Semiotic Annotation of Narrative Video Commercials: Bridging the Gap between Artifacts and Ontologies

Elio Toppano
Dipartimento di Scienze Matematiche, Informatiche e Fisiche (DMIF)
Università di Udine, Italy
e-mail: elio.toppano@uniud.it

Vito Roberto
Dipartimento di Scienze Matematiche, Informatiche e Fisiche (DMIF)
Università di Udine, Italy
e-mail: vito.roberto@uniud.it

Abstract—**Drawing on semiotic theories, the paper proposes a new concept of annotation – called semiotic annotation – whose goal is to describe the multilayered articulation of meaning inscribed within narrative video commercials by their designers. The approach exploits the use of a meta-model of the narrative video genre providing the conceptualizations and the vocabulary for analysis and annotation. By explicating design knowledge embodied in the video, semiotic annotation plays the role of intermediate level knowledge between the meta-model (an informal ontology) and practice (the concrete video artifact). In order to assess the feasibility of the approach, a test bed is presented and results are reported. A final discussion about the potential contribution of semiotic annotation in the fields of Research Through Design, Technological Mediation, and Interface Criticism concludes the study.**

Keywords-video; content annotation; ontology; semiotics; advertising; semantic web.

I. INTRODUCTION

This paper is about representation and semantic-based content annotation of narrative video commercials during production. It is an elaboration and extension of previous work presented in the IARIA Fifth International Conference on Building and Exploring Web Based Environments [1]. Narrative video commercials are multimodal artifacts used in the domains of marketing and advertising to communicate a product's features, a public service announcement (PSA), or abstract concepts such as a brand personality [2] or brand identity [3] through a story-like format. The intention is to persuade people to buy the product or, in the case of brand communication, to resonate with brand meanings (e.g., brand core values) [4].

Nowadays, video advertising covers a wide range of products differing in production quality, time length, and distribution. Standardization activities made by the IAB (Interactive Advertising Bureau) have provided several formats and guidelines to improve the development process of this genre of artifacts and to enhance the viewer experience [5][6].

A number of reasons motivate the use of stories in advertising. First, humans are storytellers. They use narratives as a natural and effective way to understand, structure, and communicate their experiences. This is because narratives evoke more meaning and emotions than bare facts. Narratives are also crucial for our understanding of time and time-based events as well as for understanding own identities and self [7]. Second, narratives have an intrinsic persuasive potential that is related to the extent that viewers/readers are "transported" into the world of the narrative and become involved with its protagonists [8]. Strong immersion into a story reduces counterarguments against story assertions, creates a lifelike experience, and provides strong connections with characters, all of which facilitate *narrative persuasion* [9][10]. Third, studies show that recall of narrative information is twice as likely as recall of expository information [9].

Narrative video commercials demand a careful design activity. A critical decision is, for example, how to integrate the persuasive message (e.g., information intended to influence the audience) with the story told by the video. Another decision is how to combine narrative persuasion with other non-narrative persuasive mechanisms such as reasoned arguments, statistical evidence, celebrity endorsement, etc. in order to achieve maximum effectiveness. Yet, another problem arising in brand communication concerns the development of a *brand-specific design language*, i.e., to decide how to communicate abstract meanings such as brand values and personality by a systematic and consistent use of expressive features (e.g., visual shapes, color scheme, auditory timbres and leitmotifs). This process - called *semantic transformation* - has been extensively studied in the field of industrial design [11] but has received much less attention in the field of multimedia design [12].

It should be evident from the above discussion that describing and annotating the (semantic) content of narrative video ads during the design process poses several problems due to the density and complexity of meanings that are inscribed into these kinds of artifacts. The annotation can be reduced neither to the description of what is depicted in the video nor to the specification of the general theme/claim of the video or the description of basic visual and auditory features. Rather, it should be possible to capture the entire range of meanings inscribed within the product including deep values, narrative structure, figurative and plastic meanings, rhetorical and persuasive mechanisms, to name only a few. Most importantly, the

annotation should be able to capture the relationships existing among all these meanings, i.e., their articulation in different conceptual layers within the artifact and their distribution across representation modalities (i.e., written words, images, sound objects).

