# Intention Based Authoring Process from Audiovisual Resources

Ludovic Gaillard, Jocelyne Nanard, Bruno Bachimont, Louis Chamming'S

▶ **To cite this version:**

Ludovic Gaillard, Jocelyne Nanard, Bruno Bachimont, Louis Chamming'S. Intention Based Authoring Process from Audiovisual Resources. RR07008, 2007, pp.10. lirmm-00138194

**HAL Id: lirmm-00138194**

**https://hal-lirmm.ccsd.cnrs.fr/lirmm-00138194**

Submitted on 23 Mar 2007

**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.

# Intentions based authoring process from audiovisual resources

Ludovic Gaillard
Heudiasyc
UMR CNRS 6599,
Université de Technologie de Compiègne
P 20529,
60200 Compiègne, France
ludovic.gaillard@utc.fr

INA
Direction de la Recherche et de l'Expérimentation,
4 avenue de l'Europe,
94366 Bry-sur-Marne, France
lgaillard@ina.fr

Jocelyne Nanard
LIRMM
161 rue Ada
34392 Montpellier cedex 5, France
Tel. (33) 4 67 41 85 16
jnanard@lirmm.fr

Bruno Bachimont
Heudiasyc
UMR CNRS 6599,
Université de Technologie de Compiègne
P 20529,
60200 Compiègne, France
bruno.bachimont@utc.fr

INA
Direction de la Recherche et de l'Expérimentation,
4 avenue de l'Europe,
94366 Bry-sur-Marne, France
bbachimont@ina.fr

Louis Chamming's
INA
Direction de la Recherche et de l'Expérimentation,
4 avenue de l'Europe,
94366 Bry-sur-Marne, France
lchammings@ina.fr

## ABSTRACT
This paper introduces a novel approach to assist the author during the production process of new audiovisual documents from audiovisual archives. Our research relies on rhetorical tradition to relate author's intentions to forms and contents of target audiovisual discourse. We present the different levels of structure which are involved in the authoring process, together with their associated components (topic structure, intentional discourse structure, concrete discourse structure, discursive and narrative components). The paper specially focuses on reuse of audiovisual resources and shows how the context of their indexation can be articulated with the new publication context, so emphasizing the interest of multiple discursive structures interaction. We illustrate our approach with an example of TV news publication based on political TV archives.

## Categories and Subject Descriptors
I.7 [Document and Text Processing]: Document Preparation – Languages and Systems; H.5.4 [Hypertext/Hypermedia]: Architectures;

## General Terms
Documentation, Design, Languages.

## Keywords
Rhetoric, Audiovisual, Intention, Structure, Genre, Transform.

## 1. INTRODUCTION
Major audiovisual archive repositories, such as INA (Institut National de l'Audiovisuel), RAI, etc. have two main missions: preserving the national audiovisual heritage and making it readily and widely available. Digitizing the audiovisual corpus is no longer a hot topic. The important problem is now to add value to digitized archives. The valorization policy first targeted the scientific and professional fields with specific technologies. Internet offers new possibilities to improve access to archives, specially regarding new contexts of publication (education, culture, family life). New contexts mean also new kinds of publication for new audiences, leading to build completely new documents by combining elements extracted from the archives. However, most of traditional authors have more or less knowledge about the topic of discourse of their intended document but are totally disabled insofar as the audiovisual authoring process is concerned. *The purpose of our work is to assist such an author (novice to the authoring process, but expert in some topic) to write multimedia and audiovisual documents.* For example, a teacher would like to prepare a lesson based on audiovisual archives regarding a particular topic and related fields. This author should be helped not for selecting archive documents on a relevant topic but on how to organize them in a way to publish *summaries*, *syntheses* and *elaborations*. In this paper, we focus on two points: *discourse structures* and *resource recontextualization*.

*Discourse structures*. Assistance to discursive creation aims at relieving the author of some stages and to enable him to focus himself on his writing work and his skills. The author should be able to handle familiar concepts about the writing process but should not have to see the complexity inherent in audiovisual temporal object manipulating. These concepts relate to the traditional structures present in all discursive creations including textual ones. These structures are independent of publication devices and are generally implicit for the author. They relate to concepts traditionally handled by rhetoric such as narration and argumentation. Though these structures are distinct, they are mixed during the writing process and ensure the global consistency of the discourse. More precisely, beside their informative goal, audio-visual documents traditionally conform to stereotypes and follow specific building rules (as is the case for cinema [Arijon 1993]).

We focus on producing audiovisual documents that are primarily able to provide information rather than designed for their aesthetic features. The produced documents will conform to stereotypes classically associated to some genre[1].

[1] Conventional definitions tend to view a genre as constituting a particular set of conventions of content and form, which are shared instances of the genre [Chandler 1997]. A genre is

*Thus, one of the goal of our work is to clarify the various levels of structure and to articulate them in order to guarantee to the author a coherent audio-visual discourse regarding the appropriate stereotypes associated with the intended document.*

*Resource (re)contextualization*. Developing publications for different situations (new public, new supports) from components selected in audiovisual archives imply to build description models of target documents which are potentially different from the documentary models used for guiding the indexing of archives. Sometimes, the descriptors used during analysis of archives are re-used directly for producing the target documents. This kind of approach presupposes that the initial context of analysis is the same as the final context of publication. However, when reusing components, it seems important to distinguish context-free components from those that are not. Consequently, separating the indexing context from publishing context allows a safe re-use of components and an evolution of publishing models without having to modify documentary models used for describing archive components.

*Thus, another goal of our work is to question the criteria and mechanisms of selection suitable for the audio-visual form allowing a possible diversion and re-use of the descriptors of the indexing documentary model.*

This paper is position paper on a rhetorical approach for implementing the audiovisual authoring process. We first briefly analyse related works on recent multimedia authoring approaches and present the specificity of our works. We then explore the rhetorical basis of our approach and introduce the different kinds of semantic structures we use in our audiovisual process. Next, we summarize the main principles we use to index the audiovisual resources. Finally, we detail how an author can build the various semantic structures of discourse with a TV news report example. The paper concludes with the state of implementation of our audiovisual process and details future works.

