



HAL
open science

Repurposing digital audiovisual resources for user communities possessing different cultural and linguistic profiles

Peter Stockinger

► **To cite this version:**

Peter Stockinger. Repurposing digital audiovisual resources for user communities possessing different cultural and linguistic profiles. [Research Report] Institut National des Langues et Civilisations Orientales; Fondation Maison des Sciences de l'Homme. 2012. halshs-00813000

HAL Id: halshs-00813000

<https://shs.hal.science/halshs-00813000>

Submitted on 14 Apr 2013

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

Repurposing digital audiovisual resources for user communities possessing different cultural and linguistic profiles

Peter Stockinger
Fondation Maison des Sciences de l'Homme (FMSH)
Equipe Sémiotique Cognitive et Nouveaux Médias (ESCoM)
54, Bd. Raspail – 75006 Paris – France
Tel. : 01 49 54 21 83 – email: stock@msh-paris.fr

1/ Introduction

In Europe and elsewhere there are more and more available digital libraries or archives composed of textual, audiovisual, visual, sound, ... objects which constitute *potential* knowledge resources for specific user communities, *viz.* individual users [CHA 02].

But such as, they actually do not (always) satisfy this mission of being a knowledge resource pool in, for instance, formal or informal education, home leisure activities, professional environments, etc. Such as they are only *potentially* interesting, they only have a *potential added value* for specific user groups.

A simple text description of an oeuvre exhibited in a museum *may be* pedagogically relevant in a specific *context*, for a given user and with respect to his/her curiosity to be satisfied. In another context, this same description will not be able to satisfy the user ... and therefore be considered as not relevant, not useful, etc. by the learner [STO 03b].

The potential relevancy of a digital object depends also on the *type of the knowledge appropriation* understood either as a process of “simple (*article* or *book* based) lecture”, or contrarily as an active interaction with a digital knowledge object, i.e. as a more complex global activity composed by and organised via different local cognitive acts implied such as comparing, testing, working actively with a resource, playing, simulating, etc ([STO 92], [STO 94], [STO 95], [STO 99]).

The potential relevancy depends also on the *evaluation of the learning* itself. For instance, the knowledge appropriation is it performed mainly with respect to the personal stilling of curiosity or, contrarily, with respect to formal or semi-formal rules of evaluation, notation prevailing in formal educational situation.

Finally, the potential relevancy of a digital object and therefore it's (more or less) added value for a targeted user or user community, depends also of the *place* that occupies the digital object within a range of other resources available for a learning process: it can be a prescribed, “imposed” one that the “learner” has to use, it can be a more or less central one, a traditionally used one, it can be a more or less good or risky “candidate” with respect to a required (intended, needed, desired, ...) output of a learning process largely speaking, and so on [STO 07]).

In looking on the web, we have very different categories of digital objects which make them more or less suitable to become generally speaking relevant resources for a user or a user community, and more specifically speaking, *relevant learning resources*. For instance, we have:

1/ A huge category of rough, more or less *unstructured* digital objects (simple pages, photos, films, etc.)

2/ Categories of digital objects organised following *traditional publishing genres* (for instance: as digital books with chapters, as digital journals, magazines, as digital dictionaries and encyclopaedias, as digital atlas, etc).

3/ Categories of digital objects that are structured, organised with respect to *specific uses* and exploitations among which more particularly for the use in formal learning settings such as schools, universities, vocational training centres, etc.

All these categories of digital objects can become user relevant resources, and more particularly resources potentially relevant for specific learning situations. The important point here is that, such as, they are maybe only partially, only more or less well attuned for a given situation. This means that they have to be (re-)processed, (re-)published with respect to the constraints and expectations of a specific user community [STO 00].

Let's take the concrete example of the French programme Archives Audiovisuelles de la Recherche (AAR)¹. This archive actually recovers actually already more than 4000 hours of streamed video, accessible for everybody, in all disciplines of human and social sciences. The records composing the AAR, contain individual interviews with researchers in all scientific disciplines, conferences, lectures, research seminars, workshops, small reportages about the daily life in research labs, scientific expositions, documentaries, travelogs and road movies, ethnographical films, etc. The offer is composed of more than 120 research seminars, symposia and workshops as well as of 260 interviews with researchers in palaeontology, archaeology, ethnology and anthropology, social history, cultural history, art history, anthropology, language and communication sciences, sociology, economical sciences, political sciences, philosophy as well as more particularly in oriental studies (language studies, culture, history, ...). More than 1150 researchers, scholars and professionals have already contributed to this audiovisual archive. Even if most of the videos are in French language, there also exist videos in English, Italian, Spanish, German and even Chinese and Hindi. The average growth per month of the audiovisual archive in social and human sciences is of about 40 hours on line, this means of about 50 hours of video taking and digitizing.

The principal user communities of the AAR are the (French speaking and international) research community itself, the educational community (3rd but also 2nd level) and particular professional communities (science journalists, stakeholders especially in non-profit organization, governmental agencies, ...). The whole composes a very specific "market niche" in the digital (audio/video) content production and communication. The daily average number of significant visits for the AAR are of about 900 to 1100 visits (statistics produced by an independent company) but together with the principal other programs in this "market niche" (*Canal U*², *Canal IRD*³, *Diffusion des Savoirs* of the ENS⁴, ...), the daily average number of significant visits may be situated between 3000 to 5000 visits for this type of very special and complex (audio/video) content as far as the French situation is concerned (this number is a rough approximation based on internal information).

¹ Adresse of web AAR web portal : <http://semioweb.msh-paris.fr/AAR/FR/>

² <http://www.canal-u.education.fr/canalu/index.php>

³ <http://www.canal.ird.fr/>

⁴ <http://www.diffusion.ens.fr/>

All available indicators show that this archive has indeed a potentially high impact for teaching in formal contexts but also for learning in informal settings (life long learning, etc.), for professional contexts such as people working for NGOs and preparing and/or documenting the field work, etc. But such as, this whole heritage has to undergo a complex process of repurposing, of adaptation in order to fit more specifically with user profiles and user contexts (cf. the second point, below).

