audiovisuel) » est celui de la textualisation du sujet du discours. Nous le traitons à l’aide de trois facettes dont les deux premiers sont obligatoires pour toute analyse :

- Le [Texte] : Cette facette permet d’identifier le média (ou la partie du média) sur lequel le sujet du discours est enregistré. Par exemple, dans le cadre de l’analyse d’une vidéo, il s’agit d’identifier le ou les *segments* dans lequel/lesquels on parle d’un sujet donné.
- Le [Paratexte] : Cette facette réunit tous les éléments nécessaires à identifier un texte (vidéo, photo, segment de vidéo, région de photo, ...). Elle intègre, entre autre, les éléments exigés à renseigner par le standard *Dublin Core*.
- Le [Péri-texte] : Cette facette (facultative) permet à l’analyste de renseigner éventuellement sur la « vie » du sujet en dehors de son texte d’origine. Concrètement, cela veut dire que le sujet traité dans un entretien peut être traité par son auteur à d’autres occasions, dans d’autres publications (lato sensu) ; le même sujet peut également être traité par d’autres auteurs ou encore le sujet tel qu’il est traité par son auteur peut lui-même être l’objet de commentaires, de critiques, etc. Toutes ces données constituent le « péri-texte » d’un sujet.

2.3. Les procédures d’analyse

La figure 3 montre un extrait de l’ontologie ASA consacrée à l’identification et définition des principales catégories de procédures d’analyse dans le cadre du projet Campus AAR :

- [Procédure d’analyse libre de l’objet média] : l’analyste saisit librement à l’aide de formulaires standards et communs à tous les projets d’archives les expressions verbales pour indexer une ressource audiovisuelle ou une partie d’une ressource audiovisuelle. Nous distinguons, à l’intérieur

de cette catégorie, différents types d'analyse dont le plus important est l'[Annotation de base]. L'[Annotation de base] est la procédure la plus simple qui demande à l'analyste de renseigner sous forme d'un texte librement saisi, d'un intitulé ou d'un syntagme (en général nominal) la thématique – la topique - du discours (i.e. ce dont est question dans une vidéo ou un segment particulier d'une vidéo).

- [Procédure d'analyse contrôlée à l'aide du thésaurus ASA] : l'analyste saisit, dans le thésaurus ASA la ou les valeurs appropriées pour indexer une ressource audiovisuelle ou une partie d'une ressource audiovisuelle.
- [Annotation spécialisée] : l'analyste saisit (d'une manière libre et/ou contrôlée) à l'aide d'un formulaire « ad hoc » et adapté soit à des besoins spécifiques d'un projet d'archive soit aux contraintes particulières d'un objet d'analyse, les expressions verbales (ou non-verbales) pour indexer une ressource audiovisuelle ou une partie d'un ressource audiovisuelle.

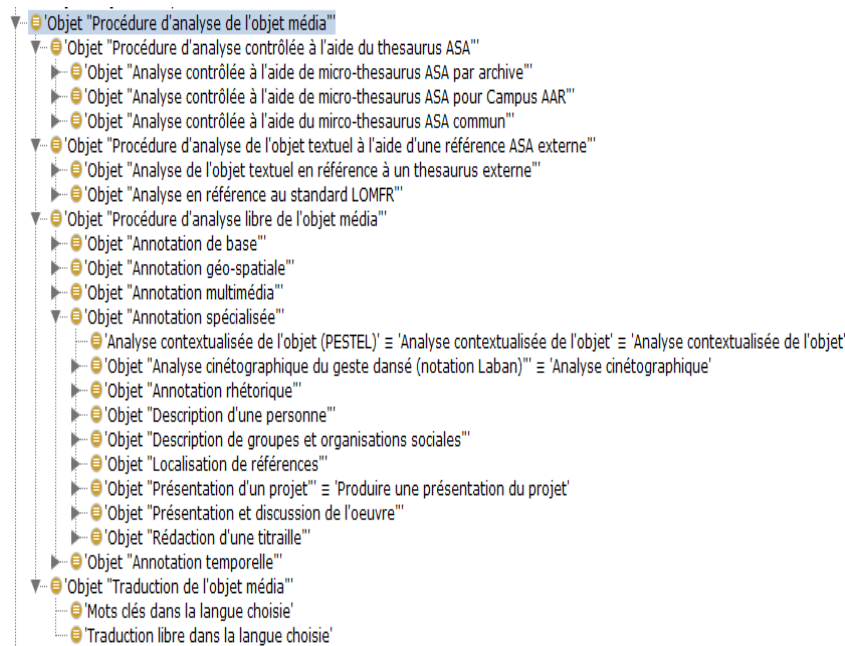


Figure 3. Ontologie ASA des procédures d'analyse sous Stanford Protégé⁷

- [Procédure d'analyse de l'objet média à l'aide d'une référence ASA externe] : l'analyste saisit à l'aide de formulaires de description configurés conformément aux spécificités d'une ressource métalinguistique extérieure au métalangage ASA (exemple : formulaires

⁷ Site de l'éditeur Protégé : <http://protege.stanford.edu/>

de description configurés conformément au standard LOM⁸ ; formulaires de description configurés conformément à un ensemble de notices d'autorité RAMEAU⁹ ; formulaires de description configurés conformément à un micro-thésaurus UNESCO¹⁰, etc.).