# 2. STATE OF THE ART

During recent years, research works on author assistance in multimedia production from resources mainly focused on automatic generation of multimedia documents based on structured answers to the user's queries. Most advanced works now consider that the author's assistance could be achieved through a progressive elaboration of discourse.

We now discuss each of these two approaches by presenting typical projects and finally explain the specificity of our approach.

## 2.1. Structured answers

In order to assist the author, structured answers can be automatically generated from the author's query, based on the modelling of various discursive structures. We present some of the projects which follow this approach, analyse their contributions and lacks compared to our approach.

DISC [Geurts 2004], after a user's query, generates a coherent discourse on the basis of the user's expertise in the field, according to narrative models of the topic. Resources are mainly textual and the topic concerns impressionist painters of RijkMuseum. Models of target documents are organized by sequencing Narrative Units. These units make it possible to arrange resources selected according to the role they play in the original story (*MainCharacter*, *Spouse*...). They use an "actant " ontology and relations of narrative progression to obtain a structured progression of discourse which ensures continuity and coherence of the different parts of the target story. This project is neither interested in the form of resources nor in their selection mechanism because it uses only texts and images. Moreover, discursive relations refer only to the story and never to any kind of argumentative coherence.

The project Vox populi [Bocconi 2005] is mainly based on an argumentative modeling applied to a purely audio-visual corpus (a hundred interviews which capture the point of view of American citizens after the events of September 11, 2001). A user's query, by choosing a point of view, generates a semantic graph which is built dynamically starting from annotations of topic as well as of provided arguments. The quality of the project is based on argumentative coherence of the obtained discourses. Inferences of argumentation are based on the Toulmin's model [Toulmin 1984]. Like in DISC, the audio-visual form processing is also relatively weak (except few rules of continuity). Contexts of indexing and of publication are exactly the same: extracts of interviews are used to generate contiguous assembly of news interviews.

The project PEACH [Rocchi 2003] consists in automatically generating a video sequence of images in order to illustrate an audio comment for museum visits. The approach is based on the use of descriptors expressed according to RST in order to organize a coherent narrative discourse. Based on discourse structure of the commentary and shot segmentation ,their engine apply relevant transition effects among shots.

Whereas these different works are particularly instructive for our approach, they completely erase the figure of the author. The author, in these systems, is only a model builder or a person able to parameter such different publishing models. So, these systems are more oriented toward building structured answer than toward ensuring the author's assistance. Furthermore, none of them consider the discourse simultaneously under the narrative, argumentative and form aspects.

## 2.2. Author's assistance

Among the most advanced works on author's assistance to multimedia production are those from CWI, within the framework of the Sample project [Falkovych 2003]. This project is interested in the various stages of construction of discourse by an author, from exploration of the topic until the final production. The resource corpus and the audience of publication are similar to those of the DISC project (Textual and visual resources about painters). The most challenging point of the project is the consideration of the notion of genre and of associated structures. Sample proposes an articulation between a narrative discursive structure, discursive functions and structural elements [Falkovych 2006]. The preliminary stage of our work is strongly inspired by these proposals. But Sample does not question the form of resources and its consequences in terms of selection rules. Another limit concerns the use of discursive functions. In Sample, the discursive function associated to a component is constrained to be the same in the indexed resources as well as in the target document.

However, generally speaking, the discursive function associated to a component depends largely on the role it plays in the

considered as an abstract concept, which is associated with a particular expectation on the part of potential readers, which, if taken into account, serves to increase the efficiency of communication [Fowler 1989].

context of some description (indexing) and it may change depending on the kind of publication where the component is reused. (de-contextualization / re-contextualization). The topic doesn't change but the discursive purpose of a component may vary within the context of use. Our approach precisely deals with the problem of resource re-contextualization.

### 2.3. Our approach

Our approach is centred on assisting an author who is beginner in audio-visual creation. With this intention in mind, we propose first to index the initial resources according to multiple points of view (set of themes, discursive functions, form). Insofar as we are working with audio-visual objects, we distinguish the sound layer, which acts as the audiovisual discourse utterance, from the visual layer which acts as a mode of significance (discourse utterance versus what is shown). Such a precision in the description enables us to consider various kinds of publications and to re-use resources in new contexts, independently from the original descriptors (see figure 1). The author is assisted in his authoring thanks to different structures and rules of selection. Each level of structure deals with a specific activity and particular processing of the descriptors resulting from the indexing. Our work first aims at articulating the indexing models and the publishing models, as well as to maintain independence of their respective associated structures. Second, it aims at providing correlation mechanisms thanks to selection and transformation rules allowing the mapping between the various structures of discourse. These two principles make it possible to consider the creation of new kinds of audiovisual documents by re-using partly or totally some structures and some sets of rules.

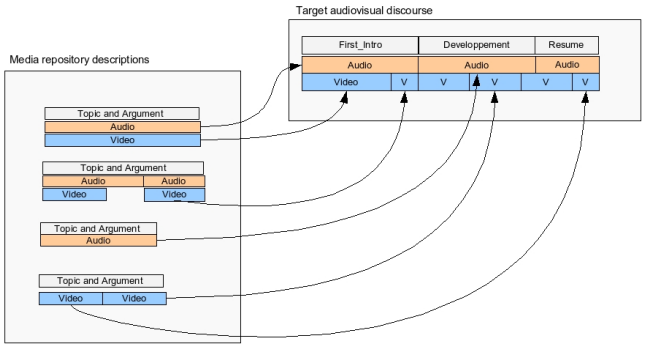


**Figure 1. Reusing resources in new context.**

## 3. DISCURSIVE PRACTICE

Our work tries to find a methodology of assistance to the creation of new discourse based on a corpus of audio-visual documents. Rhetoric provides us a theoretical and practical framework of thinking. We briefly introduce the rhetorical tradition and discuss it according to our needs in order to extract and apply rhetoric knowledge and methods to our work.