In this paper, we address this problem by exploiting contemporary semiotic theories. We wish to evaluate whether a semiotic perspective provides a useful meta-model for the analysis and annotation of audio-visual resources and narrative video commercials in particular. Semiotics studies signs, meaning, and sense-making. It addresses processes of signification by investigating the ways meaning arises from mappings among sign structures. Therefore, it can be used to model the content and expression of a narrative video commercial intended here informally as "the sum of meanings that the designer intends to communicate through the space/time composition of the audio-visual resources that constitute the video presentation".

An important assumption is that the content of the video commercial is the result of an intentional act of designing intended as meaning-making or rhetorical argumentation [13][14][15]. This assumption does not hold - or only partially holds - for other types of audio-visuals such as surveillance videos, home videos, documentary and scientific videos, news videos, etc. where the content is largely determined by what happens in the reality and its structure depends on the nature of events being recorded rather than on high control over screenplay, editing, and filming. In narrative artifacts, in particular, the meaning is strictly related to experience and its constitutive components namely the sensorial, cognitive and affective component [16]. Therefore, designing a narrative video ad means embedding, within the artifact, the conditions for affording in the viewer an intended experience. The artifact, thus, plays the role of *mediator* (of experience) between the intentions of a sender (intended, here, metonymically to represent all parties involved in the development of the video including sponsors, client, designers, producer) and the interpretation of a receiver (the user, consumer) [17].

We are interested in the production side of this framework: how meaning (e.g., projected experience) is intentionally constructed and articulated during the message construction process, how it is embedded in the video and gives form to it. Consequently, by semantic annotation, we refer to a kind of "serious" annotation performed by trained professionals in the course of video development [18][19]. Its aim is to capture the intended and inscribed meanings in order to exploit them not only for retrieval, filtering, and browsing tasks but also for explanation, critical evaluation, and content (i.e., meaning) reuse. We shall not address online user' annotation or social tagging although they obviously represent an important contribution to the development of the field.

The paper is organized as follows. In Section II we review related work about multimedia semantic-based content annotation. Section III introduces the concept of Semiotic Annotation that is at the core of our approach. This kind of annotation exploits a semiotic compliant informal ontology (i.e., a meta-model) of the artifact under consideration. A critical discussion of available ontologies of narrative videos is presented as well as some basic requirements the design of an ontology supporting the proposed approach should satisfy. In Section IV, we illustrate the meta-model we have developed for semiotic annotation. It specializes our previous work on hypermedia [20][21] for the narrative video commercial genre. An example of application is presented in Section V while possible uses and implications for research are discussed in Section VI. Finally, Section VII summarizes the strengths and limitations of the approach and draws the conclusions.

II. Related Work

The term "annotation" can denote both an activity (i.e., the process by means of which additional data - *metadata* - are attached to existing data) and the result of that activity.

Models and technologies for annotation have been studied within many communities, with different goals and perspectives. In the digital library community, for example, metadata is seen as a way of supporting cataloging and retrieving information in a large collection of documents [22]. In the knowledge representation community, the focus is, instead, on representing the underlying content of a document rather than describing the document that contains the content [23]. In the semantic web community, an annotation is viewed, first of all, as a tool for representing and linking resources together in order to support information retrieval, filtering, and browsing [24].

Figure 1 represents a basic model of an annotation A in terms of a tuple, i.e., $A=< a_s, a_o, a_r, a_c >$ where a_s denotes the *annotated data* (i.e., the subject or target of the annotation), a_o the *annotating data* (i.e., the object or body of the annotation), a_r the *annotation relation* (i.e., a predicate that defines the type of relationship existing between annotated and annotating data) and a_c the *context* in which the annotation is made [25], [26]. The context includes several facets such as, for example, *who* makes the annotation (e.g., a single individual, a group, an automatic system; an expert annotator of a casual user); *when* (e.g., during different phases of the development process of a resource; during its use); *why* (e.g., for classification, description, retrieval, filtering, explanation, browsing, reuse); *how* (e.g., manually, semi-automatically, fully automatically; using free text, controlled vocabularies, taxonomies, ontologies); and *application domain* (e.g., entertainment, news, marketing, brand management).

The two most widely known approaches towards machine processable and semantic-based content annotation are the Semantic Web Activity of the W3C [27] and the ISO efforts in the direction of complex media content modeling, in particular, the Multimedia Content Description Interface (MPEG-7) [28].