2.4. L'analyse du sujet du discours

Voyons maintenant plus en détail le premier critère de l'analyse du [Discours-objet], à savoir l'analyse du [Sujet du discours] (cf. figure 2). Comme déjà dit, celui-ci est composé des trois facettes suivantes :

- analyse de la [Topique du discours] (cette facette est obligatoire pour toute analyse) ;
- analyse du [Contexte de la topique du discours] (cette facette est facultative) ;
- analyse de la [Localisation spatiale] et de la [Localisation temporelle] de la topique du discours (cette facette est également facultative).

Les modèles sous-tendant les facettes [Contexte de la topique du discours], [Localisation spatiale] et [Localisation temporelle] restent, dans le cadre du projet Campus AAR, relativement stables. Cela veut dire qu'ils sont employés quasiment tels quels dans tous les projets d'analyse de ressources audiovisuelles, peu importe l'archive ou les archives concerné(e)s. Autrement dit, on trouvera les modèles qui sous-tendent ces deux facettes aussi bien dans le cadre de l'analyse d'objets média pour une archive en SHS (exemple : l'archive AGORA réunissant le patrimoine SHS de la Fondation Maison des Sciences de l'Homme à Paris¹¹) que dans le cadre de l'analyse d'objet média pour une archive en histoire des mathématiques (exemple : l'archive AHM¹² réunissant le patrimoine d'un séminaire de recherche dirigé par Dominique Flament, historien des mathématiques au CNRS).

Par contre, la facette réservée à l'analyse de la [Topique du discours] repose sur une très grande variété de modèles correspondant aux *spécificités référentielles* de chaque archive. Pour chaque archive (pour chaque projet d'archives), nous

⁸ LOM = Learning Object Metadata : <http://www.lom-fr.fr/>

⁹ RAMEAU = Répertoire d'autorité-matière encyclopédique et alphabétique unifié : <http://rameau.bnf.fr/>

¹⁰ Thésaurus de l'UNESCO (en format SKOS) : <http://skos.um.es/unescothes/?l=fr>

¹¹ Site du portail AGORA : <http://www.agora.msh-paris.fr/>

¹² Site du portail AHM : <http://www.ahm.msh-paris.fr/>

devons donc construire des modèles d'analyse appropriés à la ou plutôt aux topiques propres à une archive.

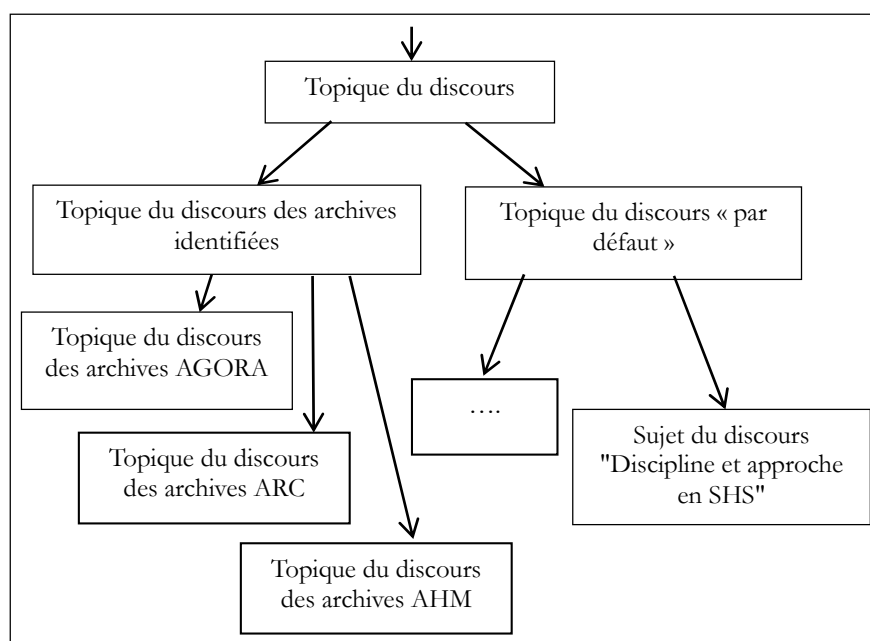


Figure 4. Univers des topiques du discours des archives Campus AAR

La figure 4 nous montre une représentation arborescente de l'univers des topiques du discours d'un ensemble d'archives qui font partie du projet Campus AAR. Sous la branche [Topique du discours d'archives identifiées] sont réunies les topiques composant les domaines du discours de chaque archive (projet d'archives) conçue et élaborée à l'aide du métalangage ASA. Dans le cadre du projet Campus AAR, il s'agit des topiques du discours des archives actuellement en cours de développement :

- AGORA¹³ dédié au patrimoine audiovisuel des sciences humaines et sociales de la Fondation Maison des Sciences de l'Homme à Paris ;
- ARC¹⁴ (A la Rencontre des Cultures) dédié à la diversité culturelle et réalisé avec l'Institut National des Langues et Civilisations Orientales (INALCO) à Paris ;

¹³ Portail AGORA : <http://www.agora.msh-paris.fr/>

¹⁴ Portail ARC : <http://www.arc.msh-paris.fr/>

- AHM¹⁵ (Archives Histoire des Mathématiques) préservant et valorisant le patrimoine scientifique réalisé dans le cadre d'un séminaire de recherche pluriannuel en histoire des mathématiques et dirigé par Dominique Flament du CNRS.