### 3.1. Rhetoric: a framework of thinking

The rhetoric of Aristotle [Aristote 1967] is regarded as the founder handbook of the art of public speaking. It defines the various components and stages to which the apprentice-orator will face. This work defines rhetoric as a practice, a *praxis*, and not as a theoretic analysis. It is used to accompany the orator in his creation of discourse from search for arguments until the speech in public. Traditional rhetoric comprises four phases [Reboul 2001]:

- *inventio*: Rhetoric is a method to help to communicate topic to others. In order to do it effectively, an orator must first be able to collect proper arguments that support his thesis. Thus, inventio is the systematic discovery of arguments, the research by the orator of all arguments and other means of persuasion relating to the topic of its discourse.
- *dispositio*: Once relevant arguments are selected, the orator should arrange them. The internal organization of discourse, as well as the argumentative and narrative coherence, come from the arguments order. This stage produces the desired plan of discourse which acts as an intentional draft.
- *elocutio*: Although today the translation of the word "elocution" connotes an oral speech, for the rhetorical tradition, it corresponds to the stage of the effective discourse writing. This stage also known as stylistics must not be reduced to simple style figures because it relates also to the final selection and incarnation of arguments in sentences resulting from previous stage.
- *actio*: Actio corresponds to the actual pronunciation of the speech, with associated effects of voice, mimicry and gesture.

Rhetoric is often wrongly restricted to one or the other of its components. Neophyte people tend to limit rhetoric only to the famous stylistic figures. Contemporary research on theses fields tends to minimize the performative aspect and the production process, so focusing only on rhetoric as an analysis method. For instance, Perelman [Perelman 1977] studied only the argumentative dimension. The works of Mann and Thompson [Mann 1988] on Rhetorical Structure Theory (RST) are particularly fruitful insofar as they provide means for a precise analysis of discursive coherence. However such works are too restrictive for the global vision of discourse creation which interests us in rhetoric.

We need to consider discursive practice in rhetoric in order to exhibit the deep discourse structures that we intend to use for assisting the author writing process.

According to this point of view, studies by the historian Carruthers [Carruthers 2002] carry out a rich lesson. By studying the practices of memorizing and writing during the medieval period, she detailed the activity of writing. Indeed she differentiates the *res* of a text from its *dictamen*. Res relates to the intention of the text, as reflected by its ideal organization elaborated during the inventio and dispositio stages. Res, also called *forma tractatus,* corresponds to the information material organized in a fully coherent textual argumentation. The result can be worked and refined, during the *elocutio* stage, in a *dictamen,* also called *forma tractandi*. We do not further detail here Carruther's studies but keep in mind the important distinction between the intentional dimension of discourse and its concrete dimension.

### 3.2. Various structures of discourse

The main stages of rhetoric use different structures, which work either at a purely logical and intentional level, or on arrangement of available relevant material. In order to assist the writing process according to the rhetoric stages, we propose to use the following structures:

- *Topic structure*: it is the place of exploration of topic and the associated concepts. The author conceptually organizes

his structure according to the set of themes under consideration. Like in inventio, the topic structure reflects the selection of topic elements likely to legitimate the matter.

- *Intentional discourse structure*: it is the place of logical organization of the previous topic structure. Various elements are organized in terms of roles (argumentative component), possibly positioned sequentially in the flow of the intended discourse (narrative component). We detail later a number of differences between argumentative and discursive functions. Like in *dispositio*, the discourse is still abstract, in terms of intentions. The intentional discourse structure can be seen as a reasoned arrangement of the various desired arguments.
- *Concrete discourse structure*: it results from the instantiation of the intentional discourse structure according to the available resources. The choice of relevant materials is guided by selection rules. This mechanism will be partially detailed in section 6. The author works directly with resources like in the *elocutio* stage. Then compromises are necessary because the wished intentional structure could be not exactly satisfied. Elements of the issue can be missing, some discursive functions cannot fit, etc. A lot of stylistic modifications are possible according to the nature of medias (in our case, mainly temporal arrangements).

The *intentional discourse structure* is relatively stable, whereas discursive incarnations can be various and in perpetual re-handling.

In the following, we show on an example how to take advantage of these different levels of discourse structure and how to articulate them in order to assist the author's writing process from the indexing until the final production of audiovisual documents.

# 4. Elements of the indexing model

Whereas our focus in this paper is not on indexing, we provide now a brief presentation of the main principles of our indexing model in order to give a better understanding of the articulation with our publishing models.

## 4.1. Separating topic from showing

We describe the topic domain with an ontology. According to our corpus, we consider that the thematic significance is mainly related to the verbal information contained in the audio channel. So, most of the time, domain indexing consists in reformulating the spoken content by using ontology terms. Regarding the visual layer, it is necessary to specifically describe "what is shown" by the different shots, which can be different from, or even not related to the audio thematic content. Generally speaking, audiovisual montage is based on this articulation between "what is shown" and "what is said".

## 4.2. Regarding audio and visual forms

At the lowest level, we differentiate the audio from the visual layers. Research in film analysis has tried to identify the minimal audiovisual description unit. The resulting so called ``audiovisual grammar'' would be universal for describing any audiovisual document. For Metz [Metz 1969], the elementary unit is a shot. Metz argues that the film sequence temporal organization is based on this minimal element. The *Grand Syntagmatic* points and classifies the narrative segments of film language in terms of shot sequences, called *syntagmas*. *Syntagmas* express the spatial, temporal and logical connections which constitute the universe of the *fibula* [Eco2001]. Metz's study is too related to the syntagmatic structure level to take into account the most elementary narrative level, which is the shot.

