A Context-Based Audiovisual Representation Model for Audiovisual Information Systems*

Yannick Prié[1,2], Alain Mille[2], and Jean-Marie Pinon[1]

[1] LISI 502, 20 av. Albert Einstein, INSA Lyon,
F-69621 Villeurbanne Cedex, France
Yannick.Prie@insa-lyon.fr
pinon@if.insa-lyon.fr
[2] LISA, CPE-LYON, 43, Bd. du 11 novembre 1918,
F-69616 Villeurbanne Cedex, France
am@cpe.fr

Abstract. In this paper we present a contextual representation model of audiovisual (AV) documents for AV information systems. In the first part, we study AV medium, and show that AV intra-document context is always related to a user task seen as a general description task. We then present the AI-Strata model for AV description: audiovisual units (pieces of AV documents) are annotated with annotation elements described in a knowledge base. The annotation elements are connected at the document level. The whole system being considered as a single graph, we define a context of one element as end points of graph-paths starting with this element. In order to control contextual paths, we define the notion of potential graphs as graph-patterns instantiated in the general graph. Finally, we show how these graphs are used in the main task of AV information system: navigation, indexing and retrieval.

1 Introduction

With the huge growth of data storage capacity and computing power, multimedia documents have become a reality (*e.g.* web pages, CD-ROMs). Among them, audiovisual (AV) documents, *i.e.*, documents composed of several sequential streams using a single temporal line (TV, video, radio...), can now be captured, edited, stored, and seen/heard in a digital form. These changes will surely induce on the long term changes in the media themselves, but on the short term, considering the increase in networking capabilities, many issues related to digital libraries have to be met, both for institutions (public libraries, audiovisual repositories) and companies (TV Channels), but also, and in an increasing way, for private users. Hence, the design of *audiovisual information systems* (AVIS) becomes an important research area.

* This work is partially supported by France Télécom (through CNET/CCETT), research contract N° 96 ME 17.

AVIS should allow users to *describe* and *index* audiovisual documents, in order to be able to manage them easily. Thus description schemes are needed[1], first for AV documents retrieval from the AVIS, that can then be played, second for the re-use and the manipulation of these documents (for instance an excerpt of a movie can be reused on a TV report). The overall design of audiovisual information systems is also crucial, and should benefit from results obtained in the study of textual and hypertextual information systems, mainly about the relations between the system and its users. For instance the user should be able to explore at his own pace the system, mainly using *navigation* (which has become a standard); the system should also help the user in his task, trying to reformulate his queries (relevance feedback), using a model of the task the user is performing, and even learning from his actions and results. In other words, the user and the system have to *collaborate* to reach a solution that matches the user's needs in various contexts.

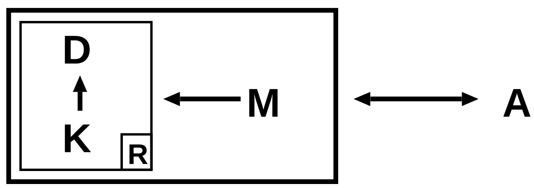

Fig. 1. A framework for identifying different views of context

Information systems, more and more heavely based on knowledge, gain from being studied along "contextual points of view", *i.e.*, considering how context, or contexts can be defined among the different parts of such a system. Used for a long time, *context* has been studied for itself only a few years ago[2]. In [4], Brézillon et al., though they admit that in fact this notion has multiple appearances — depending mainly on the ongoing task needing it — have made an attempt to retain its main characteristics. Indeed context is "something surrounding an item and giving meaning to it"; and it "cannot be considered out of its use". Moreover, "there are different types of contexts with respect to what we consider, and in which domain we are", and "all these contexts are interdependent". Restricting their purpose to information seeking systems in [5], Brézillon et al. propose an adaptation of a Newell and Simon model as a framework for the consideration of different views of the context. In figure 1, we refine this model into a documentary information processing system, with the repository R composed of two

[1] Several normalization comities like MPEG7 [10], or the EBU/SMPTE task force (see `http://www.ebu.ch/pmc_es_tf.html`) are currently working on these issues.

[2] Of course the word "context" is also often used is a very general way, for instance in their scheme for information retrieval, [13] name after context the content-theme of the searched document, which mainly reduce context to document genre *e.g.*, a cooking recipe.

interacting parts: a set of documents D and a knowledge base K (for instance metadata on the documents, thesauri, knowledge about the search task, *etc.*) A is an agent (human or machine) interacting with the system, and M is the set of "mechanisms" allowing access to R.

