

Sémiotique du document virtuel personnalisable

L'exemple du document audiovisuel

Peter Stockinger



**10 et 11 Juillet 2002 à Brest (France)
ENST Bretagne**

Peter Stockinger

© ESCOM et auteur, 2004

Maison des Sciences de l'Homme
54 Boulevard Raspail
75006 Paris

Serveur web : <http://www.semionet.com>

Mentions légales :

Les œuvres scientifiques, orales et écrites, diffusées dans la collection *E-Publication* de l'ESCOM, sont protégées par la législation française et les conventions internationales relatives au droit d'auteur.

Les utilisations gratuites et exclusivement destinées à des fins d'enseignement et de recherche de ces œuvres sont les seules autorisées par les auteurs, nonobstant les dispositions du code français de la propriété intellectuelle.

Toute autre utilisation (reproduction, représentation, traduction,...), et notamment à des fins commerciales, est interdite sans l'obtention de l'accord préalable des auteurs ou des titulaires de droits concernés. L'utilisation non autorisée (par la loi ou par l'auteur) d'une œuvre est une contrefaçon, et toute contrefaçon est un délit (article L.335-2 du Code de la propriété intellectuelle).

Ce document a été composé par Elisabeth de Pablo

Table des matières

- 1) LE DVP – définitions, objectifs et problèmes
- 2) Deux exemples : la mise en ligne de manifestations scientifiques et la description de documentaires audiovisuels
- 3) La sémiotique textuelle
- 4) De la description sémiotique aux « applications »
- 5) Le scénario sémiotique

1) Le DVP - définitions, objectifs et problèmes

1. 1) Définitions

Un document virtuel se présente sous forme de :

- « fragments (textuels) » ou « unités documentaires »;
- un « espace d'information composé, constitué par des fragments, d'unités documentaires » et de stratégies, de méthodes pour assembler ces fragments

cf. à ce propos : Sylvie Ranwez et Michel Crampes : « information brick » vs « real document »
“a Virtual Document is a composite class, IBs are *components*, and techniques and tools are *construction methods*. A *Real Document* is then a sequence of IBs generated by the methods in concordance with the user's specification.”

1.2) Objectifs multiples

- Génération et édition d'une documentation « sur mesure »;
- Création de versions (compléments, mise à jour, ...) d'une partie (séquence, segment) d'un document;
- Annotation de parties (séquences, segments) pertinents dans un document (un corpus de documents);
- Indexation de parties (séquences, segments) thématiques;
- Réutilisation de parties (de segments pertinents) d'un document dans d'autres produits documentaires;
- Recherche et exploration (de segments pertinents dans un espace d'information – ex. : bibliothèque ou vidéothèque numérique, « base de connaissances »).

1.3) Remarques

a) cette conception générale correspond, grosso modo, à celle des disciplines travaillant traditionnellement sur la problématique de la production et de l'interprétation (lecture) du texte (lato sensu) :

- distinction entre d'une part, les « éléments textuels » (existants) et le texte en tant que « produit », résultat d'une production (« écriture ») - pas de « création ex nihilo »
- distinction entre d'une part, les standards (des « modèles » : genres, « scénarios », ...) textuels et d'autre part les versions (« réalisées ») d'un ou de plusieurs standards
- distinction entre le « contenu » lui-même dans une donnée textuelle (lato sensu), l'*expression*, la *manifestation* de ce contenu et le *support* à proprement parler

b) Cependant - différence entre :

- d'une part la conception de l'« unité documentaire », de l'espace d'information » et des méthodes de construction (conception très générale et aussi très simplifiée de l'objet « texte »)
- d'autre part la *complexité structurale* (multi-dimensionnelle) inhérente à un texte (un document, un corpus de documents) - complexité qui peut être indirectement appréciée par le phénomène des « lectures multiples » du texte, de tel ou tel segment d'un texte

c) se pose donc plutôt le problème de la description à proprement parler d'une donnée textuelle en amont de toute conception-développement pratique (en général) et informatique (en particulier)

Une description présuppose une théorie, un cadre théorique qui permet l'élaboration d'un langage de description (problématique des critères d'une description) et qui justifie et explicite une description particulière (problématique de la « connaissance de l'objet » ; i.e. de l'objet textuel) . Une description présuppose également une méthodologie du faire descriptif

d) un cadre théorique possible : la *sémiotique du texte* en particulier et les différentes disciplines traitant la problématique du texte (tant du point de vue de la production, tant du point de vue de l'interprétation, tant du point de vue purement structural, ...)