The reason for this is, as already mentioned, the specific profile, the “identity” of the digital objects composing the AAR corpus – a profile or “identity” which is either too far from the expectations, needs or desires of an individual user or a targeted user community or simply not perceived, i.e. hidden either from a cognitive point of view (the content is *not understandable*) or from a physical point of view (the content is *not perceivable*). In a nutshell, there are at least three serious challenges we have to face:

1/ the *language limitation*; i.e. we have to improve the quality of existing digital content in opening and making available scientific and scholar content for an ***intrinsically multilingual knowledge market*** by the means of an extensive use of hints and aids for a linguistic understanding of a content performed in a given source language provided by user communities applying the same resources and tools;

2/ the *hidden information*; i.e. we have to improve the quality of existing digital content in eliciting, systematising and classifying the ***hidden information*** in large audiovisual databases by the means of user community centred description and indexing of digital corpora using the same description and indexing resources (ontologies) and tools (editors) ;

3/ the *adaptation to specific* □ *especially educational* □ *purposes*; i.e. we have to improve the quality of existing digital content for ***specific contexts of use*** (especially: formal and informal education) and an intrinsically ***multicultural market*** (characterized by diverging knowledge and value references, by different types of expectations and acceptations, interests and needs, ...) by the means of user community centred re-authoring or repurposing of the available content.

In order to better understand from a theoretical and methodological point of view these three challenges and how to handle them concretely, we will discuss shortly in this paper:

1. the notion of “user community” or again “user”;
2. the structural composition of a digital resource as a potential information or knowledge resource;
3. the process of repurposing (re-authoring) of digital resources.

2/ The notion of □user community□

There exist a huge variety of classifications of user communities trying to understand a user or user community as a *sociological* and *semiotic* (i.e. a meaning producing/reproducing and exchanging entity; [STO 83]). Conceptually speaking, the establishment of a typology of user communities is based on one or more criteria belonging:

- either to the *activity* and *task criteria* group (i.e. the user community understood as a collective entity sharing some interests or needs or again desires and aiming – with more or less resources, competencies and also luck – at the achievement of some goals, realisations, etc.);
- or to the *qualities* and *attributes* criteria group (i.e. a user community understood with respect to shared values, knowledge and skills, with respect to common tendencies and

preferences, with respect to the possession of social, cultural or economic resources, with respect to physical and biological attributes, etc.).

Depending on the issues of the description and modelling of user communities, the scenarios of such communities can become more or less complex, more or less sophisticated. They possess, in any case, a central impact on how to conceive and develop methodologies and tools for the production, management, exploitation, re-use, etc. of digital knowledge objects such as texts, films, images, etc. organised in digital libraries, archives or other forms of digital information and knowledge spaces (cf. below).

With respect to digital multimedia libraries or archives, user communities are habitually and basically identified and characterised with respect to their roles either as *content* [producers] (lato sensu), *content* [traders] and *content consumers* (lato sensu). This tripartite distinction, such as, is certainly too simple for being able to capture the whole complexity of user communities but they constitute nevertheless a starting point for investigating user communities and their dealing with multimedia libraries and archives.

One of the advantages of this tripartite distinction is to stipulate that there are basic activities characterizing the roles of the production, of the trading and of the consumption of digital objects or resources not matter the specific identity of the social actor or group performing them, no matter if they are performed by people belonging to a research laboratory, a classroom, a family, a group of people sharing a common hobby, a group of professionals working for a company, etc.

Central topics in the description and modelling of group-centred, i.e. community relevant and adapted digital multimedia libraries, archives and other forms of collections of digital objects are the following three ones:

- the structural organisation of the digital object or resource itself, i.e. the features or patterns that define the identity, the specific profile of a digital object or resource;
- the repurposing of digital objects or resources, i.e. the intentional manipulation of the profile of a given digital object or resource in order to adapt, to attune it to the profile of a destine community;
- the structural specificities of the “life world” of a user community understood as a social actor working with, exploiting, maintaining, etc. digital objects and resources.

Point 1 and 2 will be discussed again below. Let us have a look of what is meant by “life worlds” of a user community [SCH 79].

Intuitively speaking, this means that user communities have a specific, collective behaviour – *habitus* in the sense of sociologists such as A. Schütz [SCH 79] or P. Bourdieu [BOU 79]. They are social actors with their routine activities, their common interests, their common horizon of meaning, and so on. They are like ethnographic communities or again like “paper communities” as they exist in novels and other writings. The understanding and description of the life worlds of user communities are necessary in order to define:

- on the one hand the *structural specificities* (cf. point 3, below) of the digital objects that are of *relevancy* (i.e. *understandable, usable, useful, acceptable, ...*) for a specific community
- and on the other hand the tasks and tools adapted to the production and repurposing of content (information, knowledge, ...) by the user community (cf. point 2, above).

Very simply and concretely speaking – a user community possesses, among other resources, one or some languages for communicating but also eliciting and appropriating knowledge. This is a basic variable – the digital resources have to be within one of the available linguistic competences in order to become a consumable object for that community.

This is what is meant generally speaking under “translation”, i.e. the linguistic adaptation of an existing digital object to the linguistic horizon of the intended user community. Now, it is clear also that “translation” is not only a sort of a literal transposition of the meaning of the content of the digital resource within the horizon of targeted user community, it is much more a complex *process of interpretation and adaptation* of this meaning with respect to this horizon [STO 07].

Interpretation and adaptation of the meaning of a given digital resource can consist, concretely speaking, in the modification of the initial content itself, in its simplification, in its cognitive and axiological reorientation via comments and analysis, in the production of a truly new version of the initial content, etc.

In this sense translation, linguistic translation, has to be understood in a broader sense of what sometimes is called *cultural translation* or again (*semiotic*) *interpretation and adaptation* of a digital resource in order to transform it in an *object of value* (a *value added object*) for a targeted user community. This complex process of (semiotic) interpretation and adaptation is the *kernel intellectual or cognitive activity* of what is called the *re-authoring* or *repurposing* of an existing digital resource ([STO 92], [STO 99], [STO 07]).

Interpreting and adapting successfully an given digital resource and its content presuppose necessarily the (explicit or implicit) understanding of the expectations, needs, desires, etc. of the targeted user community as well as of its available cognitive skills for eventually appropriating a “new” version of a given digital resource – skills such as notably linguistic skills, intellectual abilities, shared values (view points), a shared baseline relevancy concerning the selection of “interesting” resources within a given set of digital resources, and finally also technical and technological, economical and other constraints. This makes the process of (semiotic) interpretation and adaptation a so complex but also exiting and concretely speaking multi-faceted activity and metier.