Depending on the research area considered, it is possible to associate several definitions of the context to elements or relations between elements of this framework, as in [5]. For instance the *interaction context* (somewhat related to pragmatics and discourse analysis) is at the level of the double arrow. The *knowledge-representation context* (linked with the reasoning context in AI and logics) lies at the K level and deals with internal inferences implying a symbolic notation of context. The *organizational context*, which contains the enunciation context in which the user is, takes place around A. Moreover, if we consider the documents as texts, the *internal linguistic context* [7] is located inside the documents in D. Of course, it is possible to define other contexts that rely on other simpler contexts.

Getting back to documentary information systems, it seems to us that the main dichotomy in context definition opposes internal textual/linguistics approach of the documents themselves, and a more cognitive approach using the document-user point of view, and the *situation* of his practice. In this rough scheme, the AI approach wanders between these two main poles, with the useful constraint of symbolic computing efficiency: at the document level, computerized knowledge helps enlighting the documents, while at the user level, it deals with the task (*i.e.*, the machine representation of the task) performed by the user.

In the first part of this article we will study audiovisual medium modelling, and show how intra-document context is always related to user tasks. In the second part, we will present AI-Strata, a generic model for representing AV documents, designed for tackling up intra-document context. Inter-document context, and context knowledge on AV documents are also supported, all these contexts being also related to tasks. We will then present our very pragmatic definition of context, and methods designed to use it. We will finally show how these methods are used in main audiovisual information system user tasks.

2 Modelling Audiovisual Documents in a Context-oriented Way

2.1 Audiovisual Medium and Modelling

An audiovisual document is composed of a superposition of streams that can be *aural* (music, voices) and/or *visual* (video, texts). The streams are both *sequential* (like text) and *temporal* (a speed rate is imposed). In the case of video stream, the 24 to 30 images per sec. frequency and the retinal persistence creates the *illusion of reality* (objects seem to live on the screen, and thus look more vivid than in still images). Above this first level of frame sequentiality, we can consider a second level, built on shot[3] sequences: *montage* of shots into sequences allows to make sense across shot cuts, just like sequences of sentences make sense into a text. Finally, superposing different streams of different modalities allows to benefit from the global effect of their union: for instance music can link totally different images using an audio metaphor, sound can reinforce the visual effects of a fight, *etc.*

Because of the proper nature of the AV media, which is mainly *non-textual*, its contents have to be reformulated in a semiotic form in order to facilitate their manipulation as symbols in a computer. In the modelization of digital sequential documents for computer representation, *annotation* is the fundamental process, which attaches an annotation (a description) to a piece of the document, each piece being delimited by two limits. For temporal media, these limits are obviously two instants in the AV stream. To characterize annotation of AV data according to that scheme, we proposed in [12] several criteria. *Time granularity* is the first; it deals with with the level of abstraction and the regularity of the cutting-up of documents into AV pieces: document level, shot or scene level with full decomposition, or video pieces as simple *strata*. The *kind of data* used to annotate the pieces is the second criterion: from the low-level features automatically extracted from the stream (color histograms, textures) to higher conceptual level characteristics like shots, keywords, or texts, everything is possible. The third criterion is the *degree of complexity* of the organization of characteristics into annotations: simple or atomic when a term or a numerical feature is attached to a piece, it can reach higher complexity like attributed structures or even semantic networks. According to these criteria, there exist many ways to describe, at different levels, with different complexities, AV material pieces that are cut following different granularity schemes. Our last and fourth criterion for annotation characterization is the structuration of pieces of the different pieces of a document, and is strongly related with granularity choices. There are two main approaches for the structuration of AV documents: *segmentation* and *stratification*.

Segmenting an AV document consists in cutting it up into *pre-defined* pieces which will be annotated later. An arborescent structural organization is also set-up to express a document structure [11] (see figure 2, left). On the contrary, the *stratification* approach [1] (see figure 2, right) means that the annotator freely defines *strata* (pieces) when needed. An *a posteriori* useful cutting-up can be derived from strata intersection. The essential difference between these two approaches lies in the definition of the temporally situated and annotated pieces of documents. In one case the cutting-up exists *before* an annotation that can be considered as *second* while in the other case it is *dynamically* created by the annotation process, annotation and cutting being tightly linked.

