

H a n d o u t

Peter Stockinger, P.U.

**La place de l'hypertextualité dans le traitement
de corpus audiovisuels et/ou multimédias
numériques**

Cours de Master « Document Numérique Multilingue »

2006 - 2007

Fascicule III

Paris, Institut National des Langues et Civilisations Orientales
Janvier 2007

Sommaire

VII/ L'APPROPRIATION DE L'INFORMATION DANS UN FOND AUDIOVISUEL – UN SCENARIO.....	3
6.1/ LE VIDEO-LIVRE INTERACTIF SUR LE PORTAIL PCI.....	3
6.2/ LA REPUBLICATION DE CORPUS AUDIOVISUELS.....	4
6.3/ L'EXEMPLE DU SITE « GUANACACHE. LES EAUX DE LA SOIF »	6
6.4/ LA PLACE DE L'HYPERTEXTUALITE DANS LE PROCESSUS DE LA REPUBLICATION	6
VIII/ LOCALISATION ET EXPLORATION D'UNE INFORMATION DANS UN CORPUS AUDIOVISUEL –	
EXEMPLE ET SCENARIO « IDEAL »	8
7.1/ « TRAVAILLER AVEC UN CORPUS AUDIOVISUEL »	8
7.2/ SITE PORTAIL ET COLLABORATOIRE – L'ESPACE NUMERIQUE DE TRAVAIL.....	9
7.3/ VERS UN SERVICE « D'INTERET GENERAL » (« COMMUNITY SERVICE ») DES AAR.....	10
VIII/ LE CADRE THEORIQUE.....	12
7.1/ LA SEMIOTIQUE DU TEXTE	12
7.2/ LE TEXTE COMME ENTITE STRUCTURALE	13
7.3/ L'ENVIRONNEMENT PERTINENT DU TEXTE	14

VI/ L'appropriation de l'information dans un fond audiovisuel – un scénario

6.1/ Le vidéo-livre interactif sur le portail PCI

L'appropriation (la lecture, la compréhension, ...) d'une information se fait, dans le cadre du PCI et de tous les autres sites appartenant à l'« anneau » AAR, sous forme d'une consultation libre et donc, selon l'utilisateur, plus ou moins systématique, de sites web dynamiquement générés (à partir des données contenues dans la base de données) pour un « événement scientifique » (un entretien, une conférence, un séminaire, un documentaire, ...).

L'élément central et principal y est la vidéo de l'événement. Cette vidéo est publiée en analogie au genre « livre ». Dans ce sens on parle du « vidéo-livre interactif », d'une vidéo consultable via une "table de matière" (un sommaire) interactif dont les titres renvoient au passages vidéos concernés (à la manière de la table de matière dans un livre « classique » qui renvoie aux chapitres du livre).

Prenons l'exemple d'un entretien réalisé à la FMSH à Paris en novembre 2005 avec l'historien bolivien René Arze sur les guerres dans l'histoire de la Bolivie et le rôle déterminant dans ces guerres joué par les combattants indigènes.

Cet entretien a été filmé dans son intégralité et enregistré sur des cassettes DV. Une partie a été ensuite re-numérisée (en format avi) et, puis, découpée, montée, indexée et, enfin, compressée et publiée en trois formats différents sous forme d'un site web dynamique accessible à partir d'ici.

Le site web dynamique (i.e. généré à partir des données relatives à cet entretien et stockées dans la base des données des AAR) ressemble, métaphoriquement parlant, à une sorte de livre.

La page d'accueil du site dédié à l'entretien avec René Arze fait allusion à la page de couverture d'un livre où on trouve d'une part des informations signalétiques relatives à l'événement et d'autre part une très brève présentation du contenu de l'entretien ainsi que du chercheur lui-même.

Sur la partie gauche de la page d'accueil, on trouve le sommaire, la table des matières de ce "vidéo-livre" interactif en ligne (vidéos, curriculum vitae, photos d'archives, ...). En cliquant sur les différents intitulés du sommaire, on accède aux "chapitres" correspondants du "vidéo-livre".