Now coming back to the notion of “life world” of a social actor such as a user community, this notion means that the life world has to be understood as a set of commonalities shared by the individuals and groups of individuals composing a user community such as, for instance, a learning community, a professional community, a media information exchanging and consuming community, etc. Among central commonalities for describing and understanding the life world of a user community, we have, for instance [SCH 79]:

- a common space and time, a common territory (which is not necessarily a physical one);
- common activities and goals;
- common interests and expectations;
- common linguistic resources;
- and also common knowledge and axiological schemes and resources shared by the members of a user community (cf. especially [SCH 79]).

This is a very central point in social sciences, i.e. to understand and describe such commonalities. Their operational understanding helps us to better control and to improve the “translation” processes called (re-)purposing or (re-)authoring of existing digital resources in order to adapt it to the concrete expressions of such commonalities that characterise the life world of a given user community, to envisage the explicit modelling of parts of this very complex “translation” process as well as, finally its partial automation.

Again quickly some words concerning the individual user, in this perspective! The individual user, in the sense of an *understandable behaving and doing variable* belongs to one or more of user communities. In this sense the conceptual shift from the individual user-centric to the collective user group approach in order to handle the (re-)purposing or (re-)authoring process of digital resources, does not at all exclude the individual or personal perspective of the use of digital resources! Contrarily, the collective or user community approach provides reliable means,

methodologies, tools enabling not only any (institutional or non-institutional) community to become a content provider, trader or consumer but also any individual who intends to play the one or the other or all of these three roles. Indeed, one of the most appealing perspectives consists, for instance, in the fact that any individual as well as any small groups should be able to produce its own digital heritage, to manage it and to exploit it for other user groups or individual users ...

The point here is, that the activities, tasks, strategies, procedures of producing, exploiting, preserving, etc. digital heritage objects (*lato sensu*) have to be – so to speak - “extracted” from the collective experiences composing the identity and the tradition of a social group or community such as teachers, researchers, professional workers, social workers, and so on.

In other words: the work on collective features and patterns such as the structural organisation of digital documents, genres, scenarios and other cultural models for information and knowledge processing or again (content producer, trader and consumer) community patterns and routine solution procedures, help us in order to provide the individual user with a much more richer but also more naturally and intuitively sensed digital environment.

3/ The structural organization of a digital knowledge resource

If we want to progress in our understanding of repurposing, adapting digital resources in order to transform them into genuine added value objects for a targeted user community, the central problem remains nevertheless the appropriate understanding of the *structural organisation of the digital resource itself as a (potential) meaningful object*.

Indeed, from a semiotic point of view, a digital object or resource such as a digital text, film, image, etc. can be characterised, described with respect to a set of *central features* such as the themes or topics encapsulated in a digital resource, the development of a topic through a digital object, the forms of expression of a topic, and so on. Like the grammar of a natural language, these features form together the specific identity or profile of a digital object or of a corpus of digital objects. Such a profile or identity is not an arbitrary thing but belongs to a tradition represented by genres or cultural models of information processing to which people refer and which people use in their own activities of writing, reading, exchanging information and knowledge.

In this sense, the understanding, description and indexing of digital objects is nothing else than the eliciting of the specific profile and identity of a corpus of digital objects and constitute a direct or indirect means for comparing and evaluating a given “source profile” with respect to the needs, interests, desires, etc. of a destined community or again a projected context of use.

Figure 1 shows the general structural picture of the (digital) resource understood as an information loaded sign or network of signs ([STO 99], [STO 01], [STO 03a]). In this picture, we distinguish three principal models of the semiotic approach of the (digital) resource:

1. the *compositional model* which sticks to the fact that a (digital) resource as a (more or less permanent or, contrarily, transient) meaningful whole can be decomposed in meaningful units (including physically independent (digital) resources);

2. the *layered feature model* which shows that the internal organization of a (digital) resource has to be understood basically, as already shown in Saussurian semiology, as a composition of a content and the expression of the content by the means of a percept conserved and conveyed by some physical support;

3. the *task activity model* showing the (digital) resource as an information and communication good belonging to one or more social practices.

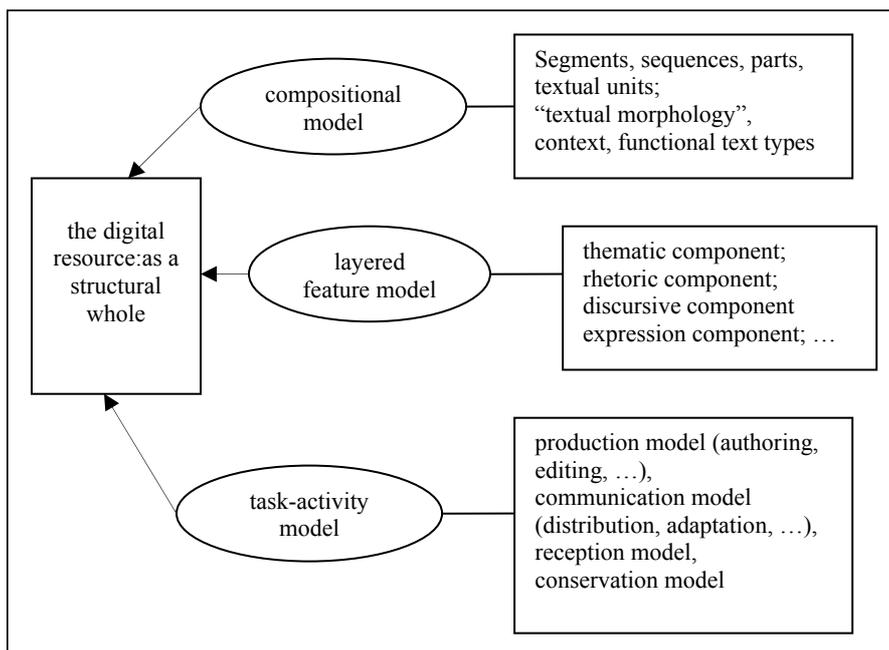


Figure 1 : *The general structural model of a digital resource in a semiotic perspective*

Based on this model, we are able to approach in a rather systematic and explicit way several of the main problems within the context of the (re-)authoring of given digital resources such as:

1. The systematic content description of a given source corpus of digital resources in order, for instance, to develop domain ontologies for the identification, description, indexing and classification of relevant parts or pieces (segments, paragraphs, ...) of the source corpus for a given target user community,
2. The eliciting of the user community's existing knowledge as well as of its expectations and needs with respect to a given source corpus of digital resources,
3. The specification of forms and genres of annotations in the sense of means and tools for enhancing, explaining, reorienting, "versioning", ... the content or parts of the content of a given source corpus,
4. The specification of publishing genres adapted to the working routines of a targeted user community as well as to its expectations and needs.