The Semantic Web approach provides a structured set of cooperating languages (e.g., RDF, RDFS, OWL) and processing tools to define ontology vocabularies. It supports reasoning with ontologies but does not specify any collection of specific metadata for multimedia products. Andrews, Zaihrayeu, and Pane [24] surveying various

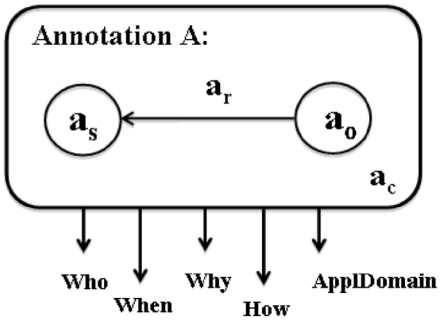

Figure 1. Annotation model: a_s is the target of the annotation, a_o the body, a_r the annotation relation and a_c denotes contextual information (i.e., who makes the annotation, when, how, and the application domain).

annotation systems for the web, investigate existing models for representing annotations and analyze their different characteristics, forms and function. Based on this analysis a classification scheme for annotation models was developed that distinguishes three main dimensions namely: i) the structural complexity of annotations (e.g., simple labels, attribute/value pairs, structured collections of concepts and relations), ii) the type of vocabulary used (e.g., free text, controlled vocabulary, taxonomies, ontologies), and iii) the level of user collaboration in sharing and reusing semantic annotation, and in the collaborative construction and evolution of the underlying vocabulary or ontology.

MPEG-7 is aimed at providing standardized means for describing audio-visual data content in multimedia environments [29]. The standard provides a set of descriptors and description schemes specifying the structure of the metadata elements, their relationships and the constraints a valid MPEG-7 description should adhere. It does not incorporate formal semantics and it is based on XML encoded metadata. As a consequence MPEG-7 is not open to standards that represent knowledge and make use of existing controlled vocabularies for describing the subject matter. Moreover, its XML schema-based nature has led to design decisions that leave the annotations conceptually ambiguous and, therefore, prevent direct machine processing of semantic content descriptions. In order to overcome these limitations, several approaches have been published providing a formalization of MPEG-7 as an ontology. A survey of these initiatives is provided by [30] where a detailed comparison of MPEG-7 compliant ontologies - such as Hunter's ontology, AceMedia, SmartWeb, Boemie, Rhizomik and COMM - is made on the base of three main dimensions namely: 1) low-level descriptors covering visual and audio features; 2) structural descriptions pertaining to the decomposition and localization of content parts; 3) subject matter descriptions expressing the semantic conveyed by a multimedia resource.

What emerges from the analysis is that most of the approaches use a modular architecture to mainly represent structural issues and low-level features, using OWL DL formal languages. Subject matter descriptions are usually demanded to external top ontologies such as SUMO or DOLCE or domain-specific conceptualizations. The COMM ontology (Core Ontology for MultiMedia) [31] constitutes one of the more recent approaches to the formalization of MPEG-7 descriptions semantics. It extends the Description&Situation pattern (D&S) and Ontology of Information Objects (OIO) of DOLCE by re-engineering the MPEG-7 description tools in order to provide a common foundational framework for describing multimedia documents.

A comparison between MPEG-7 and several other multimedia metadata standards is discussed in [32][33]. The analysis is made according to several dimensions including the specific media production process - e.g., premeditate, capture, archive, query, message construction, organize, publish and distribute [34] - during which the metadata are intended to be used.

What emerges from the comparison is that a support for describing semantic content linked to premeditate, message construction, and organization processes is generally missing. The premeditate process is where initial ideas about media production, e.g., goals, intentions, target audience, subject, genre, deep values are established; the message construction and organization processes are where the author/designer specifies the message he/she wishes to convey and media assets are organized according to the message.

MPEG-7, for example, supports premeditation in the CreationInformation description scheme and in the Classification scheme [29]. The former allows the annotator to represent date, place, action, material, staff involved in the creation of the content entity. The latter allows the annotation of the artifact with information about the subject, genre, purpose, and market segment. However, these data are completely disconnected with respect to content structure and semantics. As an instance, the genre is a label that is associated to the entire multimedia product; it cannot refer in any way to discourse structure or to expressive characteristics that are used to materialize that genre in the artifact. As a consequence, it is difficult to understand *how* the genre is intentionally constructed and articulated within the message during the design process, *how* it shapes the message and *how* users can infer the designer's intentions – both informative and persuasive. A similar argument can be stated for product purpose, intended audience and so forth. It is not possible to explain *how* a purpose is achieved in that artifact, *how* an intended audience is inscribed within the artifact itself, or *why* a specific design decision has been made.