2) Deux exemples : la mise en ligne de manifestations scientifiques et la description de documentaires audiovisuels

2.1) Le projet français OPALES (Outils pour des Portails Audiovisuels Educatifs et Scientifiques)

Objectifs :

- permettre l'enregistrement, la numérisation et la diffusion sur le web de colloques, séminaires et entretiens avec des chercheurs en SHS
- constituer un "patrimoine", une "base de connaissances" qui évolue, se met à jour, et qui sert à des exploitations multiples, notamment dans l'optique de l'"édition multi-support" (édition sous forme de sites web, sous forme de CD-ROM, sous forme de DVD, sous forme de cassettes VHS, etc.) et "édition sur mesure" (sélection de segments ou de séquences thématiquement pertinents aux besoins ou intérêts d'information d'un utilisateur et assemblage sous forme d'un seul et unique produit d'information).
- permettre aux chercheurs (enseignants, ...) de "travailler" ces vidéos, de les exploiter activement soit dans le cadre de ses recherches, soit dans le cadre de ses activités d'enseignants.



Remarques

- Travailler, exploiter activement une vidéo (un enregistrement d'une conférence ou d'un colloque) veut dire, entre autre :
- Pouvoir identifier et délimiter, selon ses propres intérêts et besoins, des séquences thématiques dans une vidéo,
- Pouvoir indexer ces séquences thématiques, les décrire et/ou annoter;
- Pouvoir se constituer ses propres "archives" (partagées ou non avec d'autres chercheurs d'une communauté) en vue de préparer une publication, une conférence, un séminaire, etc.

Problèmes qui se posent :

- Fournir des modèles de lecture, de compréhension d'une vidéo au sens d'un enregistrement d'une activité scientifique (conférence, séminaire, entretien avec un chercheur, ...)
- Produire des outils de description (d'indexation) plus sophistiquée du « contenu » d'une conférence, d'un séminaire.

Etat de l'art

Actuellement souvent utilisés pour identifier et décrire le contenu d'un segment audiovisuel :

- ontologies et thesaurus de thèmes (ou concepts) d'une pièce d'information, d'une ressource numérique (ontologies sous forme de classification taxinomiques et/ou mérologiques de thèmes ou concepts)
- outils de segmentations physiques (découpage) ou logiques (« méta-données »)

Limites :

- pas de possibilités de tenir compte des formes d'actualisation (d'hiérarchisation) d'un thème dans un segment textuel (rhétorique du thème, degré d'explicitation, ...);
- aucune prise en compte des formes d'intégration (i.e. de développement) d'un thème dans un segment audiovisuel;
- aucune prise en compte des formes de manifestation de thèmes dans un segment audiovisuel.



Par exemple un entretien peut posséder :

- un contenu historico-chronologique,
- un contenu critique (i.e. organisé autour de commentaires, etc.),
- un contenu explicatif (i.e. organisé autour d'explications et d'exemples)
- un contenu plutôt pratique et appliqué,
- un contenu plutôt exploratoire et hypothétique,
- un contenu composé d'exemples, voire d'anecdotes personnelles ou professionnelles
- un contenu qui se construit comme réplique argumentée,
- etc.

L'explicitation de telles formes sous forme de scénarios permet de développer des environnements plus sophistiqués de production, d'échange et de partage de « contenus » stockés dans des ressources audiovisuelles numériques.

2.2) La Description des documentaires audiovisuels

Domaine

- Documentaires audiovisuels commercialisés;
- Documentaires « scientifiques ».

Importance

- Offre télévisuelle de très grande ampleur;
- Utilisation dans des contextes de communication les plus divers;
- « Base de connaissances » analogiquement à l'encyclopédie classique.

Objectifs

- Développer des « formats » de description suffisamment complexes pour tenir compte (au moins partiellement) des spécificités d'un documentaire;
- Décrire (identifier, indexer, classifier et annoter) des « segments audiovisuels » non pas au sens physique (exclusivement au sens physique) mais au sens de leur pertinence pour un utilisateur.