In this sense the model represented in figure 1 is really a central one for many technological but also scientific challenges within the field of (digital) document and content engineering.

4/ Digital resource (re-)authoring or repurposing

As already stated, one of the major challenges (if not the major challenge) for the constitution of a genuine knowledge community within a multilingual and multicultural world such as Europe, is the user oriented, user-centred ("subjective") exploitation of existing digital resources such as videos, sound tracks, texts, etc. produced by the concerned users themselves or other authoring instances and stored in central or distributed digital libraries. This challenge is identified under the heading of *(re-)authoring* or again *re-processing*.

(Re-)authoring of digital objects is a more or less complex (individual or collective) process by the means of which a digital object or a corpus of digital objects (such as a corpus of video clips, a

corpus of textual elements, a corpus of still images, a corpus of sound elements, etc.) are adapted, attuned to a specific context of use and, this, following a sort of *authoring* or *publishing scenario*.

In taking the example of the above introduced and shortly presented audiovisual archive in human and social sciences (AAR), this means – quickly spoken (figure 2) – that:

1. a corpus of audiovisual files containing, for instance, information about a given or requested domain of knowledge (“the immigration policy in Europe since 1996”; “the status of French language in Western Africa”; ...);
2. undergoes a series of manipulations called “re-authoring”;
3. by the means of which they can be used as a *hypermedia folder*, a *video-lexicon*, a *courseware*, a *hyper-documentary*, etc. either in formal learning situations or in a wide range of non-formal one.

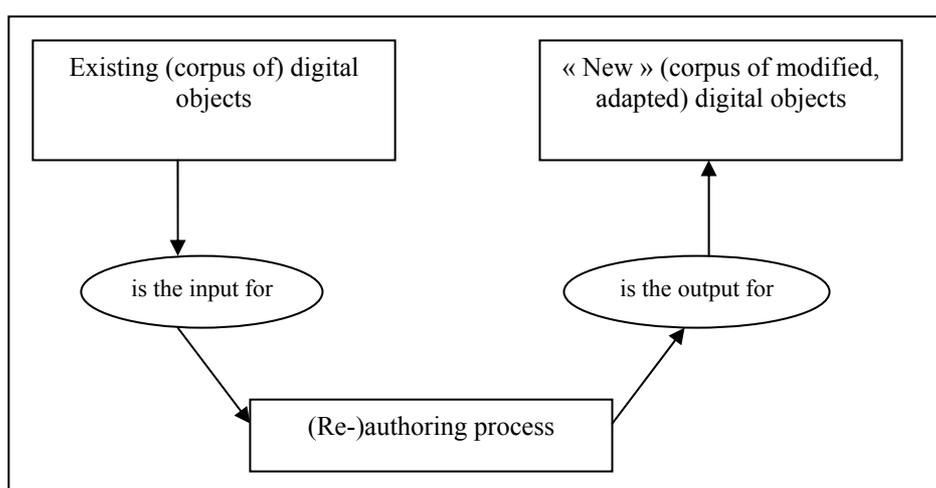


Figure 2: *The (re-)authoring process as a mediation process for (continuously and collaboratively) adapting digital objects to specific user needs*

In considering more precisely the (re-)authoring process (figure 3) itself, it has to be explained with respect to:

1. the *type and genre of (re-)authoring*: content selection in a given corpus of existing digital objects, explanation & completion of selected content; existing digital object versioning; content translation; existing digital object interlinking; creation of new content parts; (visual, sound, ...) content expression modification; etc.;
2. the specific *type and genre of digital objects*: technical and scientific texts; audiovisual ethnographic documentaries; cultural heritage images; etc.;
3. the *goals of a (re-)authoring process*: *contexts of use* (learning, teaching, science popularisation, ...); *destines* (pupils, students, any person, specific social group, ...); *publishing genre* (courses, thematic folders, glossaries, info flashes, games, ...); *forms of distribution*; etc.;
4. the *resources, means and tools of the (re-)authoring process*: *human resources* (authors, domain specialists, translators; publishers; ...); *intellectual resources* (ontologies; thesauri; publishing models; aids, hints and more developed explanations and methodologies; ...); *technical resources* (indexing tools; annotation tools; translation tools; publishing tools; ...); *economic resources* (budget, ...).

5. the *process* of (re-)authoring *itself*: the phases and tasks composing a (re-)authoring *chain* (such as “publishing genre selection and preparation”; “corpus constitution”; “description and indexing”; “translation”; etc.).