As far as message construction and organization processes are concerned, the MPEG-7 standard provides very high flexibility by introducing general-purpose descriptor schemes for representing a wide collection of story-related semantic entities such as objects, people, events, concepts, states, places, and time together with their properties and relationships [29]. In practice, this flexibility is hindered by the fact that the annotator is left alone in using these descriptors to model a narrative.

The problem is that to be a true narrative, a text must exhibit a specific quality called *narrativity* [35]. This

quality, that may be present in various degrees, is the result of a set of interrelated structural factors (e.g., clear structure, genre typicality, affective structure, dramatic mode) [36], that are not taken into account by MPEG-7 and the other multimedia metadata standards alike.

In other words, to build a narrative, it is not sufficient to have the main ingredients, it is necessary to know how to put these ingredients together in order to reflect the specific qualities of this kind of genre. MPEG-7 and other standards provide the What but do not support the How.

For the genre of narrative video commercials that are the scope of the present article, what is needed is a vocabulary (and a conceptualization) that is able to bridge the gap existing between abstract concepts such as purpose, narrativity, narrative structure, discourse structure, intended values, affective states of the story's characters, etc. and perceptual qualities such as basic visual and auditory features. The aim of the research described in this paper is to explore how it is possible to represent this kind of meanings and use them to annotate the message in order to exploit it for content retrieval, explanation, and reuse.

III. SEMIOTIC ANNOTATION

In this section, we introduce the concept of *semiotic analysis* and *annotation* that is at the core of our approach [1]. This process exploits a meta-model (i.e., an informal ontology) of the narrative video genre. Therefore, we first review some relevant conceptualizations that have been proposed in the past to describe or annotate narrative videos. The analysis and comparison of these conceptualizations allowed us to identify a set of requirements for the development of our meta-model that will be illustrated in Section IV.

A. Semiotics

Semiotic studies cover a wide range of theories, models, and conceptualizations according to the specific intention they try to achieve and the unit of inquiry they address. Classical semiotics assumes the concept of *sign* as the main unit of signification and studies languages as sign systems [37]; interpretative semiotics focuses on processes of *interpretation* (i.e., semiosis) [38]. Contemporary semiotics extends its scope to the *text* construct intended as a unit of interrelated sign structures while Social Semiotics investigates human signifying practices in specific social and cultural circumstances [39]. More recent developments studies *mediated experiences* [40], *technical artifacts* and *design* [41][42].

What is common to all these approaches, regardless the variety of perspectives, is a focus on meaning, meaning construction (sense-making) and communication. As stated by Scolari [43]: “Semiotics studies objects (texts, discourses) to understand processes (sense production and interpretation)". From this point of view, semiotics appears as a methodology. What it actually does is to reflect on the more appropriate *methods* that can be used to perform the *analysis* of communicative and physical artifacts viewed as kind of multimodal texts. As a consequence, it elaborates tools (e.g., conceptualizations, models, grids of focal queries) for making the analysis and tests the effectiveness of these tools with concrete artifacts.

B. Semiotic analysis and annotation

By *semiotic analysis*, we mean a process of knowledge acquisition based on decomposition and re-composition of a given communicative artifact. Its aim is to unfold the *articulation of meaning* inscribed within the artifact by its designer/author. The basic assumption is that the *intended meaning* is spread over different interconnected *forms* - understood, here, as structured sets of relationships among content or expression entities - deployed within the artifact. The decomposition and re-composition processes are always based on some idea or conceptualization - a meta-model or (informal) ontology - of the genre of artifact under consideration. To be effective, such a conceptualization should be capable to capture the internal articulation of interconnected forms that are inscribed within the artifact. The result of the semiotic analysis is a partial or complete instantiation of the meta-model, i.e., a description (model) of the artifact that is then used for the annotation. The process can be detailed as follows (Figure 2):

- Step-0. An object - e.g., a clip video in the current case - is selected and regarded as a text, i.e., an autonomous, multilayered and organic unity having a goal/purpose: to produce effects by means of signification processes.
- Step-1. A meta-model of the type of object under consideration is selected as a reference guide to individuate main parts and relationships. The text object is *decomposed* accordingly.
- Step-2. Constituent parts and relationships are investigated, in turn, in order to understand how they may contribute to the functioning of the whole. To this end, a *re-composition* of the parts into the whole is mentally attempted and a description (model) of the artifact is produced. Again, the meta-model is used to build the description. This step involves a back-and-forth movement between pointing out material particulars and relating them to interpreted wholes.
- Step-3. The result of the analysis is used to annotate the object.