- Permettre la production de « nouveaux » produits audiovisuels sur la base des segments décrits

Contexte

- Projet Opales
- Projet Européen Knosos (portail pour l'e-learning)

Remarques :

Notion « documentaire audiovisuel » devient ambiguë :

a) Le texte « documentaire » : « réseau » hiérarchique de segments textuels (plans, scènes, séquences, ...)

b) Le segment peut être compris :

- au sens physique (i.e. le texte du documentaire découpé en des unités documentaires « indépendants »),
- mais aussi au sens « subjectif » (i.e. c'est le spectateur qui décide en quelque sorte ce qu'est un segment et pourquoi)

c) Les segments (physiques et/ou logiques) :

- Constituent une sorte de « bibliothèque » de données audiovisuelles

d) Notion du documentaire (revisitée) :

- Est une « solution » (pour un problème de communication donné)
- Il y a d'autres « solutions » possibles

e) Problème : développer des formats de description appropriés pour rendre compte de la complexité du genre « documentaire » (cf. les travaux sur le site de l'ESCOM : <http://www.semionet.com>)

3) Sémiotique textuelle

Le texte, la donnée textuelle (lato sensu) est appréhendé(e) à travers différents « constituants » typiques dont les suivants ([cf. STO99]) :

Chaque composant est en soi une entité assez complexe et parfois encore mal maîtrisée :

le composant (inter)textuel : à la charge de partitionner et de positionner des segments textuels selon leur statut et « tâches » dans la production et la communication d'une information):

- délimitation hiérarchique
- délimitation fonctionnelle (rôle)
- positionnement des segments non seulement à l'intérieur du même texte mais aussi entre différents textes

le composant thématique : à la charge de sélectionner et l'organiser des informations dans un texte ou un corpus de textes

- identification et description des thèmes
- classification des thèmes (classification taxinomique, classification méréologique, classification par schémas d'action, ...)

le composant narratif : à la charge d'intégrer des thèmes sous forme de parcours narratifs au sens large du terme

- description des « actes de discours »
- description des schémas et genres narratifs (description, explication, comparaison, chronologie, débat, témoignage, reportage, enquête, ...)
- description des thèmes narratifs (i.e. du rôle narratif d'un thème)
- ...

le composant de mise en scène discursive : à la charge le traitement à proprement parler d'un thème sélectionné en lui conférant un « profil » qui lui est propre à l'intérieur d'un texte ou



d'un corpus textuel

- description des thèmes « manifestes » vs thèmes implicites (degré de l'implication/explicitation d'un thème)
- description du « poids » d'un thème (degré de l'expansion/condensation d'un thème)
- description du « cadrage » d'un thème (point de vue, « cadre », focalisation, ...)

le composant de l'expression des informations : à la charge le choix d'un ou de plusieurs « langages » d'expression d'une information)

- description des codes de manifestation (synchrétique, linguistique, visuel, ...)

le composant de l'organisation formelle et physique de l'information (ayant à la charge de « visualiser » au sens large du terme une information sur un support donné)

- description de la composition, organisation d'un thème traitée, développée sur une « scène » (visuelle, audiovisuelle, ...)
- description de la réalisation d'une « scène » sur un support : page, image, tableau, dessin,

le composant de l'interaction entre la donnée textuelle et son lecteur/utilisateur : à la charge la gestion de la réception, de la « lecture » et de l'exploitation d'un texte par un destinataire

- description des parcours de lecture
- description de la reconstitution (mentale et/ou physique) d'un texte sous forme d'une nouvelle « écriture »
- description de l'usabilité (Le Coadic) et des usages

Composant « cycles de vie » d'une donnée textuelle : à la charge la gestion des différentes phases et activités liées à la période de l'existence d'une donnée textuelle

- Description « généalogique » (de la « première idée » au produit fini)
- Description des phases et étapes de production-édition
- Description des mises à jour, rééditions, ...

Remarques :

a) Ces composants constituent les différentes facettes de la *connaissance du domaine de la description sémiotique*

b) Mais ils montrent aussi les facettes multiples d'une donnée textuelle (lato sensu) et donc :