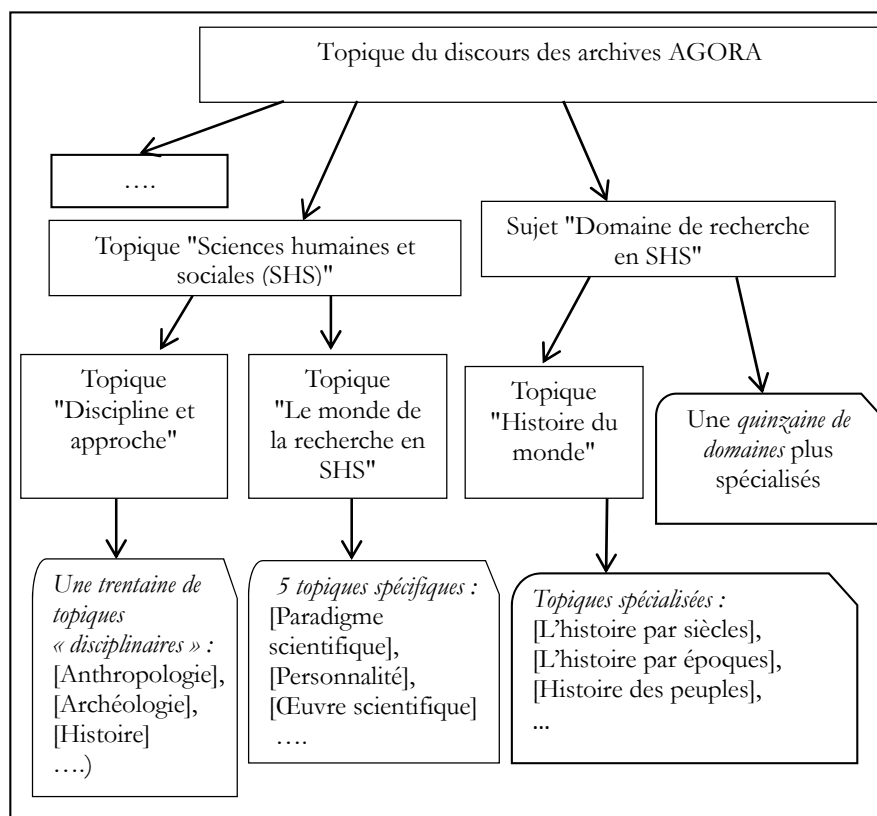


Figure 5. Vue générale des topiques composant l'univers du discours AGORA

Considérons rapidement l'organisation des topiques composant le discours de l'archive AGORA. Comme le montre la figure 5, celle-ci se différencie entre deux grandes classes de topiques :

- Une première classe consacrée plutôt à la *recherche elle-même* : disciplines scientifiques, approches, paradigmes scientifiques, personnalités, publications, projets de recherche, etc.

¹⁵ Portail AHM : <http://www.ahm.msh-paris.fr/>

- Une deuxième classe consacrée plutôt aux *domaines de connaissance*, aux objets de recherche en SHS : l'histoire, le monde contemporain, la société, les cultures, les langues, etc.

Cette distinction entre deux grandes classes de topiques repose en effet sur un choix en amont d'organiser l'accès au fonds audiovisuel de l'archive en question soit d'une manière classique et standard via une série de *disciplines* ; soit d'une manière beaucoup plus thématique mais aussi plus déconcertante pour un visiteur « novice » sous forme d'ensembles de domaines de connaissance qui organisent les sciences humaines et sociales. Prenons le cas d'un entretien consacré à la première guerre mondiale. L'analyse de celui-ci peut se focaliser une fois plutôt sur l'événement historique lui-même et enrichira ainsi une des topiques spécialisées de l'Histoire du monde (cf. figure 5) ; une autre fois, elle peut porter sur l'approche disciplinaire ou scientifique enrichissant ainsi la topique disciplinaire « La recherche historique par spécialité » (cf. figure 5).

Malgré la très grande diversité possible de modèles de description de la topique du discours audiovisuel, on y voit se dégager progressivement de *classes génériques de modèles* pouvant être réutilisés tels quels ou moyennant certaines modifications locales dans différentes archives, i.e. appartenant à différents univers de sens. Ces classes génériques sont réunies sous la branche [Topique du discours *par défaut*]. Dans notre cas précis – et étant donné l'objectif principal du projet Campus AAR – il s'agit plus particulièrement d'une *bibliothèque par défaut* couvrant les *disciplines* en SHS.