As shown in the figure, several meta-models of the type of object under consideration may be available, at a certain time, for supporting analysis and annotation; so the selection of an appropriate one depends on specific purposes and interests of the analysis. If we assume that the considered artifact is an organic unity and we seek to explain this unity, not all meta-models are equally appropriate for the task. We need a meta-model that embodies a hypothesis about the general organization and internal coherence of the elements that constitute the artifact. This coherence may have several sources, including the existing relationships between the artifact's structure and its function, the artifact's genre, author's style, the cultural meanings associated with multimodal materials. In this way, the annotation is not simply used for classifying the artifact elements into categories; we want to relate the elements and

their respective categories in order to explain its organic unity. Moreover, meta-models may change over time reflecting the interests and tastes of mainstream research communities. As a consequence, also the aspects of an object that are deemed relevant change. In the field of multimedia design, for example, we have witnessed a shift of interest from pragmatic issues related to technology, product utility, usefulness, and performance to hedonic aspects that are related to the whole human experience such as aesthetics, pleasure, fun, values. The meta-models have evolved accordingly.

It should be stressed that the process of semiotic annotation does not occur in the vacuum but within a specific pragmatic situation and a social and cultural context. The result of analysis and thus the annotation are simultaneously personal and inter-subjective. The annotation is personal because it is particular to the individual analyst, his/her knowledge, and experience. It is inter-subjective because the analysis is driven by a meta-model that represents a shared conceptualization within a community of practice, i.e., it is socially constructed. In this way, the personal dimension is balanced both by the qualities found in the object itself, and the characteristics of the meta-model. The interpretation must start from empirical qualities but is completed by the experience of the analyst and the knowledge embedded into the meta-model.

Finally, the process of semiotic analysis and annotation is also a process of evaluation of the meta-model itself that may be modified and enriched in order to better represent concrete artifacts.

C. *Ontologies of narrative videos*

Quoting Gruber [44] an Ontology is an: "explicit and formal specification of a shared conceptualization about a given domain of interest". We are interested in the *knowledge level* (conceptualization and vocabulary) of an ontology rather than its *symbolic level* (formal representation). At this level Ontologies are *meta-models*. They provide the *concepts* (of entities, properties, and relationships) and the *vocabulary* that can be used to build models (e.g., descriptions) of specific things belonging to the considered domain of interest. By specifying the conceptual primitives, a meta-model implicitly defines the set of questions (called *competence questions*) that can be answered using the conceptualization. Ontologies are always incomplete and perspectival, i.e., they partially represent the domain of interest and they do it from a specific point of view that is related to the intended purpose of the ontology.

In this subsection, we briefly review some relevant conceptualizations of digital video with specific attention to those models aimed at annotation or indexing tasks or including narrative features. Before doing that it is worth attempting to characterize the nature of a narrative video viewed as a cultural artifact (e.g., a semiotic text) rather than a technological object featuring specific digital data structures, video format, compression algorithms and so forth [45]. Therefore, in this paper, a *video* is conceived of as a tangible object, a space-time dynamic configuration (a

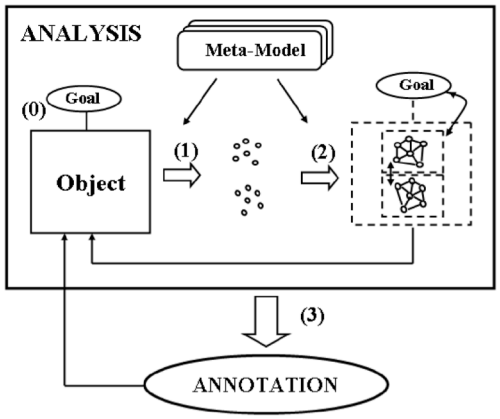

Figure 2. Analysis and annotation processes: 0) object selection, 1) decomposition, 2) re-composition, and 3) annotation steps. Object decomposition and re-composition are driven by a meta-model of the genre of object under analysis.

presentation) of visual and auditory signs that are inscribed within a material support using specific techniques, and are, therefore, available for an activity of visual and aural exploration and interpretation by a subject (the user or viewer/listener). A narrative video is a video that satisfies the conditions of narrativity stated by Ryan [35] or, alternatively, that is compliant with one of the definitions of narrative proposed by Grimaldi [46].