[3] Visual shots: stream of n contiguous frames continuously recorded by a single camera; extended to aural shots: n seconds of an audio stream, with an internal semantic coherence.

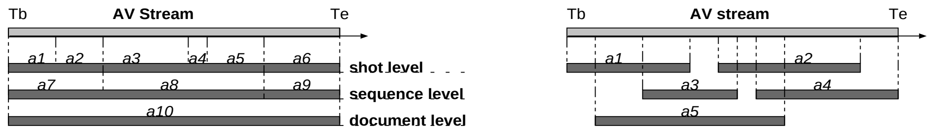

Fig. 2. Segmentation and stratification approach, the annotations are the a_i

2.2 Contextual Necessities

Our liminary remarks show that the *intra-document context* plays a very important role in an audiovisual document. At the montage level for instance, it has been proved a long time ago that a shot doesn't convey the same meaning when seen alone, or nearby another shot[4]. At the superposition level, sound can also totally modify the meaning of an image.

We can at least define two different types of contexts in an audiovisual document: the *temporal* context is linked with the temporality of the medium (for instance in the stratification approach the annotations of two strata that temporally overlap can be mutually influenced); the *semantic* context deals with the others contextual relations. In the semantic context comes first the *structural* context, dealing with the structure of the document. Indeed, a document is generally composed of sequences, themselves composed of shots: this hierarchy allows mutual influences between annotations from different structural units (for instance a sequence is enriched by the annotations of its shots [6], or a shot inherits the annotations of the document to which it belongs [15]). More generally, any co-reference meaningful context between any parts of a document is possible and belongs to the semantic context (for instance "each time this little character appears on the screen, I can hear that funny tune, and here the tune is used alone, thus represents the character").

If the temporal context — related to the fact that the audiovisual superposed streams are temporal — is shared in every usage of audiovisual material (*i.e.*, based on visualisation), *semantic context is related to the circumstances of the contextualization by the spectator*. As a consequence, in the more specific case of an AVIS, contextualization is related to the different tasks of the users.

Moreover, performing these tasks always entails the description of pieces of audiovisual documents. As the description of AV material is nothing but the elicitation of symbolic annotations and their placement in contexts where they are meaningful, "inserting in context" is the basic operation of any describing task, such as:

- Indexing : each symbolic annotation chosen for the representation an elicited object in the AV document is "explained" by its context.

[4] In 1920 years, a soviet director, Kuleshov, made the following experiment: he showed to an audience a shot of a man expressing no feelings, preceded by 1- a shot of a burial, 2- a shot of a child playing, 3- a shot of a soup plate. For the audience, the man was first sad, then happy, and at last, hungry.

– Searching : a searched item is described by a set of annotations together with their meaningful contexts.
– Navigating : surfing through the AV document, from one piece to another following meaningful links defined by the context where they make sense.
– Analysing : starting from a given studied, finding pieces of the video that are reachable in such a context.

As a conclusion, we claim that there is a clear link between task oriented contexts and intra-document contexts, as becoming aware of a contextual relation in an AV document is done in a description process. In next part we propose a description model of AV document able to operate with these contexts, but also with inter-documents context, and knowledge context in information systems.

3 The Annotations-Interconnected Strata Model

In this part we describe our model for the description of AV documents (for a more detailed presentation, see [12]). In our model, we privilege the stratification scheme for the following reasons: first, it is obviously more adequate to the representation of the dynamic aspects of AV material ; second, we consider strata and atomicity of the annotation as primordial, before any *a priori* segmentation (moreover, stratification approach is more general than segmentation: shots are just strata described as shots).

Objects of interest, analysis dimensions. We call *object of interest* any object (in the general sense of the term) that can be spotted when watching/listening to an AV stream. Objects of interest can refer to any kind of characteristic, at any level of abstraction; there are as many of them as there exists analyses of the stream. We group these analyses into *analysis dimensions* that allow to spot the same kinds of objects. For instance an analysis dimension can be related to shots, faces, people, moves, or President Clinton detection.

Audiovisual units, annotation elements. As soon as an object of interest is detected, it defines a temporally extended *audiovisual unit* (AVU) representing a stratum (the name of the stream and two temporal limits[5]), and at least one *annotation element* (AE) as a term, symbolic expression of its meaning. The annotation element annotates (is in relation R_a with) the audiovisual unit it has defined (we stay in the stratification approach). For instance, spotting a shot leads to define an AVU annotated by the AE $\langle Shot \rangle$, spotting for instance a well-known face leads to the creation of another AVU annotated by $\langle Mandela \rangle$, and so on for any object of interest: $\langle Zoom \rangle$, $\langle Round\_shape \rangle$, $\langle Sad \rangle$... A second description level is provided by *AE attributes*, for instance the numerical value of the histogram for $\langle Color\_histo \rangle$, a text as a speech excerpt for $\langle Script \rangle$ or a representative image for $\langle Shot \rangle$ (see figure 3).