Le "chapitre" le plus central est évidemment le chapitre donnant accès aux différentes parties de l'entretien avec l'historien bolivien. Ces différentes parties composent, pour rester avec la métaphore livresque, les différents sous-chapitres du chapitre "vidéos". Ces sous-chapitres sont organisés essentiellement selon un point de vue thématico-chronologique, c'est-à-dire qu'un sous-chapitre :

- possède une certaine unité d'un point de vue du contenu;
- est positionné par rapport aux autres sous-chapitres en respectant l'ordre temporel du déroulement de l'interview.

Il s'agit ici, en d'autres mots, d'un genre particulier de publication (ou de republication) d'un événement scientifique (i.e. d'un entretien). Il constitue actuellement la seule clé offerte aux visiteurs du site PCI (et de tous les autres sites faisant partie de l'anneau AAR) pour s'approprier (lire, interpréter, ...) le contenu des ressources audiovisuelles.

Pour diversifier les possibilités d'appropriation, un travail de R&D important est actuellement entrepris portant sur la problématique de la republication de corpus audiovisuels (et, plus largement, multimédias).

Références:

- Pour plus d'informations sur le genre "vidéo-livre interactif", cf. le cours en ligne "The interactive hypermedia book" (IHB) dedicated to the Huarpe civilisation" (Lisbonne, septembre 2006)

6.2/ La republication de corpus audiovisuels

Cette problématique concerne l'utilisation d'un même corpus audiovisuel (multimédia) pour des genres de publication différents où un genre de publication est supposé répondre à des contextes d'usage particulier, voir à des profils spécifiques des lecteurs des documents audiovisuels.

Par exemple, le corpus audiovisuel du site portail PCI peut servir à la fois à l'enseignement (secondaire ou supérieur), à l'apprentissage "libre" ou personnel, à la recherche, à un certain type de journalisme, à l'éveil de la reconnaissance et du respect de la diversité des cultures et des moeurs, à des usages professionnels comme, par exemple, au personnel travaillant pour les ONG, et ainsi de suite.

A ces types d'usages correspondent grosso modo et plus ou moins bien des genres de publication tels que des cours, des documentaires, des reportages, des vidéo-lexiques, des "highlights", des débats virtuels, des versions plus ou moins libres et dans différentes langues, etc.

Les principaux enjeux ici sont les suivants:

1. description et modélisation des genres de publication et mise en place d'une sorte de "grammaire de production" de modèles de publication audiovisuelle et/ou multimédia;
2. identification des principales activités et tâches composant le processus de la republication;
3. développement d'un environnement technique, d'un outil informatique de republication de corpus audiovisuel.

Ces trois enjeux constituent une partie essentielle des deux projets de R&D déjà mentionnés Saphir et Logos.

Références:

- Présentation pratique de l'outil Interview (co-développé par l'INA Recherche et l'ESCoM) utilisé actuellement pour republier des versions multilingues d'un entretien publié sous forme d'un vidéo-livre.
- Présentation technique de la republication d'un corpus audiovisuel s'appuyant sur le modèle du "vidéo-livre interactif" (rapport ESCoM réalisé dans le cadre du projet de R&D Logos).
- Cours (en anglais) dédié à la question de la republication ("re-authoring") de corpus audiovisuels dans un contexte multilingue et multiculturel.
- Accès à l'excellent projet Vox Populi du CWI visant la publication automatique de documentaires audiovisuels intégrant, selon un schéma d'argumentation donné (du philosophe Toulmin), des points de vue divergents et opposés à propos d'un événement donné tel que la guerre américaine en Afghanistan.
- Accès à un autre projet fort intéressant du CWI dédié à la publication de données multimédias suivant des schémas narratifs spécifiques (ici: biographie, portrait, chronologie de l'oeuvre, ... d'un artiste).