Attempts to exploit narrative theory for describing and annotating audio-visual resources date back to the early 90s. An example is the research work by Davis [47] at MIT who proposed a set of base categories for narrative video representation including action, character, object, mis-en-scene, cinematography. Davis explores annotation for video repurposing. More recently [48] illustrates a multilayered conceptualization of digital videos for indexing tasks that distinguishes among three main levels of analysis: layout, content and semantic index. For each level, a core set of descriptive concepts has been introduced. A similar proposal is the meta-model discussed in [49]. In this paper, video analysis and description is articulated into three dimensions: a spatiotemporal dimension representing the artifact at different levels of structural aggregation; a semantic dimension focusing on content (e.g., objects, events, plot structure) and an interpretative dimension inspired to film semiotics featuring different levels of interpretation (e.g., perceptual, cinematic, diegetic, connotative and sub-textual). Other relevant contributions are the research work by Stakelberg [50] and by Lombardo et al. [51] in the field of transmedia production and storytelling. The former contribution represents a very rich conceptualization (not yet an ontology) of transmedia narratives. The latter, developed within the CADMIO project, is focused on semi-automatic annotation of narrative artifacts and illustrates two computational ontologies devoted to characters and story respectively. Generative Theory by Greimas [52] is actually one of the

most widely used frameworks - in the field of contemporary semiotics - for the analysis of commercials video [53]. The framework distinguishes four interrelated levels of analysis: i) the textual level representing the concrete/physical manifestation of narrative content in terms of audio-visual features ii) the discourse level referring to thematic, figurative, rhetorical aspects iii) a shallow narrative level describing a story in terms of abstract roles (called actants) and narrative schemas (e.g., the narrative canonical schema) and iv) a deep narrative level that uses a specific tool called semiotic square to articulate deep semantic meanings such as narrative values (axiology). Signification unfolds by crossing these levels from shallow features of a video to the most abstract and deep ones. Although this framework is highly popular in the field of semiotic studies there are few attempts to transfer it (or parts of it) in the technical fields of multimedia analysis, design, and annotation. A notable exception is represented by the work illustrated in [54]. The authors exploit a framework of consumers' values proposed by Floch [55] within generative semiotics for automatic classification of videos and retrieval. Finally, the work by Bateman [56] and Tseng [57], although not strictly related to annotation tasks, are important contributions focusing on the application of Metz's semiotics of film and social semiotics, respectively, to the analysis of audio-visual artifacts.

What emerges from the analysis of relevant literature can be summarized as follows. A common objective of the considered works is the attempt to represent the complex and multilayered syntactic and semantic structure of video artifacts at different abstraction and aggregation levels. However, the number of layers, their meaning, the links existing between layers, and the conceptualizations proposed to describe the semantic content of each layer vary from one approach to another. There are differences and ambiguities in the use of core terms such as for example, narrative, narrativity, discourse, plot, story, as well as their conceptual meanings. As an instance, the concepts of story and narrative are often used interchangeably and defined in various ways along a continuum ranging from the easiest definition (e.g., a narrative is a representation of one or more events) to the hardest one (e.g., a narrative is an emotion-evoking and value-laden representation of one or more characters in a series of chronological events that are connected by causality or agency and which progress through conflicts toward a climax). As a consequence, meta-models of video artifacts vary in complexity and expressive power. The same concept of "event", which recurs in all definitions of story or narrative, is actually intended in different ways. Sometimes the term is used as a synonym of action or happening, i.e., something that - intentionally or unintentionally - occurs in time and space and produces a state transformation; other times it refers to the effect of an action (e.g., the state transformation itself); or to a state of affairs. These conceptual ambiguities hinder the development of usable and shared conceptualizations as the basis for interoperable ontologies.

It seems that no one of the forenamed approaches is able to exploit the benefits of contemporary semiotic conceptualizations in their full potentiality. This is the aim of the present work as discussed in the next sections.

D. Requirements for Semiotic Annotation

That said, we present a list of requirements for the design of a semiotic compliant narrative video annotation. Our aim is to integrate concepts belonging to the MPEG-7 standard with concepts that are drawn from contemporary semiotic fields, namely, visual semiotics, social semiotics, and generative semiotics. We have taken inspiration from classical theoretical papers in order to provide a conceptualization that is widely shared among experts in these fields. To this end we have aggregated the requirements into three main classes namely: syntactic, semantic and pragmatic requirements.