3. Les modèles d'analyse de l'environnement Campus AAR

3.1. Introduction et exemple

Les objets média composant le fonds d'une archive audiovisuelle sont analysés – décrits et indexés – à l'aide d'une bibliothèque de modèles d'analyse appelés aussi *scénarios*. L'analyste accède, utilise ces modèles sous forme de formulaires interactifs qui forment la partie la plus importante de l'interface de l'outil d'analyse du Studio Campus AAR.

La figure 6 montre une interface qui permet à l'analyste d'utiliser un scénario spécifique à l'aide duquel il peut identifier les segments d'une vidéo qui sont pertinents pour son analyse. La figure 7 montre l'accès aux scénarios utilisés pour analyser soit *une vidéo dans son intégralité* soit *tel ou tel segment spécifique*. Voici une liste très utilisée de scénarios qu'offre le Studio Campus AAR pour analyser et décrire une vidéo :

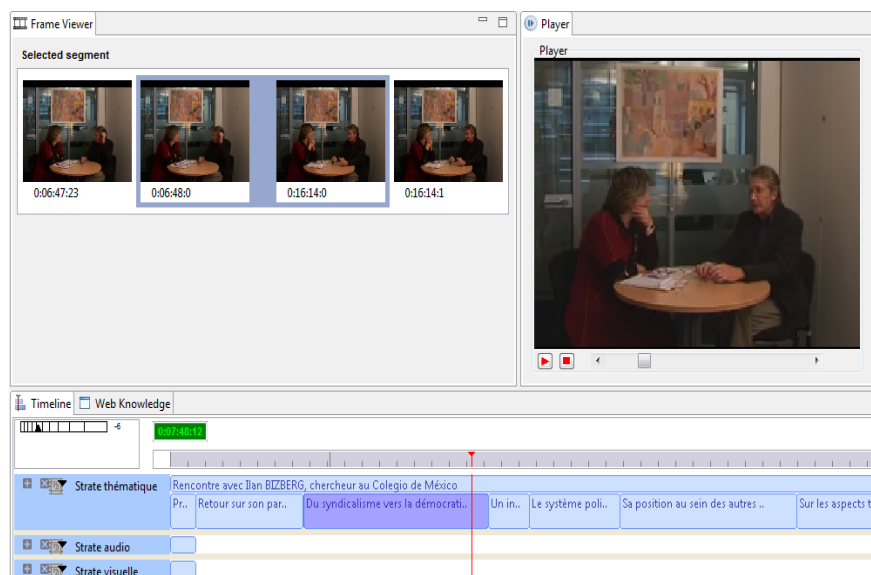


Figure 6. Interface donnant accès au modèle d'identification et de segmentation – thématique, visuelle, sonore, ... - d'une vidéo)

- Informations signalétiques d'une vidéo.
- Mentions légales.
- Classification de la vidéo à l'aide d'un thesaurus propre à une archive ou faisant partie d'un langage tel que RAMEAU.
- Identification et présentation des différentes personnes ou groupes de personnes intervenant autour ou dans une vidéo.
- Description du plan visuel d'une vidéo.
- Description du plan sonore d'une vidéo.
- Description du contenu à proprement parler d'une vidéo.
- Description du genre discursif ou rhétorique d'une vidéo (ou d'un segment d'une vidéo).
- Présentation des usages possibles d'une vidéo.
- Enrichissement d'une vidéo sous forme de sa mise en relation avec d'autres ressources.
- Traduction du contenu d'une vidéo.

Tous les projets de production, d'analyse et d'éditorialisation-diffusion d'archives partagent cette bibliothèque. Certains scénarios restent plus ou

moins *invariables*, d'autres subissent des modifications plus ou moins importantes d'une archive à une autre. Parmi les scénarios qui ne subissent que peu de variations, on trouve plus particulièrement le scénario pour produire une *méta-description*, i.e. une présentation de l'analyse en elle-même. Mais c'est aussi le cas pour les scénarios d'analyse textuelle (de segmentation d'une vidéo sur une ou plusieurs lignes temporelles ou en une ou plusieurs zones visuelles de l'espace 2D d'une image), de description du plan visuel et acoustique, des mentions relatives à un objet média, d'informations signalétiques (auteurs, responsabilité, édition, ...) ou encore de description des usages d'un objet média.