At the opposite, our purpose is to describe the specific narrative level of the shot itself, mainly by differentiating the audio and visual layers.

**Semiotic functions.** We describe the semiotic nature of each layer (audio, video, textual, graphical, etc.) and shooting form attributes (camera movement, scale shot, etc.) [Zettl 1990]. Let us observe that this description level is context and interpretation free.

**Narrative functions.** We can also help the indexing process by using a more interpretative approach. Related to the thematic content, a narrative approach makes it possible to define *general narrative functions* or *thematic roles*, which organize the thematic discourse coherence and the "fabula" development (*see* Greimas theory [Greimas 1966]. In this view, Stockinger's work provides a relevant example of application in the audiovisual field [Stockinger 2003]. There are some particular narrative functions, like preceding greimassian roles (*actants*), which can characterize a specified genre and reciprocally are pre-formatted and determined by the genre itself. For example, in a TV news report, they can be defined, according to the audio layer, as : *newscaster, journalist, expert, quidam, spokesperson, representative*, etc.

**Discursive functions.** The role of discursive functions is to bring out the main argumentative forms in charge of the discourse construction. These functions are dependent of the discursive structure context.

In our work, we use the analysis and description of these different indexing structure levels which are mainly the result of ISA[2] research about the semiotic [Chamming's 2003], narrative, and discursive description of audiovisual objects.

# 5. Typical author's activities and engagement

In this section, in order to characterize the kind of needed assistance, we present three typical activities of an author engaged in producing a target audiovisual document. They show a continuum in the degree of author's engagement.

- *Summary* consists in a simple contiguous editing of audiovisual resources about a specific topic. It uses only extracts without any recombination of audio and visual layers. The most basic case doesn't even change the extracts order. This kind of document is typically close to *structured answers*. The author only selects the extracted segments.
- *Synthesis* organizes a range of different points of view on the same main subject, in a coherent discourse flow. The author tries to respect the document points of view and doesn't intend to make any personal judgments. For example, she uses interviews with witnesses to exemplify the subject and interviews with experts to develop her purpose.
- *Elaboration* leads to a new well organized discourse which aims at persuading, and not only at presenting. The author is committed in the result and wishes to change the future watcher's point of view. During this activity, the author

[2] ISA (*Interprétation Sémiotique de l'Audiovisuel*) is a team of the Research Department of INA.

adds comments to selected resources in order to compare and to criticize them. Many new resources are created by the author (texts and audio comments). They are mixed later with archives segments to build a coherent audiovisual discourse. This kind of publication may be a simple linear audiovisual discourse as well as a hypermedia interactive product.

Insofar as we want, in this paper, to illustrate the interest of using the various levels of discourse structure to assist the author in building a new coherent audio-visual discourse from resources, let us point out the reverse relationship between author engagement and resource usage.

- **Increasing of author engagement.** Going from *summary* to *elaboration*, the author engagement is increasing. In the first kind of publication, the author is a simple user. At the opposite, in elaboration, the author is a real writer, giving his opinion on the topic and arguing about it. In a synthesis, the author engagement arises from the various levels of structures to combine.
- **Decreasing of archive resource usage**. Going again from *summary* to *elaboration*, the resource usage is decreasing. In the first kind of publication, most of discourse materials come from archives with very few new elements added by the author. On the contrary, in the \textit{elaboration, most of the discourse results from personal contribution of the author, and archives are then more considered as quotations. In a synthesis, the author is more engaged in resource re-contextualization from rhetoric and semiotic points of view, which is important for discourse coherence, than in selecting a huge quantity of resources.

Consequently, in order to keep our example simple, we choose a synthesis activity to detail how the author can build a target audiovisual discourse in the manner of a TV news report. As we show it later, this case study allows us to illustrate the different levels of structures previously presented. The elaboration activity would only add new material which may not be in the initial resources.

# 6. An example: TV news report synthesis

Our corpus is composed of morning, midday and evening bulletins of French TV news (TF1, France 2 and M6) during a three days period. The sample period is voluntarily narrow in order to keep a homogeneous set of themes. The main subject during this period was about the political reformation of *Contrat Première Embauche* (CPE). We extracted a subset of 17 news reports relating to this subject. We suppose that this corpus is indexed according to the proprties described in section 4

In this section, we explain the different steps followed by an author to articulate the use of the different levels of structures and of their components. We discuss later which tools we use to support the process.

The authoring process is based on three distinct steps, which are related to the specific discourse structures, as previously described in section 3. The author should first establish the topic structure in order to specify which kind of material should be used in the target document, according to the topic of discourse. Second, he builds the *intentional structure,* which organizes topic elements according to their role in a narrative structure. Finally he fills up this structure with the relevant resources and assembles them in an appropriate manner. We call this final structure the *concrete structure* to highlight the material content component.

## 6.1. Building the topic structure

The author has a general idea about the topic of his discourse. He should first explore a domain to select every topic related to his subject. Domain ontological modelling helps exploring conceptual material. In order to increase his domain knowledge, the author can specialize or generalize a specific concept. During topic exploration, he manually selects relevant topics and elaborates a conceptual topic draft. This approach is close to *conceptual exploration stage* used in the Sample project [Falkovych 2003]. According to Falkovych's comments, exploration of large topic repository is a time-consuming process for the logical inferences, and a friendly and readable user interface. See Figure 2 for an example of top level conceptual graph query on *French Prime Minister D. de Villepin about political reformation during CPE crisis.*

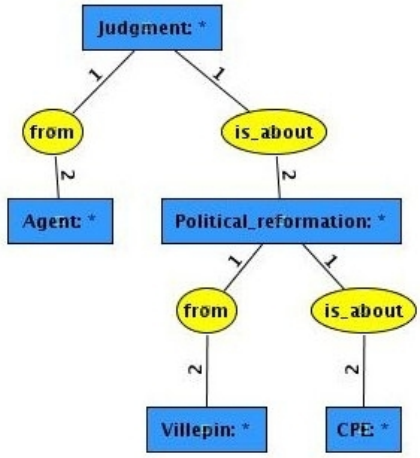


**Figure 2. A query example expressed as a conceptual graph.**

This query aims at selecting material which illustrates some agent judgments on the CPE politics by Villepin. Figure 3 illustrates schematically the possible results of such a query as a set of resources providing judgment by some relevant agents, which are *the opposition*, *the poll opinion* and *the majority.* Different kinds of judgment are related to them. In a recursive way, the author could decide to apply a query on specific results to discover related topics.