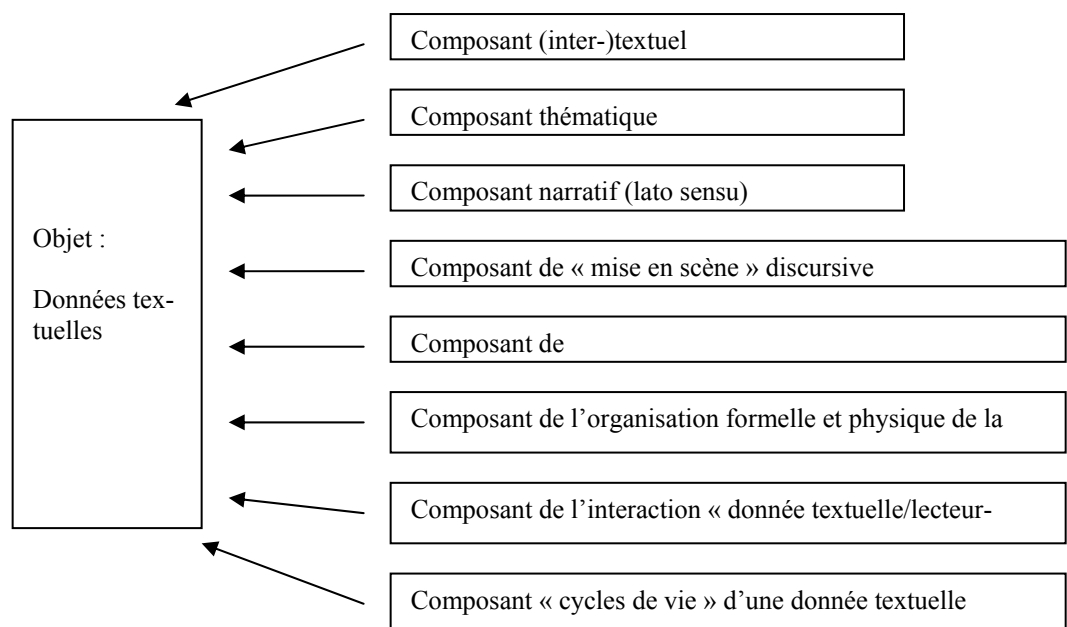
- la nature très diversifiée, voire diverse des unités documentaires ou fragments composant un espace d'information

c) autre aspect important : de la notion « *composant* (d'une donnée textuelle) » à la notion de « *service* » :

Un composant est là, parce qu'il sert, parce qu'il joue un rôle (dans la communication) :

Un document, une donnée textuelle = une sorte d'espace, d'organisation d'un ensemble de services de communication. Par exemple :

- le composant thématique comme service de filtrage, de sélection, de hiérarchisation, d'organisation d'informations
- le composant narratif comme service de conseil, d'avertissement, d'explication, ... d'une information
- le composant de l'expression comme service de « visualisation » d'information.



Cf. la notion du document *non pas* au sens éditorial (« imprimé, « livre ») mais au sens d'une « institution couchée sur papier », d'une « organisation »

Cf. l'intérêt de cette vision pour la compréhension des sites web (à la fois document au sens classique et sorte de lieu social, d'institution)

Remarque au sujet de la description sémiotique d'une donnée textuelle :

Selon ses objectifs, une description concrète privilégiera tantôt telle, tantôt telle autre facette.

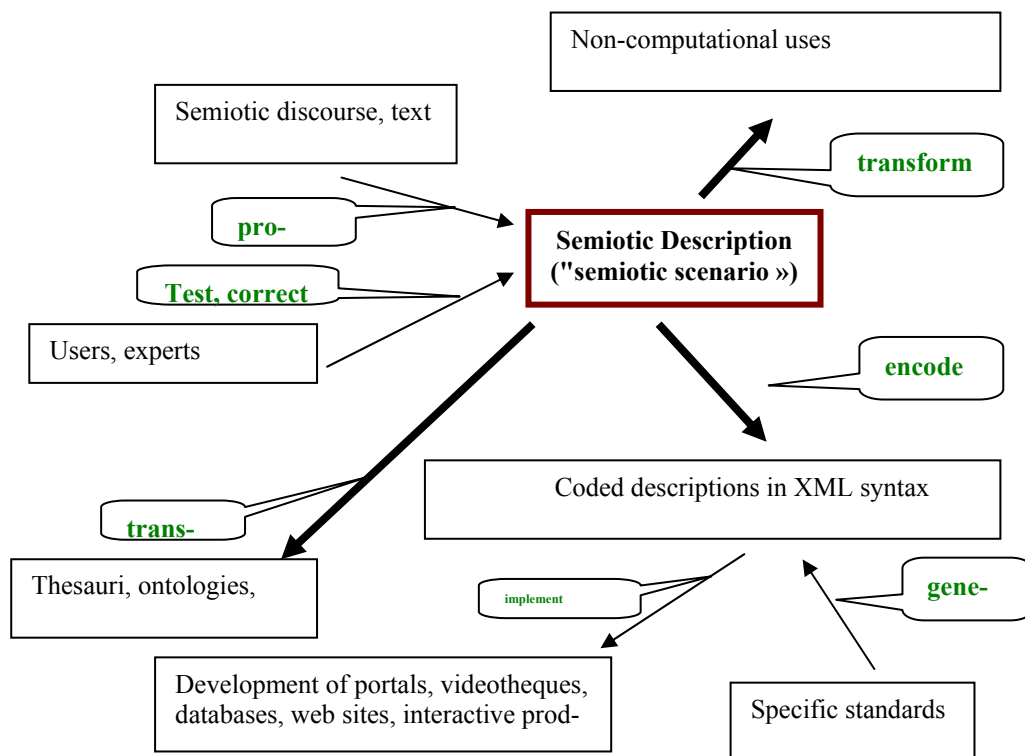
Ainsi peut-on avoir :