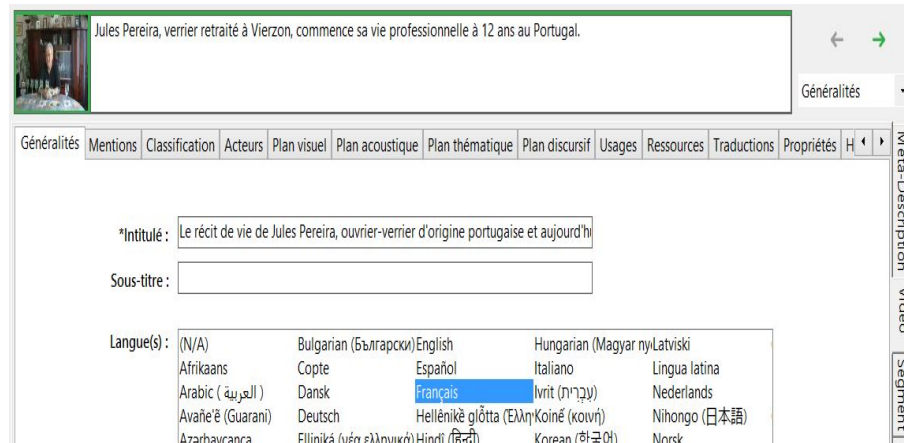


Figure 7. Interface donnant accès à une bibliothèque de modèles d'analyse servant à l'analyse d'un objet média/ d'un corpus d'objets média

En d'autres termes une grande partie des scénarios d'analyse peut être réutilisée quasiment telle quelle dans tout projet concret. La grande exception est constituée ici par la classe des scénarios d'analyse du contenu à proprement parler. Ces scénarios, en effet, varient ou – plutôt – peuvent varier énormément d'une archive à une autre.

3.2. La structure de base d'un modèle d'analyse

D'une manière élémentaire, un scénario d'analyse n'est rien d'autre que la réunion d'un **objet d'analyse** et d'une **procédure d'analyse** (cf. figure 8).

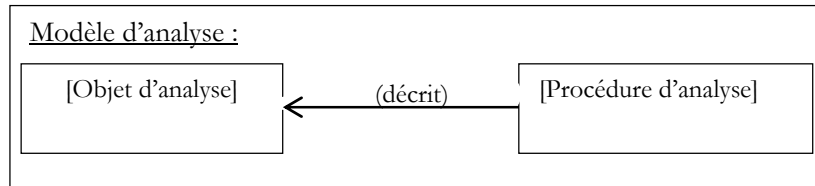


Figure 8. Représentation graphique de la structure d'un modèle d'analyse

[Objet d'analyse] et [Procédure d'analyse] sont *interprétés* (spécifiés) par des *termes conceptuels* (concepts) qui composent le vocabulaire (ou *lexique*) conceptuel (appelé *méta-lexique ASA*) :

- « méta-lexique » pour le distinguer des lexiques des langues naturelles ;
- « ASA » en référence au projet « Atelier de Sémiotique Audiovisuelle¹⁶ » ayant permis une première élaboration du méta-lexique en question).

Le *méta-lexique ASA* forme l'*ontologie de base - l'ontologie générique* - de tous les projets de production, d'analyse et d'éditorialisation-diffusion d'archives réalisés à l'aide de l'environnement Campus AAR (cf. Stockinger 2012, pour une présentation plus détaillée). Les objets d'analyse de l'univers du discours Campus AAR forment ainsi les référents d'une des deux branches du méta-lexique ASA – branche dont le terme de base s'intitule [Objet d'analyse] ; les procédures d'analyse reconnues dans l'univers du discours Campus AAR, forment la deuxième branche intitulée [Procédure d'analyse] (cf. figure 9).

Dans les cas les plus simples, l'objet d'analyse est représenté par *un seul* terme conceptuel. Par exemple, le terme conceptuel [Pays] peut suffire pour identifier tel ou tel pays concret dont est question dans une vidéo, un texte ou une image. Dans des cas plus complexes, cependant, on aura recours à plusieurs termes conceptuels pour décrire un objet d'analyse. Par exemple, pour exprimer le fait que dans un enregistrement audiovisuel d'une conférence est question de la *traite d'esclaves au 18^{ème} siècle entre l'Afrique subsaharienne et les Amériques*¹⁷, on aura recours aux termes conceptuels suivants dont chacun représente (exprime) un aspect particulier de l'objet d'analyse :

¹⁶ <http://asashs.hypotheses.org/>

¹⁷ Il s'agit d'une conférence d'Ibrahima Thioub sur la [traite atlantique des esclaves](#) dans le cadre du CEM (Collège d'Etudes Mondiales) à la Fondation Maison des Sciences de l'Homme à Paris (réalisée le 8 novembre 2012)

- [Pratique d’esclavage] pour décrire la traite d’esclaves ;
- [Région géopolitique] pour décrire l’Afrique subsaharienne et les Amériques ;
- [Période] ou [Siècle] pour décrire le 18^{ème} siècle.