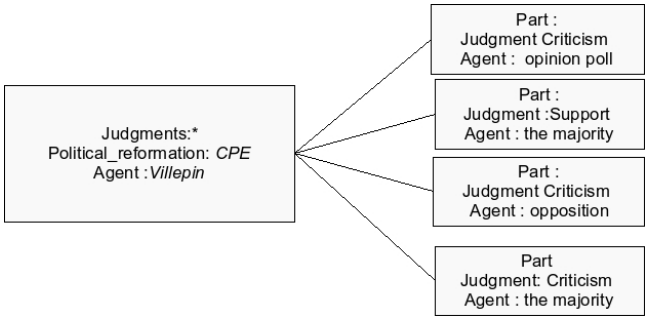


**Figure 3. Schematic results of the query of figure 2.**

During this step, we use a conceptual graph plate-form [Chein 1992]. Approximated query mechanisms are in progress to allow

the author to rank his more relevant element types (*agent*, *localization*, *action*, *object*...). In this way, the topic exploration step results in a collection of several conceptual graph queries elaborated by the author.

This collection of queries defines the topic structure in an intentional way. Query execution actually gathers from indexed resources the corresponding material that will be potentially used to populate the target document. However, let us observe that this material does not embody any intentional discourse structure related to the target document.

## 6.2. Building the intentional structure

After he defined the topic structure, the author selects an appropriate genre of publication to apply to. We detail now some genre components, which compose the *intentional structure.* This structure specifies a set of structural elements able to carry the author's intended discourse. The elements are organized thanks to some chosen *discursive functions* and to a *specific narrative positioning*.

### *6.2.1.Discursive functions*

Similarly to text, audiovisual discourse relies *on general discursive functions* to define the structural organization between elements according to the role they play in the discourse. In addition, audiovisual discourse uses *discursive functions specific to the audio and visual forms.*

Furthermore, insofar as we are working with audiovisual material, it is also necessary to decide which of the verbal content or the visual forms plays the most important role in relation to discursive functions used to select the actual material in the topic structure.

**General discursive functions.** They are relations which provide role to structural elements. Thus, within a *part* of TV news report summarization, an element could respectively be used *to present, to justify, to develop, to summarize, to give background, to disagree, to contextualize...*

Most of theses functions proceed from Rhetorical Structure Theory. As seen before, RST is more an analysis theory than a practice. Here, we use general discursive functions in a constructive manner to specify how the author wants to organize the elements of his discourse. See figure 4.

**Specific audio and visual discursive functions.** In addition to the classic vocabulary used in RST, we introduce specific relations due to the simultaneous usage of audio and visual forms. For instance, a visual element can respectively *illustrate*, *suggest or symbolize* the spoken statement contained in the sound track. Each function is associated to a specific pattern of query, able to select among resources the best elements which satisfy the considered function, according to criteria expressed in the query. For example, the *suggest* function is more precisely about selecting a concept such as *agent* than is the *illustrate* function. On the contrary, the *suggest* function is more related to selecting other concepts like *localization, period*, *etc*. For this phase of association between a discursive function and concept selection by a query, domain logical inference is very useful to select the relevant concepts.

**Leading media.** Each audiovisual production is usually semantically staged on one leading media. Depending on the chosen genre, the leading media can be respectively the audio or the video track. For example, *interviews* and much more *documentaries* use audio as the leading media, whereas *live reports* are rather based on video elements, with audio comments. Our modelling of genre takes into account this important aspect of audiovisual production by introducing form selection constraints related to the discursive functions.

Figure 4 gives an example of discursive functions involved in our selected genre and their associated form selection constraints. The structural elements are organized thanks to discourse functions (*justify*, *present*), which constraint future form selection (*Audio*, *Visual* or *Audiovisual*).

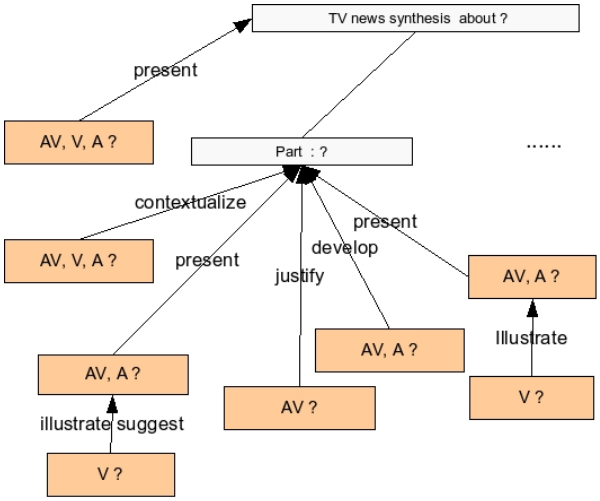


**Figure 4. Example of discursive functions allowed by genre.**

### *6.2.2.Narrative positioning*

The previous structure can be considered as a logical tree, with no information about discourse flow order. The author doesn't know yet how he will order the elements. For our example, he could decide that the flow must begin with a statement made by a journalist's point of view, followed by a *background presentation* and some *justification* or *development* and end with an *exposition*. The narrative model details the cardinality and the sequential order, which are allowed, in terms of number of elements selected by a discursive function and their relative positioning (s*uccession* and *precedence).* Two points deserve consideration during the process of building the narrative model

### *6.2.3.The resulting intentional structure*

**Selecting forms.** The position of the elements in the discourse structure influences the content selection in the resources but also selection of specific forms. For instance, an element used to present a topic, placed at the beginning of a part by the author, usually means an introduction of this part. The traditional montage pattern for an "introduction" uses audio comments from the journalist, illustrated by some visual elements related to the topic. The same element placed differently in the discourse structure will have the same function but a different pattern of visual audio montage. We will detail further this example with a difference between the current part *introduction* and the *first part introduction*. in section 6.3.