[5] An important fact is that we do not represent temporal knowledge except in the audiovisual units: we consider that the use of this knowledge will proceed from the link between annotations and strata.

Elementary relations. To complete the primitive annotation that defines an AVU, it is possible to add as many AEs as necessary. The first way to do so is to add AE with the same temporal range, for example adding to an AVU defined by $\langle Document \rangle$ an AE regarding its $\langle Author \rangle$. The second and the most important way is derivated from the *structuration* of the annotation: in order to express more complex information than simply symbol-terms with a temporal extension, we allow *relations* between them. For instance, to express that "this shot has Mandela as video focus", we define relations between annotations elements *in the way* we already annotated. As on the example figure 3, we can use a third AE $\langle VideoFocus \rangle$ that acts as a relation term and then connect them with two elementary relations R_e. This method can be used to express any relation between any annotations elements as "this character is doing this action", and even between AEs in different AV stream, for "this shot is reused in that document". The model of the *Annotations Interconnected Strata* (AI-Strata) is named because of those relations.

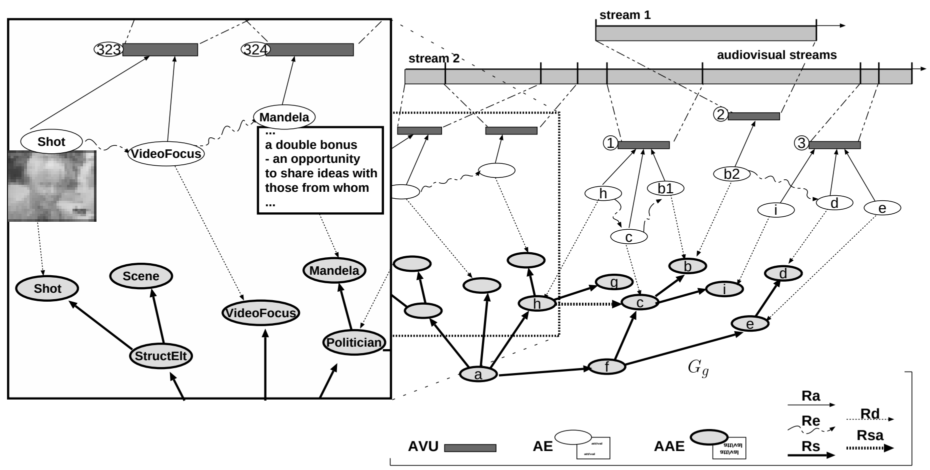

Fig. 3. General overview of AI-STRATA

Abstract annotation elements. In the same way as representation units and indexes, annotation elements and their relations are the operating substra that supports every access to AV material (thanks to AVU). The annotation process leading to this graph is done by a user who can be a professional (an archivist) or anyone. In order to facilitate and monitor further access, it is necessary to consider AE as terms issued from a controlled vocabulary in a knowledge base. An AE is then issued from an *abstract annotation element* (AAE), it is in *decontextualization relation*[6] R_d with it. The knowledge base is in fact a network

[6] This name comes from an analogy with linguistics.

of AAE with classical thesaurus relations (hierarchical: specialisation relation R_s and others, like see-also link R_{sa}), to which is added information regarding possible attributes of AE, or privileged relations. It acts as a shareable ontology, which is also necessary for reusing knowledge, serving as means for integrating problem solving, domain-representation and knowledge-acquisition modules (cf. [14]). Analysis dimensions are sets of abstract annotation elements.

4 Context in AI-Strata

4.1 Definition of context and contextualization in AI-Strata

As seen before, the basic elements of AI-Strata are the audiovisual units, the annotation elements, the abstract annotations elements and the relations between them. An AI-Strata system can then be considered as a unique oriented attributed *graph* G_g. This graph being connected, we can consider for any pair of elements $(x, y) \in G_g$ at least one *path* allowing to place x into relation with y. One can say that y belongs to the context of x (for instance on figure 3, the EA `d` belongs to the context of the AVU 2 considering a path containing `b2`).