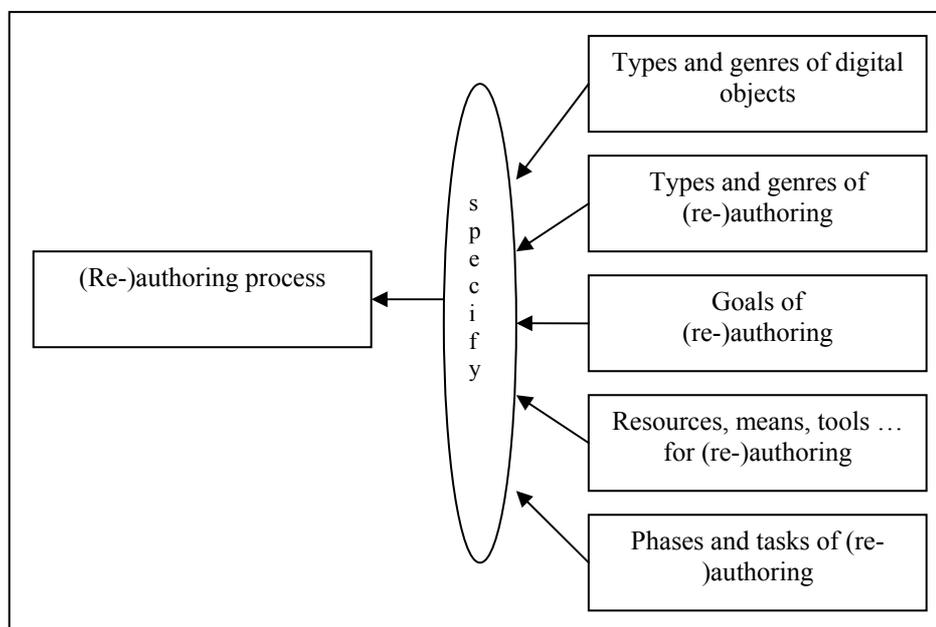


Figure 3: Central components specifying the (re-)authoring process

The (re-)authoring process, can be seen as a – in a broad sense – *cultural translation* process. Indeed, a given set or corpus of digital objects (figure 2) are meaningful entities (and hence *re-authorable*) with respect to context(s) of uses, destines, social practices, etc. for which they have been designed. In other words, they belong and refer to a *culture* embodied in the life-world of social actors, viz. user communities.

For instance, a corpus of audiovisual files in the above quoted audiovisual archive of human and social sciences are dealing with linguistic diversity from a *socio-linguistic* point of view and for a group of destine of experts in this field (“life-world”). This corpus can be (re-)authored for a high diversity of potential contexts of uses and destines: for instance as an *introductory seminar* (a “Proseminar”, in German) in philological studies, as a *complementary course* in anthropology or sociology, as a *reportage* for a wider interested public, as a set of *info flashes* for people only with local interests in this domain, as a *discovery game* for children, etc.

Each one of these examples shows (re-)authoring as a process of – so to speak – *cultural adaptation* of an initially given set of digital objects, i.e. the content, form and function of the given digital object, to the target culture determining the interests and lacks but also the uses of that object by the destine. The (re-)authoring process of a digital object can definitely be understood as the opening of the cultural specificity of a given digital object with respect to a given diversity of target cultures and the circulation of a digital object or artefact in its given cultural identity within a “multi-cultural (knowledge) space”. Linguistic translation, in this sense, is only a *specific case* of cultural translation.

In this sense, figure 8 and 9 identify also of one of the principal actions of the 7th framework of research and development of the European Community, i.e. to take into account the *linguistic and cultural diversity* that determine Europe as a knowledge society as well as a “market” for (cultural, scientific, ...) content services and resources. This is also, without any doubt, the most

central challenge of (re-)using existing digital objects for (u-)learning situations which occur as specific social practices intended and performed by social actors and with the help of specific resources within an embedding culture (professional culture, personal culture, educational culture, linguistic culture, etc.).

Let us repeat once more, the re-purposing (also “re-authoring”) is the central activity of the *intentional manipulation of the source profile, the source identity of digital objects* or corpora of digital objects in order to attune them to the expected profile of a user community and/or contexts of uses. For instance, the use of available digital objects such as videos or images for educational purposes, very often presuppose a whole series of such repurposing tasks which may more or less drastically change the content of these objects, their “look”, their purposes and goals, etc. in order to transform them into adequate resources for teachers and learners.

Technically speaking, the repurposing process is composed of a series of typical and more or less standard activities and tasks such as, for instance (cf. also [STO 03a]):

1. the constitution of “working corpora”,
2. the production or reuse of thematic grids and maps (“ontologies”) for describing and classifying digital objects or parts of them,
3. the identification and segmentation of specific parts of digital objects which are of specific relevancy for a given purpose,
4. the thematic or topic description of identified and extracted segments,
5. the key-wording of extracted segments,
6. the commenting (analysing, interpreting, evaluating, ...) of identified and extracted segments,
7. the enrichment of segments via hypermedia links to other resources,
8. the (broadly speaking) translation of identified and extracted segments,
9. the publishing of repurposed digital objects or parts of them with respect to specific publishing genres that fit to the specific requirements of the destined user communities,
10. the delivery of repurposed and published digital objects as hypermedia web sites, by the means of mobile devices, of iTV or again for digital supports such as DVD,
11. etc.

Let us stress also the point that the repurposing activity can be understood as a *centralised activity* – performed by one and only one community (in general the content production community of an archive or a library) – or as a *distributed one* (performed by any user community interacting with a digital object repository (library or archive)). It is certainly the second figure – the distributed understanding of the repurposing of digital objects – which will develop massively in the coming years.

In any case, the semiotic interpretation and adaptation of a given source corpus of digital objects understood as a repurposing or re-authoring activity is “configured” and instrumented in what we call the *semiotic atelier* (i.e. *semiotische Werkstatt*”, in German).

A “pre-figuration” of this semiotic atelier is the software application *Interview* developed since 2005 conjointly by the R&D Department of the French Institut National de l’Audiovisuel (INA) and ESCoM (Equipe Sémiotique Cognitive et Nouveaux Medias), the R&D laboratory of the Maison des Sciences de l’Homme (MSH) in Paris. Actually, and thanks to the European R&D project LOGOS as well as the French funded project SAPHIR, this environment undergoes profound changes in order to integrate many different technologies and tools in a common “authoring studio” conceived *expressis verbis* for the above presented repurposing and re-authoring objectives of digital audiovisual corpora such as those that compose the AAR program in Paris.

In order to conclude this paper, let us present only very roughly the overall purposes of the Interview tool, actually again used for repurposing and re-authoring video corpora within the AAR program.