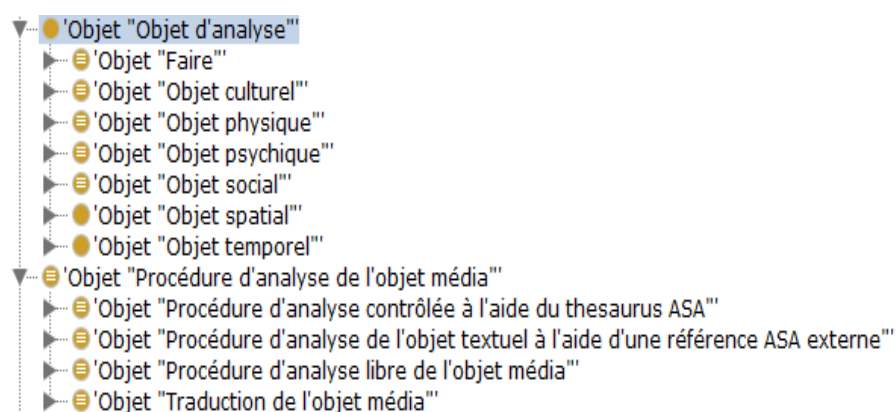


Figure 9. Extrait du métalexique ASA : la base canonique des termes conceptuels des deux branches [Objet d'analyse] et [Procédure d'analyse] sous Stanford Protégé)

On aura également recours à des *relations* entre ces termes pour exprimer notamment le fait que le terme [Région géopolitique] exprime une fois une *région* « d'origine » (de la pratique en question) et une autre fois une *région* « bénéficiaire » de la même pratique.

Du côté de la procédure d'analyse, on précisera la ou les procédures (faisant partie des quatre grandes catégories) à l'aide de laquelle ou desquelles on traitera l'objet d'analyse. Du côté de la procédure d'analyse libre, l'activité *la plus élémentaire* est appelée [Désignation minimale] et consiste en la saisie obligatoire d'une *expression dite minimale*. Par [Expression minimale], on entend une expression dont l'analyste juge lui-même qu'elle soit suffisamment claire (« évidente ») pour être comprise *et* distinctive pour le public qu'il vise. En d'autres termes, une « expression minimale » peut être un simple mot, un syntagme nominal, mais aussi une expression linguistique plus élaborée – tout dépend du seuil de la pertinence dont seul l'analyste est juge. Bien entendu, à côté de la procédure élémentaire de description libre se réduisant à la seule saisie d'une expression minimale, il existe toute une série de procédures de description libre plus élaborée, plus complexes.

3.3. Modèles d'analyse à structure interne plus complexe

Ceci dit, la structure interne d'un modèle d'analyse peut se présenter d'une manière plus élaborée que celle représentée par la figure 9.

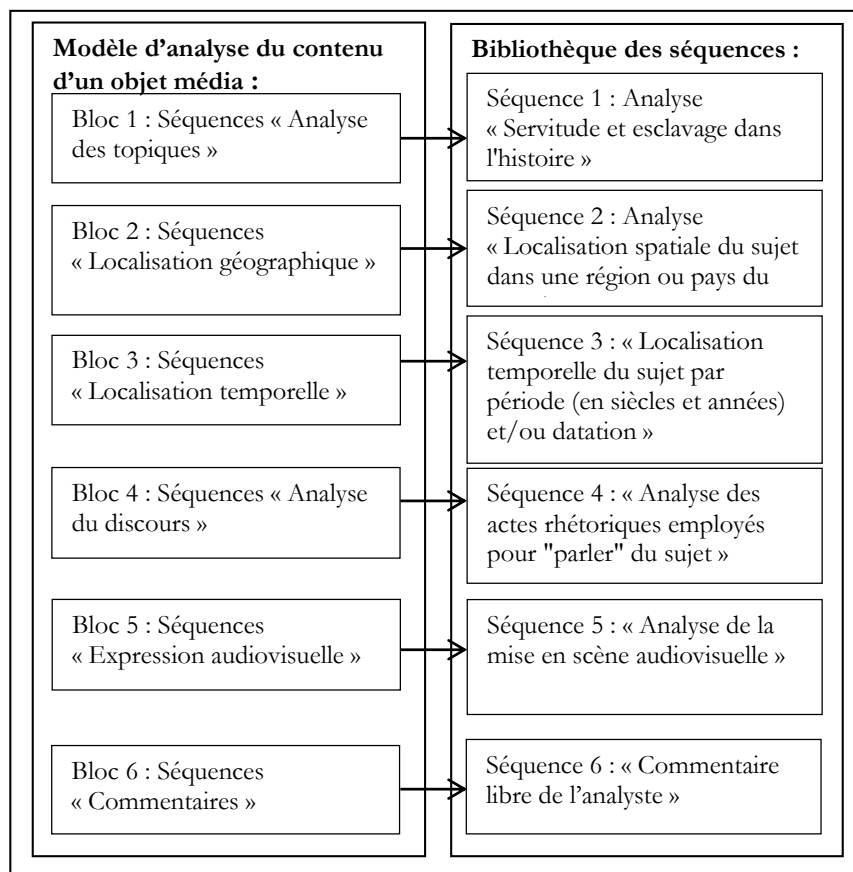


Figure 10. Bibliothèque de séquences de description organisées en plusieurs grands blocs analytiques à partir desquels sont spécifiés les scénarios ou modèles d'analyse (ici : exemple d'un scénario utilisé pour analyser des vidéos traitant la problématique de la servitude et de l'esclavage dans l'histoire ...)