We define the notion of *context* of an element x of G_g as any element $y \in G_g$ that can be placed into relation with it through a path in G_g. This definition is strict, general and well defined: a context is always a context of something that is known, in the knowledge-based system represented by the graph.

The process of *contextualization* consists first in the choice of one element of the graph as a beginning node for paths in the graph, and then in the search for the extremitates of these paths. For instance, if we consider an audiovisual unit as a beginning node (representing a Mandela shot in raw AV material), we can look for other audiovisual units contextually related to it (*e.g.*, representing all the documents where this shot has been re-used), as belonging to its audiovisual context when performing an analysing task. In the navigation task, we could search for annotation elements endpoints, and navigate towards them. Considering an abstract annotation element as beginning node, in a searching task we could look for audiovisual units in its context (*i.e.* AVU that are annotated with it). We will analyse these tasks in a more detailed way later.

4.2 AI-Strata methods for Manipulating Contexts

As we have seen, the context of any element of the system graph is potentially composed of the whole graph. This state of fact is indeed normal (after all, it is possible to find a semantic relation between any two concepts), but not very useful. So we need to gain *control* over the context, *i.e.*, over contextual paths in the graph. Keeping this in mind, we define two closely related types of graphs. A *fully attributed graph* is an oriented attributed graph with vertices (resp. edges) name attributes taking their values in a vocabulary V_V (resp. V_E). For instance, in our system graph G_g, the vertices take their values in $V_V = T \times E$, with T the types of nodes (here AVU, AE and AAE), and L the labels (for instance *Shot* or

$Mandela$ for AEs, and 324 for AVUs); while V_E contains the possible relations between elements (R_e, R_a, R_s ...). In a very similar way, a *partially attributed graph* is an oriented attributed graph defined under the same constraints as a fully one, with the exception that the vertices and edges names can also take a special value $*$. These pattern graphs are also called *potential graphs*. The name attribute value $*$ acts as a "wildcard" when the process attempts to *instantiate* a partially attributed graph g_p into a fully attributed one g_f. The instantiation of g_p into g_f consists in finding a partial subgraph g_s from g_f such as there exist an application between g_p and g_s that preserves the structure of the graph (syntactical constraint) and the name of the vertices and edges provided that $*$ acts as any other attribute value (semantic constraint)[7]. On figure 4, g_p is a potential graph, g_{i1}, g_{i2} and g_{i3} are the three instances of g_p in the fully attributed graph G_g of figure 3.

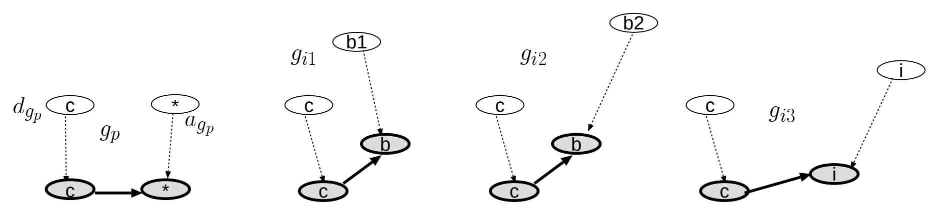

Fig. 4. Partially attributed graph and its instances

The potential graphs are characterized by some of their vertices, they can be named, manipulated, joined, *etc.*. As a first example of a potential graph, let us consider *designation graphs*. Designation graphs designate abstract annotation elements in the knowledge base; they are composed uniquely of AAE, and are characterized by one node. For instance, on figure 5, g_{p1} is instantiated into two subgraphs of G_g, the designation nodes in the instances designate the AAE $\langle Scene \rangle$ and $\langle Shot \rangle$.

Potential contextual relations express contextual paths. They are characterized by a source node and a final node. The skeleton of the graph (in grey on g_{p2} figure 5) expresses the path the graph designates, whereas other branches bear information about some intermediate nodes of the path (this could be related to the simple context notion of [8]).