*Topic selection* organizes the topic domain, which the author would like to talk about. *Discursive functions* and *narrative positioning* detail the role and the order of structural elements in the discourse flow. The resulting *Intentional discourse structure* is the crossing of these two components : a topic domain put into a discourse flow. Figure 5 shows an example of this structure.. The discourse structure is always in terms of

intentions because no physical resources are required. The author has only specified his wishes and his intentions on how to process the topic. The nodes of discourse structure contain the appropriate queries and not yet the concrete relevant resources.

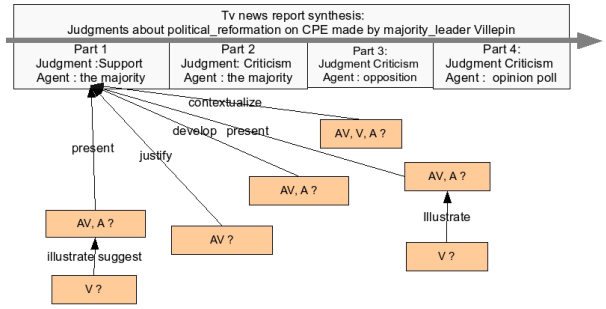


**Figure 5. Intentional discourse structure.**

## 6.3. Building the concrete discourse

The *intentional discourse structure* must be filled with the relevant resources, which will be assembled in an appropriate manner depending on genre. Going from *intentions* to *resources* imply selection mechanisms, which must take into account a set of topics organized with discursive functions and narrative positions of structural elements. This selection mechanisms are based on declarative rules specified in a specific language called TTL which allows complex element targeting and pattern matching, in order to reuse resources in new publishing context. The final structure is called the *concrete structure* to underline the material content component.

matching, in order to reuse resources in new publishing context. The final structure is called the *concrete structure* to underline the material content component.

We distinguish two steps in the instantiation process:

1. Potential candidate selection regarding domain query.
2. Evaluation of the most relevant candidate from preceding selection depending on narrative position and discursive functions.

### *6.3.1.Selecting sets of potential candidate*

Domain query is solved using logical inference as explained in section 6.1**.** Each node of *intentional discourse structure* is instantiated. We plans to used approximation mechanism of evaluation to rank the relevant domain value. Then, the author can keep the query in the same state, rank particulars materials or compel a specific media to build a set of potential candidates for each structural element.

### *6.3.2.Relevant candidate selection rules*

The previous step produces a set of the potential candidates according to the topic. The discursive functions and the narrative positions triggers selection rules, which apply on the indexing resources. Different types of indexing descriptors can be concerned from form attributes (visual or audio) to indexing discursive functions.

Rules detail which indexing components should be used and when (see Figure 6). They are organized in grammars, which are libraries of rules with a common purpose. Selection constraints can be so strict that no resource match. Then backtracking mechanisms provide, in case of failure, trigger of another rule.