C'est notamment le cas pour l'analyse du contenu d'un objet média, i.e. l'analyse du ou des sujets développés dans, par exemple, une vidéo ou un corpus de vidéos. Pour reprendre l'exemple de la traite d'esclaves au 18^{ème} siècle entre l'Afrique subsaharienne et les Amériques. Pour pouvoir rendre compte d'une

manière plus exhaustive du sujet traité par Ibrahima Thioub, il est nécessaire d'avoir recours à une série de *blocs analytiques* qui nous permettent de spécifier un scénario d'analyse non seulement du contenu de la conférence en question (i.e. de la conférence d'Ibrahima Thioub) mais aussi de toutes les autres conférences qui traitent de sujets similaires (cf. figure 10). Les trois premiers blocs analytiques dont nous avons besoin, sont :

- L'objet dont parle Ibrahima Thioub (i.e. la traite d'esclave)
- La localisation géographique de l'objet (i.e. l'Afrique et les Amériques)
- La localisation temporelle de l'objet (i.e. la période du 18^{ème} siècle, notamment)

The image shows a web-based form for describing video content. The main title is "Sujet "Servage et esclavage dans l'histoire"". The form is organized into two columns.

Left Column (Navigation/Tree):

- ▾ Sujet "Servage et esclavage dans l'histoire"
 - ▷ 1) Description de la topique "Pratique de servage/d'esclavage" (PROP)
 - ▷ 1) Identification et description du/des actes-directeur (PROP)
 - ▷ Localisation du sujet par période (PROP)
 - ▷ Localisation du sujet par siècle (PROP)
 - ▷ Localisation du sujet dans un/plusieurs pays du monde (PROP)
 - ▷ Localisation du sujet dans une/plusieurs régions du monde (PROP)
- ▾ Sujet "Servage et esclavage dans l'histoire"
 - ▾ 3) Description de la topique "Thème de recherche" (PROP)
 - Thème de recherche
 - ▾ 1) Description de la topique "Pratique de servage/d'esclavage" (F)
 - ▾ Pratique d'esclavage
 - ▾ Rédiger une annotation descriptive (PROP)
 - Annotation simplifiée
 - Annotation par mots clés
 - Pratique de servage
 - ▾ 2) Description de la topique "Statut de serf/d'esclave" (PROP)
 - Serf (statut de)
 - Esclavage (statut d')
 - ▾ 1) Identification et description du/des actes-directeur (PROP)
 - Bilan historique
 - Exposé historique
 - Critique
 - ▾ Localisation du sujet par période (PROP)
 - ▾ Période temporelle à spécifier
 - ▾ Produire des informations temporelles (PROP)
 - Saisir le nom de la période
 - ▾ Localisation du sujet par siècle (PROP)
 - 04 - XVIIIe siècle ap. J.-C.
 - 05 - XVIIIe siècle ap. J.-C.
 - 03 - XIXe siècle ap. J.-C.
 - 02 - XXe siècle ap. J.-C.

Right Column (Form Fields):

Sujet "Servage et esclavage dans l'histoire" :

- 3) Description de la topique "Thème de recherche" (PROP) :
 - Thème de recherche
- 1) Description de la topique "Pratique de servage/d'esclavage" (PROP); Pratique de servage(*) :
 - Pratique de servage
- 1) Description de la topique "Pratique de servage/d'esclavage" (PROP); Pratique d'esclavage(*) :
 - Pratique d'esclavage
- 2) Description de la topique "Statut de serf/d'esclave" (PROP); Serf (statut de) :
 - Serf (statut de)
- 2) Description de la topique "Statut de serf/d'esclave" (PROP); Esclave (statut d') :
 - Esclavage (statut d')

Figure 11. Exemple d'un formulaire de description de vidéos traitant des sujets relatifs à la servitude et à l'esclavage dans l'histoire

Or, ces trois **blocs** composent en quelque sorte une *grille générale d'analyse* des *topiques* abordées dans un objet média. Certains sujets sont « insensibles » à une localisation spatiale ou temporelle, d'autres sujets demandent des formes de localisation plus spécialisées (par exemple, des localisations dans un contexte

social, institutionnel, linguistique, etc.). Toujours est-il qu'un scénario d'analyse du contenu (des *sujets*) d'un objet média doit offrir à l'analyste la possibilité de produire, s'il le souhaite, une *description structurée* d'un sujet prenant en compte d'une part l'*objet* lui-même et d'autre part sa ou ses *localisations* dans un contexte (géographique, historique, ...).

Par ailleurs, les *trois blocs analytiques* identifiés ci-dessus ne permettent pas réellement d'exprimer bien d'autres paramètres auxquels pourraient s'intéresser l'analyste. Pour reprendre l'exemple de la conférence d'Ibrahima Thioub, l'analyste pourrait souhaiter expliciter davantage la « mise en discours » du sujet de la traite d'esclaves dans la conférence en question : le ou les thèmes plus spécialisés abordés par le conférencier ; le ou les points de vue selon le(s)quel(s) il traite ce sujet ; le développement discursif sous forme, par exemple, d'un exposé historique ou d'une série d'exemples, d'une série de comparaisons, etc.