6.3/ L'exemple du site « Guanacache. Les eaux de la soif »

Le site "Guanacache. les eaux de la soif" (titre d'un livre du journaliste et auteur argentin Gregorio Manzur) est dédié à la valorisation et diffusion de l'héritage culturel lato sensu du peuple Huarpe - un peuple indigène vivant dans des conditions extrêmement difficiles dans le nord de la région de Mendoza en Argentine.

Ce site proposera des informations sur le peuple Huarpe et son monde via plusieurs genres de publication:

- "highlights"
- dossiers thématiques ou encyclopédie thématique;
- vidéo-lexique;
- dossiers pédagogiques;
- documentaires virtuels;
- versions multilingues;
- clips vidéos pour "nomades (mobile phone, ipod, psp, ...)
- etc.

Il s'agit d'un chantier de travail dont l'objectif est de développer localement un ensemble de services d'information (scientifique, didactique, ludique, professionnelle, multilingue, ...) qui doivent être disponibles pour le site portail PCI dans son ensemble (première version prévue pour le mois de septembre 2007).

Références:

- Présentation du site Guanacache. Les eaux de la soif et de son intérêt dans le contexte de la republication de corpus audiovisuels.

6.4/ La place de l'hypertextualité dans le processus de la republication

La problématique hypertextuelle dans le contexte de la republication de corpus audiovisuels est concernée notamment:

1/ par la mise en place de syntagmes ou encore par l'intégration syntagmatique d'un ensemble de segments audiovisuels en un "tout cohérent";

2/ par l'"enrichissement" d'un contenu donné sous forme de commentaires ("annotations") ou encore sous forme de mise en relation du contenu avec d'autres ressources d'information.

Dans le premier cas, la problématique hypertextuelle est concernée plus particulièrement par la dimension narrative dans la republication de corpus audiovisuels. Tout genre textuel (cours, documentaire, dictionnaire, livre, reportage, ...) possède son ou ses plans plus ou moins stéréotypés d'intégration selon lesquels des parcours de lecture, d'appropriation sont soit offerts soit imposés aux lecteurs (visiteurs, apprenants, ...).

Dans le deuxième cas, il s'agit de transformer un texte (un segment, un paragraphe, ...) donné et appartenant au corpus "de départ" en une "nouvelle" version - nouvelle version qui est plus ou moins proche ou, au contraire, éloignée de la version initiale mais qui n'est rien d'autre que le résultat partiel du processus de republication.

Dans les deux cas, la fonction hypertextuelle permet, si elle est bien contrôlée, d'avancer sur le chemin de la republication d'un corpus donné.

Références:

- Conférence sur la sémiotique narrative dans le processus de la republication d'un corpus audiovisuel (et, plus généralement, multimédia).

VII/ Localisation et exploration d'une information dans un corpus audiovisuel – exemple et scénario « idéal »

7.1/ « Travailler avec un corpus audiovisuel »

"Travailler avec un texte" (dans notre cas: avec un document audiovisuel) veut dire donner au visiteur/utilisateur d'un site tel que celui des AAR la possibilité de le "manipuler" à ses fins, i.e. selon ses besoins et ses intérêts.

L'exemple vraisemblablement le plus parlant est celui de la "table de travail" traditionnelle du chercheur, de l'enseignant ou de l'étudiant - table de travail qui constitue un espace à la fois physique et social de l'acquisition et de la production de connaissances. Concrètement parlant, le travail avec un texte s'inscrit toujours dans un projet quelconque dans le cadre duquel le texte constitue soit une ressource, un moyen pour atteindre un but quelconque soit un but en soi.