For instance g_{p2} designate a contextual path between an annotation element and another one, with two intermediate AEs and one AAE. The branch ensures that the AAE needs to be a specialization of the AAE c. Using this potential

[7] Finding a subgraph isomorphism is a very difficult problem in the general case, and we put reasonable restrictions on our potential graphs: in a valid potential graph, at least one unambiguous association of the isomorphism must be known. For instance a AAE is known (it is unique by definition), or a AVU or a AE is precisely defined. The resulting trivial associations act as initialization for a propagation algorithm (multi-propagation if there are several initialization associations).

graph means *applying* it to an AE element of the graph in order to find the context of that element corresponding to the designated contextual path. For instance, if we apply g_{p2} to the annotation element `c` from figure 3 by the attribution to the beginning node of g_{p2} of a c value instead of $*$, we will find no instance of g'_{p1} in G_g; on the contrary, by applying it to `h`, we will find one instance of g''_{p2}, and the AE `b2` designated by the end node will belong to the context of `h` across g_{p2}.

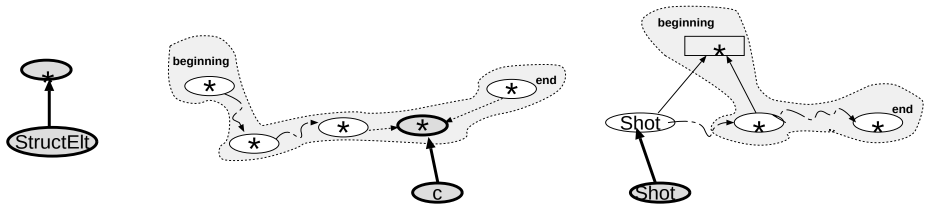

Fig. 5. Examples of potential graphs

Another important example of a potential contextual relation deals with the context of audiovisual units, *i.e.*, when the beginning node of the potential path is an AVU. For instance, the graph g_{p3} expresses a path dealing with the context of an AVU that is a $\langle Shot \rangle$. Applied to the AVU `323`, it allows us to get back the EA $\langle Mandela \rangle$ as belonging to the context of the AVU. We should also note that potential graphs can be manipulated and joined; for instance we could build a potential graph g_{p4} by linking the beginning node of g_{p3} and the end node of g_{p2}, thus describing a new contextual path. These manipulations allow to create contexts from others contexts.

4.3 Exploiting an AI-STRATA based AVIS : managing contexts according to different tasks

As seen earlier, *navigation* has become a mandatory feature of any information system. We consider navigating as going from one element of the graph to another, using any path. In AI-Strata, navigating means then simply applying a contextual potential graph to the current element, and selecting among the end nodes results which one is to be explored. We should note that this type of navigation is a generalisation of the standard one: the contextual path of navigation is controlled at any level, it is not just a selection of a predefined link. Navigation can occur at any level of the graph:

– inside the knowledge base as exploration of the annotation vocabulary and knowledge. The context of any concept in the knowledge base is an indication of its meaning (for instance abstraction and difference relationships with siblings in the hierarchy). These contexts are also indices about the knowledge base creation task;

– inside the document base, as intra- or inter-stream navigation. Intra-stream navigation is related to internal contexts, while inter-stream navigation can be related to hypertext jumping and "intertextuality" study;
– from the documents to the knowledge base, for instance for a better understanding of a term that is explained with its concept relations as in [9].
– from the knowledge base to documents, for instance for explaining a concept with its use in a real case.

Indexing uses *contextual annotation* in a large manner, because any audiovisual unit is considered as annotated with the annotation elements it is directly in relation with (local context) but also with annotation elements that are in other contexts. For instance, in figure 3, the AVU 323 is directly annotated with $\langle Shot \rangle$, but also, and to the extend controlled by potential graphs, by $\langle Mandela \rangle$ or even $\langle StructElt \rangle$. This means that the semantic content of an AVU depends on the context around it that we consider. This context (*i.e.*, potential graphs) depends of course on the will of the user.

Searching or *querying* audiovisual units in AI-Strata can take several forms. In a *precise* query, the user describes the AVU u he is looking for by designing a potential graph with a *virtual* AVU (with a $*$ name attribute), in relation with AEs, which have in turn relations, and so on. Answers to the query are then the u AVUs from the instantiations of the graph in G_g. For instance, g_{p3} itself can be considered as a query, whose answer is the AVU 323. In a general AI-Strata query, the user just describes which annotations elements should annotate the AVU he is looking for, and from what context (expressed as contextual potential graph) these annotations should come. The system then transforms the query into potential graphs, instantiate them, and gives back the results.