The purpose of the “semiotic atelier” is to allow any registered user (user group) to deal following his/her interests/desires with a previously selected corpus of video files. “To deal with” (i.e. “processing”) means here especially:

- the composition of a video file corpus and its importation in the Interview environment;
- the “subjective” identification and selection of a set of for the user relevant video segments;
- the (more or less) systematic and rich description, indexing and enrichment (via comments, interlinking with other resources, creation of tests and exercises, ...) of :
 - the referential content of the segments (i.e. the object of a segment, the domain which constitutes its topic);
 - the audiovisual language of the selected segments (i.e. the expression means of a topic) ;
 - the rhetorical level of the selected segments (i.e. the manner of how a topic is viewed and developed in a segment – this is a very important level because, in the case of the audiovisual archive of human and social sciences, most of the video files contain conferences and interviews with researchers);
 - the narrative relationships between the selected segments (i.e. the possible semantic relationships between chosen audiovisual segments);
- the (broadly speaking) translation (“versioning”) of selected segments and their preparation for personalised publishing;
- and especially the publishing of the results of these authoring activities following (adaptable) publishing genres (for instance as thematic folders, as video-dictionaries, hyper-documentaries, courses, etc.) and with the help of publishing assistants.

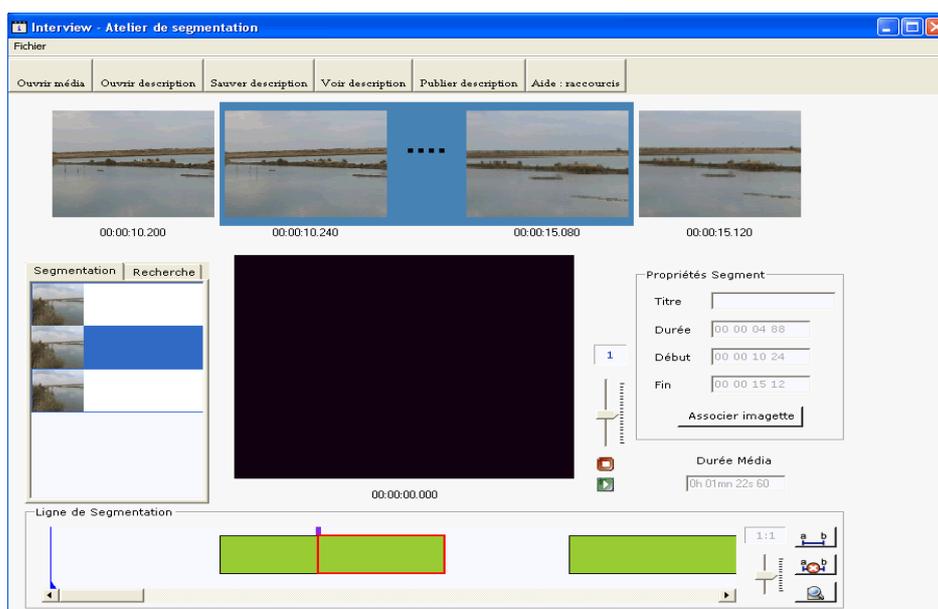


figure 4: The Interview tool and its main options

Figure 4 shows the main screen of the Interview tool with its principal functionalities:

- *Ouvrir Media* means to import a video file or a corpus of video files in the (re-)authoring environment of Interview;
- *Ouvrir Description* means to charge the xml file containing the state of the art of an already ongoing (re-)authoring project;
- *Sauver Description* means to save an ongoing (re-)authoring project;
- *Voir Description* means to start the different (re-)authoring activities (cf. below);
- *Publier Description* means to publish the results of a (re-)authoring activity in the form of a web site (actually the genre of the web site is a thematic folder but other publishing genres – video-dictionary, hyper-documentary, course, ... – are just under development);
- *Aide:raccourcis* recovers helps, hints etc. for using the Interview tool.

Following the (re-)authoring motives or needs as well as the specific publishing genre, (re-)authoring activities can be very different, build – so to speak – different *processing chains*.

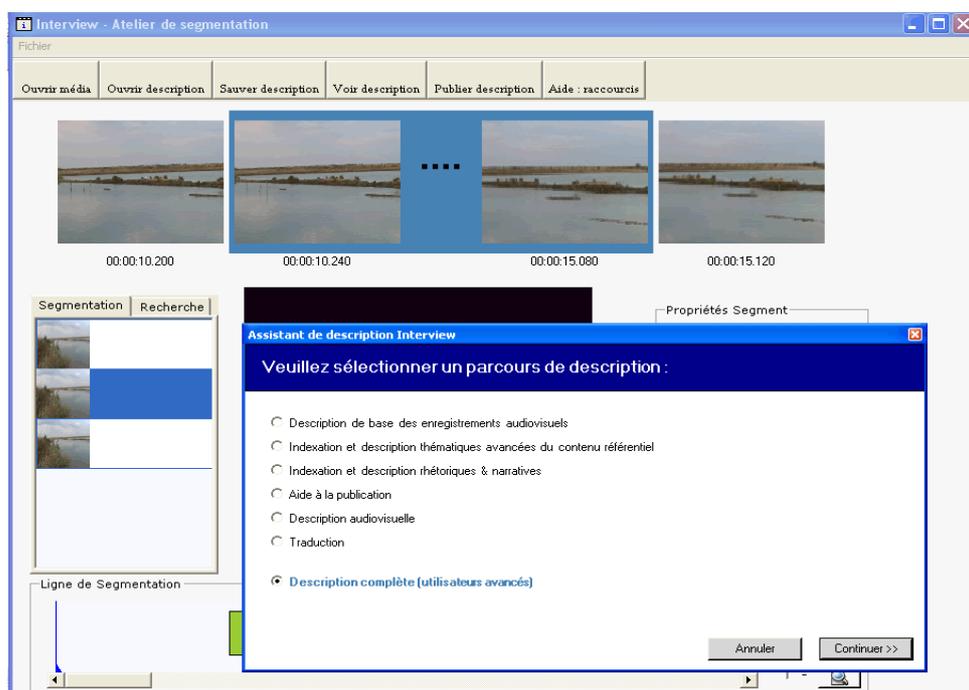


figure 5: A set of pre-defined options for processing, i.e. (re-)authoring a corpus of video files

After clicking on the “Voir Description” button, the (habilitated) user can make a choice, as figure 5 shows it, between different such processing chains. For the moment, the principal differences are between:

1. *basic processing activities* such as the identification, selection and indexing of (subjectively relevant) video segments and their publishing as an annotated corpus of video segments;
2. *more and even very sophisticated activities* including the rhetorical analysis of the content of chosen video segments, their commenting and (narrative) their re-arrangement as a new, completely virtual video;
3. and, finally, the *linguistic and cultural versioning* of video segments for adapting them to a public that does not possess neither the necessary linguistic nor the domain specific skills and competencies to understand a video file.