Toujours est-il que le travail avec (et autour) du texte se décompose en un ensemble d'activités typiques telles que:

- l'enregistrement et le classement (personnel) d'une (nouvelle) ressource textuelle;
- la prise de note (i.e. l'extraction des parties textuelles particulières pertinentes pour le projet de l'utilisateur);
- la production de résumés, de synthèses, de compte rendus, ... d'une partie textuelle;
- la production de commentaires sur telle ou telle partie en relation avec le projet;
- la vérification de telle ou telle partie textuelle quant aux informations fournies (sous forme, par exemple, d'une recherche de nouvelles informations, d'informations complémentaires, ...);
- l'ajout de références, de "liens" à d'autres ressources avec comme but de rendre plus clair le propos d'une partie textuelle traitée, de le "relativiser", de le critiquer, de le compléter, etc.;
- la traduction (telle quelle, résumant, sous forme d'une version plus ou moins libre, ...) de telle ou telle partie textuelle;

- l'enregistrement et le classement (personnel) de toutes ces activités de "traitement d'information";
- la mise à jour de s résultats de telle ou telle activité spécifique de "traitement d'information";
- et, enfin, la publication (lato sensu) de ces activités (ou d'une partie de ces activités) sous forme d'une simple réunion de pages manuscrites ou imprimées à usage purement personnel, sous forme d'un modèle préétabli de republication, voire sous forme d'une véritable publication originale.

7.2/ Site portail et collaboratoire – l'espace numérique de travail

Comme on le sait, toutes ces activités de travail sur et autour d'un texte peuvent être réalisées soit par une personne seule, soit par un groupe de personnes travaillant ensemble pour réaliser un projet commun.

Peu importe s'il s'agit d'un travail individuel ou d'un travail collectif, fournir ce genre de services aux visiteurs/utilisateurs d'un site portail tel que celui des AAR, présuppose de passer d'une conception de type "portail d'information/de connaissances" à une conception de type "collaboratoire" ou encore "espace numérique de travail".

Un collaboratoire ou un espace numérique de travail réunit donc, très généralement parlant, d'une part des ressources numériques d'information (sous forme de documents écrits, audiovisuels, ... organisés en des bibliothèques électroniques) et d'autre part des services et outils/instruments nécessaires pour effectuer un travail tel que celui présenté ci-dessus.

Références:

- une présentation synthétique du "collaboratoire" sur Wikipedia;
- l'ouvrage collectif Portails et collaboratoires pour l'enseignement et la recherche (coordonné par P. Stockinger); Paris Hermes Science Publications 2003.

7.3/ Vers un service « d'intérêt général » (« community service ») des AAR

Dans le cadre concret du Programme AAR, cela veut dire qu'un visiteur/utilisateur enregistré devra avoir la possibilité:

- d'obtenir son propre espace personnel de travail (i.e. sa "table de travail");
- de se constituer son propre corpus audiovisuel (et, plus généralement, multimédia) de travail;
- de créer son propre cadre de classement (son propre "thesaurus", voire sa propre "ontologie");
- de réaliser à distance (!) l'ensemble des activités citées ci-dessus (découpage/segmentation des vidéos, annotations, enrichissement, traduction, ...);
- de procéder à distance (!) à la publication personnalisée des résultats de ses activités;
- de publier les résultats de ses activités sur un site personnel accessible uniquement à ceux identifiés par le travailleur.

En comparant ces exigences avec le processus de republication d'un corpus audiovisuel afin de permettre une meilleure appropriation des connaissances contenues dans le corpus, alors on peut aisément se rendre compte qu'il s'agit ici essentiellement de passer d'un modèle "centralisé" d'exploitation d'un corpus source à un modèle totalement décentralisé:

- dans le cadre du modèle "centralisé", c'est une institution, un service qui - comme traditionnellement l'éditeur ou le producteur/metteur en scène - règle pour le lecteur/le destinataire visé le processus et les modalités de l'appropriation de l'information;
- dans le cadre du modèle "décentralisé", c'est la communauté concernée (dans notre cas: les enseignants, les étudiants, les chercheurs, ...) qui prennent en main leurs besoins et intérêts d'appropriation sous forme d'un travail adapté sur et autour du texte.

C'est ainsi qu'un portail tel que celui des AAR peut se transformer en un service d'intérêt général qui non seulement fournit des informations ou des connaissances à une communauté donnée mais permet également à celle-ci de l'utiliser à ses propres fins.