The queries we have just evoked only interfer with at the first level of description, *i.e.*, the surface knowledge of annotation given by AE name values. But other queries have to do with the deeper knowledge represented in other AE attributes, like image features, or texts. In such cases the instantiation process has to take into account not only binary name similarity, but also other ones, adapted to other attributes. The important point here is that AI-Strata, because it provides a way to represent any audiovisual characteristic at the same description level, allows to mix many searches in a natural way, using potential graphs. For instance, a potential graph mixing surface knowledge (a shot must be annotated by a politician) and deeper knowledge (face recognition knowledge) can help to reduce the search space — the context — in a tremendous way for the application of expensive feature similarity computations (for instance the face recognition will only be used for shot annotated by politicians). It is also a possibility to consider the context of an AVU as a pre-annotation that monitors image-processing methods, as in [3]. For instance, if a document is already annotated as $\langle TVNews \rangle$, the presence of this AE in the close context of the AVU to be studied could trigger a better image processing method than the general one.

Finally, some words about the notion of *valences*, which are possibilities of relation for annotation elements, represented as contextual potential graphs, and stored as attributes of abstract annotation elements. Valences are useful for annotation and are used when an abstract annotation element is used to annotate, as annotation element, an audiovisual unit. If there is a valence, the system looks for instantiation of the potential graph in order to detect in the context of the new AE if there is another AE that could be placed into elementary relation with it. For instance, when using the AAE $\langle Mandela \rangle$ as a character to annotate an AV with the AE $\langle Mandela \rangle$, one could use its valence attribute to detect if in the context, there were no $\langle Action \rangle$-related AE, able to express the fact that $\langle Mandela \rangle$ is doing this action (for instance $\langle ToDance \rangle$). These steps are illustrated in figure 6: (1) Instantiation of the potential graph in the valence attribute, here representing a simple co-occurence context in an AVU. (2) An instance of the potential graph designate with its endpoint the AE $\langle ToDance \rangle$ in the context of the AE $\langle Mandela \rangle$. (3) An elementary relation is set up between $\langle Mandela \rangle$ and $\langle ToDance \rangle$.

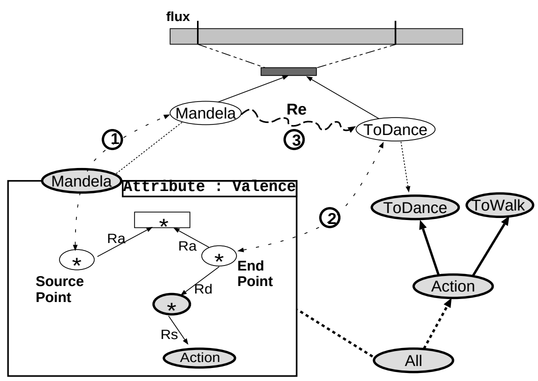

Fig. 6. Example of a valence of the AAE $\langle Mandela \rangle$

Valences, as part of the knowledge base, are useful for learning possibilities of relations from local and idiosyncratic relations and contexts. A context that has been set up by a user can eventually be generalized in a valence, and a valence that is much used for the annotation could entail the creation of a concept relation in the Knowledge Base.

5 Concluding discussion

In this article, we have presented an original approach for the modelization of audiovisual documents in a context oriented way. As computer representation means eliciting symbolic annotations, and as any AVIS-related task entails describing AV document pieces and putting them in context, we have shown that a context-suited representation model was needed. The Annotation-Interconnected

Strata approach allows to take into account both temporal (related to the fundamental temporality of audiovisual streams) and semantic contexts (related to the task the user is performing). The annotation graph is considered as a whole, and we have given a strict definition of a context of a graph element. Apart from this simple modeling of context, we have presented the potential graphs as patterns enabling to control contexts, and how these graphs (hence the contexts they represent) are used in the main description tasks of an audiovisual information system. Contextualization hence appears related to a description task, which is itself related to potential graphs. Such an approach and modelling of context is original in the multimedia document research area.

The AI-Strata approach represents an original approach in the multimedia representation field, indeed it subsumes segmentation, does not make any assertion about the document structure and allows to represent any different characteristics at the same description level. Having been designed with context in mind, many tasks of an AI-Strata based AVIS can be thought of as annotation elements description tasks in meaningful contexts. Moreover, since our notion of the context is graph-based, this approach could be extended to any other information systems, with the limitation that information should be represented with graph-based representation allowing contextualization.

A first prototype was developped, demonstrating the feasability of the ideas we have presented, and the performances of our potential graph instanciation algorithm. We are now working on a second prototype, for which we are studying the representation of AI-Strata in XML-designed documents (works like [2] are precursors in that field), and how the instantiation algorithms can be adapted to the XML world. In collaboration with France Télécom CNET, we are also working on a proposal based on AI-Strata for the MPEG-7 audiovisual content description standard.