These three main options will be completed with other (re-)authoring activity chains attuned to previously chosen publishing scenarios (i.e. video-dictionary, hyper-documentary, thematic folder, ...).

A specific option shown in figure 5 is the “expert option”. In choosing this option, the (habilitated) user is prompted to all proposed possibilities of working on a given corpus of audiovisual files (figure 6). This means that it’s up to the user to choose the processing activities which are important for him/her and to produce his/her own processing chains. To be able to work with this option, presuppose a user who is a real specialist in semiotic processing of audiovisual documents.

Figure 6 shows the principal processing activities proposed by the Interview tool. These activities have been specified and developed with respect to a *semiotic vision* of (audiovisual) documents. This vision can be explained with respect to the five principal assumptions:

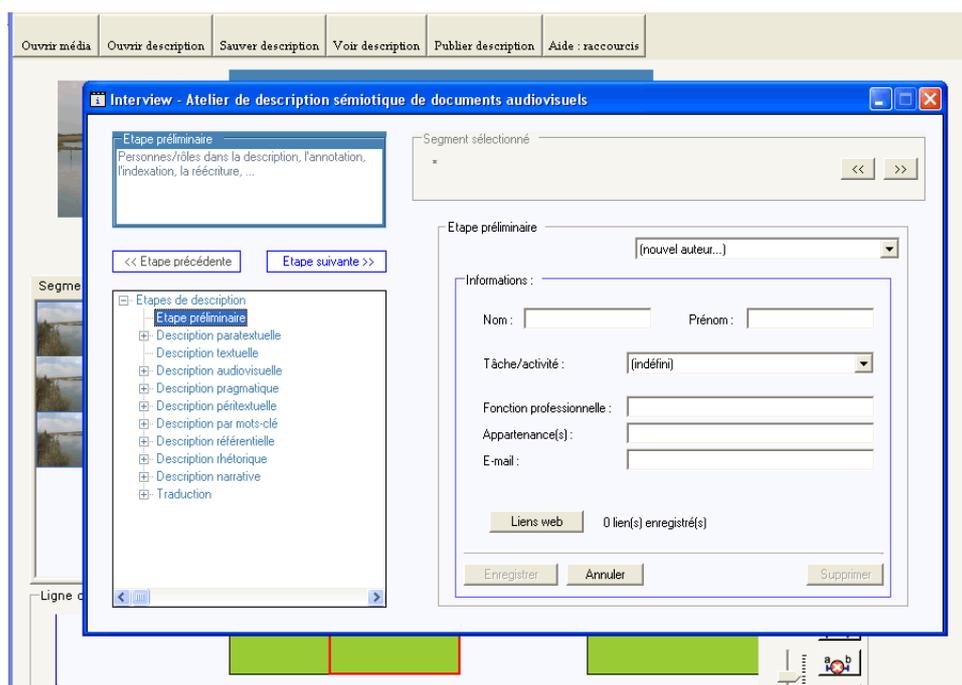


figure 6: The principal processing activities proposed by the Interview tool

1. The (audiovisual) document as a *layered entity* (this goes back to the semiotic notion of the sign as intrinsically composed of a content layer and an expression layer of the content) – both layers can be differentiated in more specialised layers ...).
2. The (audiovisual) document as a *composed entity* which can be differentiated in several entities (such as, for instance and very trivially speaking, in the *introduction*, the *development* and the *conclusion* part of a book-like document) and which itself can be a part of a network of other entities and documents (such as, for instance, an audiovisual documentary as a *functional* part of a whole hypermedia documentation of the life-world of a group of people; cf. chapter 6).
3. The (audiovisual) document as a *typified entity* (this means that any document can be (at least partially) characterised as belonging to one or more types or genres of documents (for instance, a film document on the daily life of a group of people as a descriptive documentary, a narrated documentary, a historical documentary, a fictitious reconstruction, a staged drama, etc.).
4. The (audiovisual) document as an *affordable entity*. This means that any document is provided by traces that “certifies” its real or possible uses, contexts of uses - traces that

are indications for its more or less high relevancy for specific contexts of uses, for groups of destinies, for specific objectives, etc. For instance, a fictitious reconstruction of the daily life of a group of people may be a wonderful document for specific purposes (make people aware of the cultural personality of an other group of people; make understandable the cultural differences by the means of an intrigue, ...), specific destinies (people far away from any understanding of existing cultural differences, people full up of stereotypical judgements on other habitudes and traditions, ...), specific uses

5. The (audiovisual) document as a *produced entity*. This means that any document is the result of an authoring process, a more or less formalized or, contrarily, free activity of producing and transmitting information about a given domain or object. Concerning (lato sensu speaking) pedagogical or potentially pedagogical document resources, it has to be stressed the importance of the vision (the view point) by the means of which an author (i.e. an individual person or a social group assuming the role of an author) handles a knowledge domain or object in order to explicit the author's culture with respect to the learner's cultural background (i.e. with respect to his/her knowledge and values as well as to his/her motivations and expectations as a learner).

It's not the purpose of this paper to develop in a more systematic and technical way the implications of these five assumptions. Let us only stress here the fact that they are congruent with the vision of (re-)authoring as a linguistic and intercultural processing or versioning of corpus of documents as well as with the technical system named *Interview*.

Briefly speaking, figure 6 shows a list of principal groups of processing activities which have been identified and elaborated with respect to the above quoted five principal assumptions:

1/ A *preliminary processing phase* consists in the identification and presentation of the author or the authors of the (re-)authoring activities of a corpus of already given audiovisual files.

2/ The *paratextual processing activity* refers to the description of the identity of the existing audiovisual corpus, to its so-called meaning territory. This activity recovers more specific tasks such as the overall identification of the corpus itself (title; summary; global duration; language(s); etc.); the identification and presentation of the author(s) of the files composing the corpus; if appropriate: the identification and presentation of main protagonists (scenario writers, producers, actors, ...) and, finally, copyright issues concerning on the one hand the existing audiovisual files and on the other hand the (re-)authored products as well as their translations and adaptations.