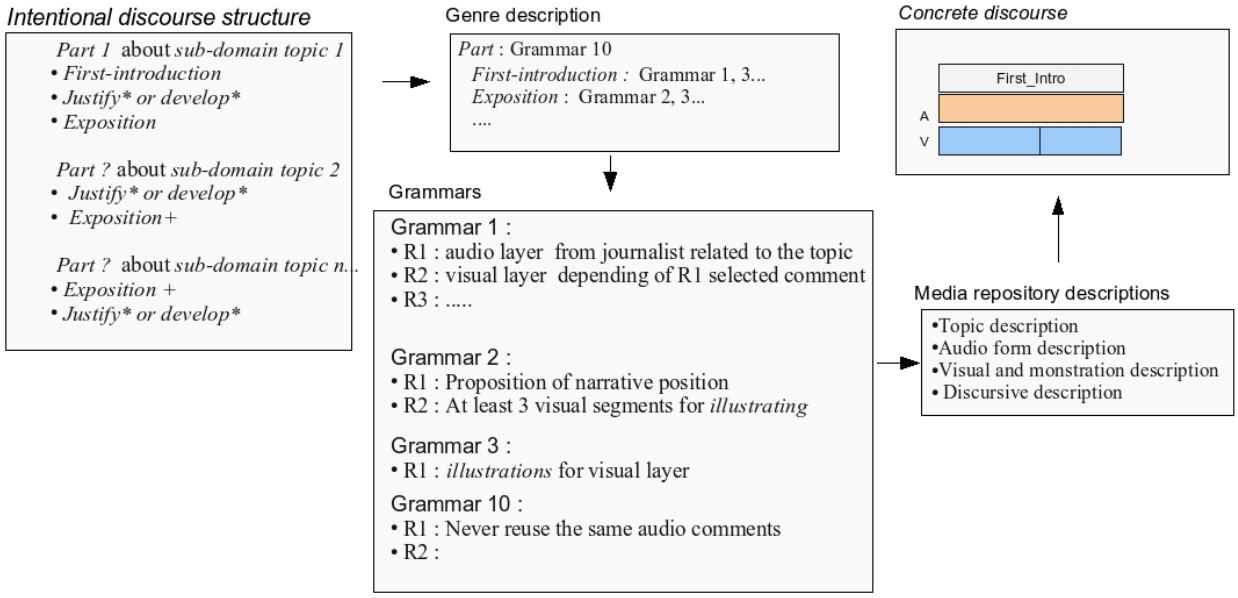


**Figure 6. Selection and transformation rules processing.**

Figure 7 give an example of different assembly patterns depending on discursive functions. The *first-introduction* of a TV news report is a specific montage pattern, which is different from the other introduction. Thus, `*first-introduction*' is a montage composed of a speaking journalist, who appears in a visual shot. The best choice of comments would be to find a newscaster's one, but every journalist speaking could also be relevant. After a few seconds playing, the visual layer must change to show an illustration which contextualizes the subject. The other kind of *introduction* looks very different: a

journalist never appears on the visual layer but some shots related to the topic must illustrate or symbolize the subject. Even if the discursive role is the same (in this case: *exposition*), the narrative position in the discourse changes the audio and visual patterns for relevant material selection.

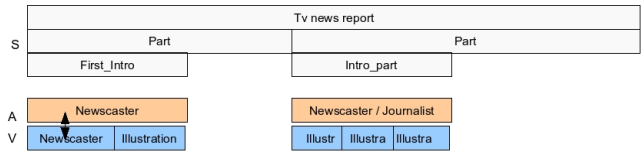


**Figure 7. Audio and Visual publishing pattern of *first-introduction* and *introduction.***

We will describe more precisely this process in another paper, including full rules examples and implementation using the TTL language [Nanard 2006].

### *6.3.3.Montage propositions*

The concrete discourse results from rules executions on potential candidates. Publishing process tries to best satisfy audiovisual patterns. These structures reach compromise between the wished intentional structure and the available resources. Author can also preview the audiovisual montage. As shown in Figure 8, this editing is not a simple contiguous set of audiovisual elements but rather a real montage, with specific treatment about audio and visual layers. Depending on the elements role, some archive extracts can be used as they were, without any recombination.

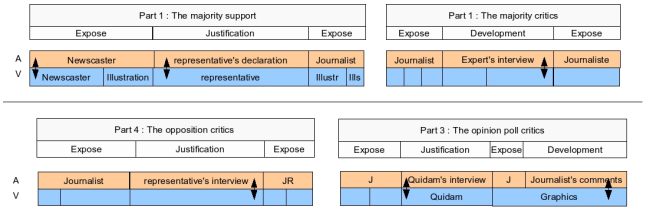


**Figure 8. An audio and visual montage based on TV news model.**

## 6.4. Stylistic

Stylistics relates to the final aspects of working. For multimedia resources, it means the spatial, temporal and interactive layout. In our example, it regards mainly the temporal dimension, such as video or audio *launching*, *duration*, *fading transition* and so on. Major structural arrangements as parts inversions, or roles modifications are not allowed in this stylistic final step. These structural dimensions must be treated by the intentional structure exposed previously.

# 7. DISCUSSION AND FUTURE WORKS

In this paper, we have shown how to articulate the indexing models and the publishing models, thanks to the independence of their associated structures. Our authoring process provides correlation mechanisms, which are based on selection and transformation rules, between the various structures of discourse.

The main advantage of this two principles makes it possible to create news kinds of audiovisual documents by re-using some structures, partly or totally; importing existing sets of rules and writing necessary rules. A new rule consists in establishing links between the publishing semantic discursive functions and the different components of the indexing models. A new model of audiovisual document can therefore re-use preexistent materials, without having to index again the repository.

We discuss now the state of progress of our work and draw future plans on the following aspects:

## 7.1. Process implementation

All the features described in this paper are currently under implementation in the Saphir framework based on Java Eclipse technologies and plug-in interface. The Indexing studio for indexing audiovisual resources is already available. It uses conceptual graphs. We plan to fix soon the ontological vocabulary and pattern graphs for achieving annotation. Thetopic query step use the Cogitant engine [Genest 1998] and Conceptual Graph [Chein 1993] query mechanisms. Transformations are specified in TTL in the SYGXML environnement [Nanard 2006] and are applied by the Sygmart engine [Chauche 1984] to produce publishing formats such as SMIL [SMIL 3.0 2007], XTL [XTL 2002], etc.

## 7.2. Device and target public component

Depending on target public (expert, novice), we plan to adapt the knowledge complexity level of target document. This component should be modelled through genre declarations and thereby should influence material selection and arrangement. Few other projects working in the same field have ever made propositions [Kim 2002]. We also intend to focus on using various target devices such as mobile phone, PDA and computer. Using such various devices will have consequences not only in terms of publishing format but mainly on discourse structures and grammar selection rules. The actual device will influence not only assembly rules of reused audiovisual material but also future navigational rules.

## 7.3. Libraries of rules

We have presented a general mechanism of selection and transformation. The next step, now in progress, is to write most of the rules and manage them into grammars. The aim is to obtain different libraries of grammars, which could be selected and reused for new kind of publication. The advantage of our declarative approach is to permit other expert authors to create their own kinds of multimedia documents. We plan to apply our authoring process (discourse structures and related grammars) to other types of publication and corpus. Thereby, we plan to explore multimedia target documents based mainly on audiovisual resources. We'd like also to develop more subjective model like *thesis*. For example, personal *multimedia critics* about famous movies. This kind of discourse, based on use of personal resources from one of the authors mixed with archives, is currently explored.

## 7.4. Graphical user interface

Most of multimedia authoring tools rely on a sequential authoring process. Even if we use specific structures, we observe that the writing process is not linear and consequently requires constant modifications of each step [Hayes 1980]. A linear sequence of inalterable steps would risk to be too constraining for authors. Using graphical interfaces like map mind tools (specially the Osmose project [Esposito 2005]) seems to be a relevant perspective. We currently are working to adapt this kind of visual representation according to an authoring audiovisual process.