On a more theoretical side, the ability of the model to tackle up any sequential media, like texts, could open fruitful uses in textual information system, and linguistics fields. We intend to deepen the study of the relations between internal linguistic contexts, and internal document context (in the reformulation approach for modelling, elementary relations between instances of annotations provide linearity to terms). Indeed, the semantic content associated with each annotation element depends on the position of the abstract annotation element which it is extracted from in the knowledge base (in the conceptual world), but also from its position in the annotation network. We should also remark that audiovisual units, as semiotic units of any temporal length (from one image to the length of the stream), can represent both local and global contexts for other audiovisual units, depending on the chosen contexts. This illustrates the fact that global meaning determines local meaning, while local meaning in turn has an influence on global meaning, in a very natural way.

Finally, it appears that an interdisciplinary research becomes more and more necessary for the design of representation models for future information systems; as document models should now be constructed in cooperation with humanities scientists (linguists, semioticians), while sufficiently general computerized models can in turn help scientists from these fields to design and test new hypotheses.

References

1. T. G. Aguierre Smith and G. Davenport. The stratification system, a design environment for random access video. In *Proc. Network and Operating System Support for Digital Audio and Video - 3th International Workshop*, La Jolla, 1992.
2. G. Auffret, J. Carrive, O. Chevet, T. Dechilly, R. Ronfard, and B. Bachimont. Audiovisual-based hypermedia authoring: using structured representations for efficient access to AV documents. In *ACM Hypertext'99*, Darmstadt, Germany, Feb. 1999.
3. F. Brémond and M. Thonnat. Issues in representing context illustrated by scene interpretation applications. In *Proc. of the first Int. and Interdisciplinary Conf. on Modeling and Using Context*, Rio de Janeiro, 1997.
4. P. Brézillon and M. Cavalcanti. Modeling and using context: Report on the first international and interdisciplinary conference context-97. *The Knowledge Engineering Review*, 12(4):1–10, 1997.
5. P. Brézillon and I. Saker. Modeling context in information seeking. In *International Conference on Information needs, Seeking and Use in Different Contexts, ISIC'98*, Sheffield,UK, 1998.
6. T.-S. Chua and L.-Q. Ruan. A video retrieval and sequencing system. *ACM Transactions on Information Systems*, 13(4):372–407, October 1995.
7. J.P. Desclés, E. Cartier, A. Jackiewicz, and J.L. Minel. Textual processing and contextual exploration method. In *Proc. of the first Int. and Interdisciplinary Conf. on Modeling and Using Context*, Rio de Janeiro, 1997.
8. B. Edmonds. A simple-minded network model with context-like objects. In *European Conference on Cognitive Science (ECCS'97)*, Manchester, 1997.
9. J. Nanard and M. Nanard. Adding macroscopic semantics to anchors in knowledge-based hypertext. *Int. J. Human-Computer Studies*, 43:363–382, 1995.
10. F. Pereira. Mpeg-7 : A standard for content-based audiovisual description. In *2nd Int. Conf. on Visual Information Systems*, pages 1–4, San Diego, Dec. 1997.
11. D. Ponceleon, S. Srinivasan, A. Amir, and D. Petkovic. Key to effective video retrieval: Effective cataloging and browsing. In *ACM Multimedia 98 Proc.*, Sept. 1998.
12. Y. Prié, A. Mille, and J.-M. Pinon. Ai-strata: A user-centered model for content-based description and retrieval of audiovisual sequences. In Springer Verlag, editor, *First Int. Advanced Multimedia Content Processing Conf.*, number 1554 in LNCS, pages 143–152, Osaka, Nov. 1998.
13. S.W. Smoliar and L.D. Wilcox. Indexing the content of multimedia documents. In *2nd Int. Conf. on Visual Information Systems*, pages 53–60, San Diego, Dec. 1997.
14. E. Walther, H. Eriksson, and M.A. Musen. Plug and play: Construction of task specific expert-system shells using sharable context ontologies. In *Proc. of the AAAI workshop on Knowledge Representation Aspects of Knowledge Acquisition*, San Jose, 1992.
15. R. Weiss, A. Duda, and D.K. Gifford. Composition and search with a video algebra. *IEEE Multimedia*, 2(1):12–25, 1995.