3/ The *textual description phase* consists of two principal sub-phases: a/ of the identification and delimitation of relevant audiovisual segments within the corpus of existing audiovisual files, and b/ of the basic description of these segments (title, summary, principal language, audiovisual type of the segment).

4/ The *audiovisual description phase* recovers – if necessary or relevant – the systematic description of the audiovisual expression of the content developed in a previously identified segment. Extremely speaking, each visual, sound, speech, ... plan composing a segment can be identified and described such as or in relation with the other ones. Such a description can be very useful in the constitution of a library of reusable visual, sound, speech or again syncretic (visual + sound) plans.

5/ The *pragmatic description phase* recovers a set of for (u-)learning purposes very useful sub-phases. Indeed, it classifies the content of a given audiovisual segment with respect to its potential contexts of uses (such as in formal or informal learning situations), to its possible destinies, to its possible links with a curriculum, to possible previous intellectual requirements for understanding its content (i.e. recommendations of consulting first other segments, other online resources, ...); and finally to its added value for other – especially professional – issues.

6/ The so-called *peritextual description phase* recovers all tasks concerned with the annotation, enrichment or enhancement of a given audiovisual segment – by the means of interlinking it with other resources, of producing structured comments, etc.

7/ The *key-word description phase* recovers the production of a semantically structured index by the means of on the one hand freely chosen key-words and on the other hand (if available) a constituted base-vocabulary of the domain of reference.

8/ The so-called *thematic description phase* belongs definitely to the advanced description phases of an audiovisual segment and deals with the semantically rich description of an audiovisual segment by the means of a domain ontology. This means that in order to produce a systematic and coherent semantic description of one segment, the author either has to refer to a given thematic classification of the domain of knowledge or has to produce it himself. Even if this process is, intellectually speaking, rather complicate (and, as well known, always controversial), it is very often required for a good semantic classification of the segments of a given corpus, for defining semantically motivated access points to the chosen set of segments and exploration paths through them.

9/ The *rhetorical description phase* is dedicated to the identification of the principal speech acts and discourse genres by the means of which the content in a given segment is developed.

10/ the *narrative description phase* is dedicated to the positioning of a chosen segment with respect to other segments of the given corpus of given audiovisual files. This positioning is produced by a set of pre-determined so-called narrative (temporal, causal, rhetorical, ...) relations by the means of which the author is able to produce a truly new narrative, a truly new audiovisual document out of a corpus of given audiovisual files.

11/ Finally, the *translation/adaptation phase* recovers different sub-tasks divided on the one hand in more linguistic oriented activities (i.e. the translation of the content of a segment from the source language in a target language) and on the other hand more cultural oriented activities (i.e. the adaptation of the content of a segment to the cultural background of a target destine or user group). This means that there are different possibilities to perform these tasks: summarizing of the content; meta-linguistic translation by the means of the freely chosen key words or a controlled vocabulary; more or less free versioning of a given segment, etc.

These eleven phases recover the (re-)authoring process of existing audiovisual files within the Interview working environment. Naturally, this does not mean that a concrete (re-)authoring process has to pass through all these phases and sub-phases. As already mentioned, there are different –more or less basic or, contrarily, more or less advanced options proposed by the Interview environment for re-processing given digital audiovisual documents.

In this sense, Interview is a good example of a working environment based on a systematic approach of the (audiovisual) document and its uses, which is needed for a totally user-centred (re-)authoring of existing digital objects enabling *any habilitated user* to transform them in – *from the habilitated user's point of view !* – relevant (broadly speaking) learning resources. What has to be stressed here is that:

- the *criterion of relevancy* (i.e. of the belonging of a given digital object to the “class” of relevant learning resources) is determined by the habilitated user him/herself;
- the *learning qualification attribute* has to be understood in the intuitive sense given to this term, once more again, by the user.

BIBLIOGRAPHY

[BOU 79] BOURDIEU, P., *La distinction*. Paris, Editions de Minuit, 1979.

[CHA 02] CHARTIER, R., Histoire du livre et textualité électronique. Entretien dans le cadre du Programme des Archives audiovisuelles de la recherche en sciences humaines et sociales, Paris ESCoM-MSH 2002 (entretien en ligne sur le site <http://semioweb.msh-paris.fr/AAR>)

[GRE 83] GREIMAS, A.J. *Zur aktuellen Lage der Semiotik*. Ein Gespräch mit Peter Stockinger, in : *Zeitschrift für Semiotik* 1983,

[SCH 79] SCHÜTZ, A. et LUCKMANN, T., *Strukturen der Lebenswelt I*, Frankfurt/Main, Suhrkamp Verlag, 1979.

[STO 92] STOCKINGER, P., *La conception-réalisation d'une application multimédia*, Support en ligne d'une conférence, Paris, site ESCoM (<http://www.semionet.com>) 1992

[STO 94] STOCKINGER, P. *Le Kiosque d'Information*, Présentation en ligne d'un scénario pour une application multimédia interactive, Paris, site ESCoM (<http://www.semionet.com>) 1994

[STO 95] STOCKINGER, P. *Le Dictionnaire encyclopédique d'histoire Murre de Larousse*, Présentation d'un scénario pour une encyclopédie hypermédia, Paris, site ESCoM (<http://www.semionet.com>) 1995

[STO 99] STOCKINGER, P., *Les nouveaux produits d'information. Conception et sémiotique du document*, Paris, Hermes Science Publications, 1999.

[STO 00] STOCKINGER, P., et Charlotte Nikitenko (éds), *La publication en ligne*, Paris, Hermes Science Publications, 2000.

[STO 01] STOCKINGER, P., *Traitement et contrôle de l'information*, Paris, Hermes Science Publications, 2001.

[STO 03a] STOCKINGER, P., *Le document audiovisuel*, Paris, Hermes Science Publications, 2003.

[STO 03b] STOCKINGER, P. (éd.), *Portails et laboratoires pour l'enseignement et la recherche*, 2003.

[STO 05] STOCKINGER, P., *Les sites web. Procédures de description, d'évaluation comparative et de conception*, Paris, Hermes Science Publications, 2005.

[STO 07] STOCKINGER, P., La republication de corpus audiovisuels numériques dans un contexte multilingue et multiculturel, in : M. Oustinoff and J. Novicki (eds), *Traduction et communication*, Hermes 49, Paris, CNRS Publications, 2007.