- des descriptions plus orientées vers une analyse du composant thématique (c'est le cas, très typique, si on souhaite produire un thesaurus, un index ou encore une ontologie pour un corpus textuel);
- des descriptions plus orientées vers une analyse du composant rhétorique (c'est le cas, par exemple, si on s'intéresse davantage à la question du genre narratif et/ou textuel);
- des descriptions plus orientées vers une analyse de la mise en scène discursive (c'est le cas, par exemple, si on s'intéresse plus spécifiquement à des traitements spécifiques d'une information thématisée dans un texte);
- des descriptions plus orientées vers une analyse de l'expression, de la « visualisation » (au sens large du terme) d'une information (c'est le cas, par exemple, si on s'intéresse aux types de langages utilisés pour exprimer et communiquer une information, aux rapports entre les différents langages utilisés, etc.);
- des descriptions plus orientées vers une analyse de l'organisation logique et physique d'une information sur un support donné (c'est le cas, par exemple, si on s'intéresse non seulement à la « lisibilité », voire à l'« impact » visuel d'une information mais aussi à la standardisation de sa production-édition);
- des descriptions plus orientées vers une analyse des rapports intertextuels entre les différentes parties – séquences – d'un texte et/ou différents textes (c'est le cas, très typique, si on s'intéresse non seulement à l'hyper-textualité mais aussi aux rapports qu'entretient un texte avec son « environnement » culturel et historique);
- des descriptions plus orientées vers une analyse de l'axe « texte/lecteur-

utilisateur » (c'est le cas si on s'intéresse non seulement à la « compréhension » d'un texte, aux interprétations parfois multiples qu'un texte peut donner mais aussi à son « utilité », à l'usage qu'on en fait).

Sous la couverture de « description sémiotique » peuvent se cacher des descriptions fort diverses, fort variées mais qui partagent néanmoins un cadre commun de référence théorique qui est la sémiotique textuelle – cadre de référence pour un type général de connaissances du domaine textuel.

4) de la description sémiotique aux « applications »



5) le « scénario sémiotique »

Différents types de scénarios :

- Scénarios de produits ou services interactifs (sites web de toutes sortes, produits pédagogiques, produits d'information, ...)
- Scénarios pour l'indexation et la classification de données textuelles
- Scénarios pour décrire l'accès à une « bibliothèque de données textuelles » et son exploration
- Scénarios de visualisation (lato sensu) d'information
- Scénarios d'identification, de localisation d'informations
- Etc.

Ces scénarios :

- peuvent être de simples aides ou guides (par exemple dans la conception de sites web ou dans la mise en place d'un thesaurus)
- peuvent être aussi codés ou codables dans un standard technologique tel que XML ou intégré dans un logiciel spécialisé (par exemple comme un nouveau "knowledge object" dans la bibliothèque du logiciel d'auteur Authorware, ...)
- peuvent être aussi accompagnés d'outils et de services d'utilisation et de gestion, ce qui les rapprochent de ce que l'on appelle maintenant les *services web* qui sont des "composants" qu'un utilisateur peut télécharger afin de développer sa propre application.

Remarque finale : distinction entre « méta-description » et « descriptions ». Il faut distinguer entre :

- Ce que nous appelons la "*description de la description*" (cf. le cadre conceptuel cité) qui est moins une description à proprement parler d'une donnée textuelle particulière (par exemple d'un fonds documentaire, d'un corpus de textes, etc.) que l'explicitation des critères guidant telle ou telle description d'une donnée textuelle
- Et les *descriptions* (parfois concurrentes, parfois complémentaires, ...) à proprement

parler d'une donnée textuelle (d'un fonds documentaire, d'un corpus, ...)

Les « scénarios » relèvent plutôt de la méta-description : ils « imposent » un cadre commun à des descriptions à proprement parler de données textuelles

- Les « scénarios » définissent donc des formats qui sont partagés par des descriptions de données textuelles – descriptions qui peuvent être, cependant, différentes, voire concurrentes.

Références:

S. GARLATTI, S. IKSAL : Documents virtuels personnalisables pour des systèmes d'informations en ligne (DVP 2001)

Sylvie Ranwez, Michel Crampes : Conceptual Documents and Hypertext Documents are two Different Forms of Virtual Document (DVP 1999)