Références:

- le projet OPALES (programme RIAM: 2001 - 2004), coordonné par INA Recherche, ayant eu comme but à définir un outil de travail personnalisé sur des corpus audiovisuels;
- la conférence de P. Stockinger "Digital audiovisual archives in humanities" (Chania, Grèce 2003) expliquant les objectifs du projet OPALES.

VIII/ Le cadre théorique

7.1/ La sémiotique du texte

Comme déjà signalé dans la présentation générale du cours, les activités de R&D autour de la publication/republication de corpus audiovisuels (et, plus généralement, multimédias) et plus spécifiquement autour de l'hypertextualité, font référence à un cadre théorique et méthodologique constitué par l'analyse du discours et la sémiotique du texte.

Le texte est le support nécessaire à toute communication pour véhiculer, faire circuler une information entre les membres d'une communauté de communication.

Les approches en sciences du langage qui s'occupent de l'étude du texte l'étudient d'une manière analogue à la linguistique (structurale) des langues comme un objet qui

- possède une organisation autonome interne
- s'inscrit dans et se positionne par rapport à un environnement (culturel, social, technique et naturel).

Il existe toute une diversité de traditions et courants (linguistiques, rhétoriques, sémiotiques, cognitifs, ...) qui s'occupe du phénomène "texte" dans sa globalité ou en en privilégiant tel ou tel aspect particulier.

Références:

- analyse (sémio-linguistique) du discours (P. Charaudeau; D. Maingueneau)
- la sémiotique de l'écriture d'Yves Jeanneret.
- la sémiotique de la langue (Michael Halliday);
- la théorie de la structure rhétorique (William Mann, Sandra Thompson, ...)
- la "critical discourse analysis" (Norman Fairclough, ...);

7.2/ Le texte comme entité structurale

Le texte entendu au sens d'une entité structurale est supposé de posséder une organisation interne spécifique qu'il partage avec un ensemble d'autres textes faisant partie du même type ou - comme on dit - genre.

Afin de pouvoir expliciter et décrire plus systématiquement l'organisation interne d'un texte (d'un corpus de textes), on fait appel à un ensemble de plans constitutifs dont les suivants:

- plan thématique du texte: sélection et thématisation des informations traitées et développées dans le texte (le corpus de textes);
- plan narratif lato sensu: intégration syntagmatique des informations sélectionnées et traitées dans le texte (le corpus de texte);
- plan rhétorique et discursif: choix d'un acte ou genre de discours pour parler d'une information sélectionnée et thématisée dans le texte (le corpus textuel);
- plan de l'expression: choix d'un langage et de medium d'expression dans le texte (le corpus textuel);
- plan de l'organisation formelle et physique de l'information traitée dans le texte (le corpus textuel);
- plan du support physique et technique de la "matérialité" (R. Chartier) du texte (du corpus textuel).

Chaque plan constitue l'objet d'importants travaux à la fois théoriques et empiriques impossibles à résumer ici en quelques lignes.

Ceci dit, une connaissance opérationnelle des différents plans est indispensable pour les différents secteurs des technologies de l'information et de la connaissance.

Références:

- livre de Peter Stockinger, *Le document audiovisuel*; Paris, Hermes Science Publications 2003

7.3/ L'environnement pertinent du texte

D'une manière analogue à l'acteur social, un texte possède son "monde vie", c'est-à-dire qu'il est:

- participe à des activités et pratiques sociales;
- interagit et communique avec d'autres textes en formant un "textscape", une "communauté de textes" plus ou moins typique, plus ou moins stable ou, au contraire, éphémère;
- se "nourrit" et nourrit d'autres textes en constituant ainsi un élément dans l'évolution textuelle et culturelle.

Description et Modélisation des relations intertextuelles sont à la fois théoriquement et empiriquement nécessaires :

- pour comprendre le sens des rapports hypertextuels dont est pourvu un document numérique
- et pour un projet de publication/republication d'un document numérique (audiovisuel).