L'identification précise de ces différents paramètres pouvant intéresser l'analyste, constituent ensemble un *quatrième bloc analytique* (cf. figure 10), qui se rajoutent au trois premiers et dont la spécificité est celle de permettre à l'analyste de procéder (s'il le souhaite, bien entendu) à une *expertise discursive et rhétorique* d'un sujet abordé et développé dans un objet média.

D'autres blocs analytiques peuvent s'avérer nécessaires pour pouvoir accompagner l'analyste dans son travail de description. Ainsi par exemple, l'analyste peut souhaiter introduire des commentaires externes (les siens ou provenant d'autres sources) à propos du sujet traité dans un objet média (nous appelons ce *cinquième bloc analytique* « Commentaires »). Enfin, un *sixième bloc analytique* concerne moins l'analyse de la mise en discours d'un sujet traité dans un objet média mais plutôt l'analyse de sa mise en scène audiovisuelle.

La figure 11 nous montre, enfin, d'une manière schématique la composition d'un scénario d'analyse du contenu par un ensemble de *séquences* de description qui elles-mêmes font partie d'une bibliothèque de *blocs descriptifs*. L'exemple est *une partie* du contenu de la conférence d'Ibrahima Thioub, i.e. la traite d'esclaves au 18^{ème} siècle.

Conclusion

Retenons qu'un modèle d'analyse (que nous appelons également *scénario*) se présente sous forme d'un *texte* (i.e. plus précisément sous forme d'un *méta-texte*) qui se compose d'une ou de plusieurs parties appelées *séquences* (*sous-séquences*, ...). Chaque séquence est à son tour caractérisée par la présence d'un objet

d'analyse et d'une procédure d'analyse provenant du métalexique ASA (figure 9) qui constitue, comme déjà dit, l'ontologie générique que se partagent toutes les archives réalisées et gérées avec l'environnement Campus AAR.

Ces constructions métalinguistiques sont actuellement testées et affinées sur un ensemble de projets pilote dont notamment les trois projets cités AGORA, ARC et AHM qui sont publiquement accessibles sur le web.

Pour terminer, soulignons qu'un des objectifs scientifiques et techniques du projet Campus AAR est faire converger en un *tout cohérent* 1) la vision sémiotique des archives, 2) des *modèles d'analyse* d'objets média très sophistiqués et adaptés aux besoins les plus variés, 3) la *formalisation* de ces modèles sous forme de *graphes conceptuels* et dans un *format rdf/owl 2*.

Un autre objectif – technique – du même projet est de mettre en place un environnement de travail sous forme d'outils d'analyse et d'éditorialisation de corpus audiovisuels qui sont exclusivement basés sur ce genre de modèles exprimés en format rdf/owl 2 et intégrant une vision sémiotique des archives audiovisuelles. Enfin, un dernier objectif du projet est de montrer la viabilité et la pertinence d'une telle *approche innovatrice* à la fois au niveau théorique qu'au niveau technique sous forme de projets pilote concrets (nous en avons cité les trois principaux : AGORA, ARC, AHM).

Même s'il reste encore toute une série de problèmes à résoudre, même si les résultats de ce travail collectif restent toujours perfectibles, il nous semble que ceux réalisés jusqu'au mi-parcours du projet Campus AAR¹⁸ nous permettent d'être – en les comparant plus particulièrement à l'état de l'art relatives aux travaux et diverses « solutions » techniques pour mieux instrumenter et accompagner les pratiques de l'analyse d'archives audiovisuelles – fiers et optimistes pour l'avenir.

Bibliographie

Craven L. (éd.) (2008). *What are Archives. Cultural and Theoretical Perspectives*. A Reader, Ashgate.

Greimas A.J., Courtes J. (1979). *Sémiotique. Dictionnaire raisonné de la théorie du langage*, Hachette, 1979.

¹⁸ Précisions que le projet Campus AAR a démarré en janvier 2014 et se terminera début 2017 ...

Greimas A.J. (1966). *Sémantique structurale. Recherche de méthode.* Larousse.

Sowa J. (1984). *Conceptual Structures: Information Processing in Mind and Machine.* Addison-Wesley 1984

Stockinger P. (2015). The semiotic turn in digital archives and libraries; in :. LCN - Les cahiers du numérique, N° 1 2015 (DOI:10.3166/LCN.11.1.55-80)

Stockinger P. (2012). *Analyse des contenus audiovisuels. Métalangage et modèles de description* ; Paris/Londres, Hermes Science Publishing 2012 (350 pages) – traduction en anglais chez J. Wiley & Sons (NY, 2012)

Stockinger P. (2011) (éd.). *Nouveaux usages des archives audiovisuelles numériques.* Paris/Londres, Hermes Science Publishing (trad. en anglais aux éditions John Wiley & Sons, NY).