## Acknowledgements

This research is supported by the Saphir RIAM project. The authors wish to thank their colleagues of the ANA team of INA: Patrick Courounet, Steffen Lalande, Claude Longerinas, for discussion and help during this work. The authors would also like to thank all the Saphir partners (LIRMM-IHM, LIRMM-RCR, LERIA, MSH, NETIA). Finally, we'd like to thank Madjid Ketfi and Quentin Goureau who actually implement the operational framework.

## REFERENCES

[1] Arijon, D. *Grammaire du langage filmé*, Encyclopédie de la mise en scène, Editions Dujarric, 1993.

[2] Aristote. *Rhétorique*, trad. M. Dufour, Les Belles Lettres, Paris, 1967.

[3] Bocconi, S., Nack F., and L. Hardman. Supporting the generation of argument structure within video sequences. *In HYPERTEXT '05: Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, pages 75–84, New York, NY, USA, 2005

[4] Carruthers, M. *Le livre de la mémoire*. 2002. Macula.

[5] Chamming's, L. *Pour une description sémiotique de l'audiovisuel*. Actes du colloque international "sémiologie 2003". Paris, Sorbonne, Novembre 2003.

[6] Chandler, D. *An introduction to genre theory*. 1997, Available online at: www.aber.ac.uk/media/Documents/intgenre/intgenre.html (accessed 24 February 2006)

[7] Chauché, J., Un outil multidimensionnel de l'analyse du discours. In *Proceedings Coling84*, 10th International Conference on Computational LINGuistics, Stanford University, California, 1984, 1984, 11 – 15.

[8] Chein, M., and Mugnier M.L. Conceptual Graphs: Fundamental Notions*, Revue d'Intelligence Artificielle*, volume 6-4, pages 365-406, 1992.

[9] Duff, D. *Modern Genre Theory*, London: Longman, 2000.

[10] Eco, U., *Lector in fabula*, Le livre de poche, 2001.

[11] Esposito, N. *Osmose : un logiciel libre de présentation multimédia basé sur une nouvelle approche*, Journée du libre 2005.

[12] Falkovych, K., Nack, F., van Ossenbruggen J., and Rutledge, L. Sample: Towards a framework for system-supported multimedia authoring, in *10th International Multimedia Modeling Conference* (MMM 2004), January 2004, Brisbane, Australia,

[13] Falkovych, K. and Nack, F. Context aware guidance for multimedia authoring: harmonizing domain and discourse knowledge. in *Multimedia Syst*, vol 11, 2006.

[14] Fowler, A., Genre, In Erik Barnouw (Ed.): *International Encyclopedia of Communications*, Vol. 2. New York: Oxford University Press, pp. 215-7,1989.

[15] Genest, D., and Salvat, E. A Platform Allowing Typed Nested Graphs: How CoGITo Became CoGITaNT". in *Conceptual Structures: Theory, Tools and Applications, proceedings of ICCS '98,* Montpellier, France, Springer, 1998.

[16] Geurts, J., Bocconi S. , van Ossenbruggen J. , and Hardman, L. Towards ontology-driven discourse: From semantic graphs to multimedia presentations, in *Proceedings of the 10th International Multimedia Semantic Web Conference*, p.362, January 5-7, 2004.

[17] Greimas, A.J., *Sémantique structurale*, Paris, Larousse, 1966.

[18] Hardman, L., *Canonical Processes of Media Production*, In Proceedings of the ACM Workshop on Multimedia for Human Communication - From Capture to Convey (MHC'05), 2005. Available on line at http://homepages.cwi.nl/~media/~media/conferences/mhc05/hardman.pdf

[19] Hayes, J.R. and Flower, L.S. *Identifying the organization of the writing process*, in L.W. Gregg, E.R. Steinberg (Eds), Cognitive process in writing, Lawrence Erlbaum Associates, 1980, 3 – 30.

[20] Kim, S., Alani, H., Hall, W., Lewis, P., Millard, D., Shadbolt, N. and Weal, M. Artequakt: Generating Tailored Biographies from Automatically Annotated Fragments from the Web. *In Proceedings of Workshop on Semantic Authoring, Annotation & Knowledge Markup (SAAKM'02)*, the 15th European Conference on Artificial Intelligence*, (ECAI'02)* -(-), pp. 1-6, Lyon, France.

[21] Mann, W.C., Thompson, S.A. *Rhetorical structure theory: towards a functional theory of text organization*. Text 8 (3*)* (1988), 243-281.

[22] Nanard, M., Nanard. J., Chauché, J., and King, P.R. A structural computing approach to the production of multimedia document series, *New Review on Hypermedia and Multimedia*, Vol. 12 (2), 2006, 165-190.

[23] Metz, C. *Essais sur la signification au cinéma*. Klincksieck, 1969.

[24] Perelman, Ch., *L'Empire rhétorique*, Vrin, Paris, 1977.

[25] Reboul, O. *Introduction à la rhétorique*. PUF, 2001.

[26] Rocchi, C. and Zancanaro, M. Generation of video documentaries from discourse structures, in *Proceedings of the 9th European Workshop on Natural Language Generation*, 2003.

[27] SMIL 3.0, http://www.w3.org/AudioVideo/ (accessed 4 February 2007)

[28] Stockinger, P. Le document audiovisuel. Hermes, Lavoisier, 2003.

[29] Toulmin, S., Rieke, R., and Janik*, A. Introduction to Reasoning*. MacMillan Publishing Company, 1984.

[30] XTL reference http://doc.51windows.net/Directx9_SDK/?url=/Directx9_SDK/dir_8.htm

[31] Zettl, H. *Sight sound motion*. Wadswoth publishing company, 1990.