At the *syntactic level*, the conceptualization should enable the annotator:

- to structurally decompose the video presentation using different spatiotemporal *aggregation levels*. As an instance it should be possible to focus on single regions within a representative frame; on moving regions crossing a sequence of adjacent frames; or look at the video as a temporal sequence of more aggregated entities such as shots, scenes, sequences, and episodes. We need specific mechanisms for the univocal identification of the anchor of annotating data;
- to describe the video using multiple structural decompositions. As an instance, a decomposition representing the video as a sequence of shots, another as a sequence of scenes and another one representing the same video as a sequence of homogeneous sound objects (e.g., music, silence, speech, effects) or combination of sound objects;
- to relate together structural entities belonging to the same or to different decompositions (e.g., to represent the relationships existing between adjacent shots; between shots and scenes or between scene boundaries and sound objects);

Structural decompositions constitute the scaffolding for the semantic level. They allow the annotator to represent compositional (or organizational) meaning that tells us what goes with what, what smaller units belong to what larger unit, how parts are related together, and how semantic meanings are distributed across the whole video.

At the *semantic level*, the conceptualization should enable the annotator:

- to associate basic kinetic and plastic features (e.g., shapes, colors, positions, textures, sizes, cinematic movements, visual contrasts, rhythm) to visual regions or structured groupings of regions within a frame or across frames [58]. To describe spectro-morphological features of sound objects (e.g., time features such as amplitude, envelope, loudness, tempo, and spectral ones such as pitch, timbre, harmony) [59].
- to associate a semantic construct (e.g., a figurative sign such as an object, subject, action, event, or abstract concepts such as goals, deep values, emotions) to visual or auditory fragments, and, indirectly to the plastic or spectro-morphological features that characterize them [60]; to link the semantic constructs by several types of relationships

(e.g., spatial, temporal, logical, rhetorical, typological, mereological, causal, teleological relationships);

- to associate dramaturgical patterns (e.g., the canonical narrative schema by Greimas [61], the Hero's Journey by Campbell [62], the Dramatic Arc or Three Acts Model [63]) to visual or auditory segments of the video, and, indirectly, to the semantic constructs and expressive features that represent these segments at the syntactic and semantic/figurative levels;

Semantic annotation allows the annotator to describe representational (or ideational) meaning that tells us what recognizable existents are represented, who is doing what, to whom, and with what means, what is happening and what is related to what and how.

At the *pragmatic level*, the conceptualization should enable the annotator:

- to identify the images, called simulacra, of all the participants involved in the production and use of the video (i.e., addresser, addressee, narrator, observer, actor) that are inscribed within the artifact and specify their interrelationships [64];

- to describe the kind of relationship the designer/author of the video wants to evoke between the various subjects inscribed within the video and the intended user.

Pragmatic annotation allows the annotator to represent interpersonal (or orientational) meaning that refers, for example, to social distance and intimacy, image acts and gazes, narrative engagement and power relationships [65].

Table I exemplifies focal questions that can be used to direct the attention of the analyst to key perspectives and issues related to the three main types of meanings taken into account by social semiotics [39][66].

Finally, the conceptualization should provide the annotator with a set of relationships that can be used to link all the above aspects together in order to build the desired means/ends ladder: deep values with the storyline, the elements of the story with discourse segments and expressive qualities; expressive qualities with interpersonal meanings.

TABLE I. EXAMPLES OF FOCAL QUESTIONS RELATED TO TYPES OF MEANING AND LEVELS OF ANALYSIS

Type of meaning	Focal questions (examples)
Compositional (Syntactic level)	What constitutes the whole video text under analysis? How do you know what is and what is not a part of it? What is the most salient visual/auditory element? Which parts are related together? Which parts are separated? How are parts related together?
Representational (Semantic level)	What recognizable actors or participants in actions and relationships are presented? Persons? Concrete things? Abstract ideas, qualities? What relationships are presented among these participants? In what common or shared action, event, happening are they presented? What are le locations, settings, causes of, temporal location of, these relationships, actions, and events? How are actions, events synchronized or sequenced in time? What logical, rhetorical relationship among actions, are presented? What emotions are described? How emotions evolve over time? What values drive the story?
Interpersonal (Pragmatic level)	Who is the intended viewer/receiver of this video text? What internal features index anticipated qualities of the receiver? What qualities of the sender/author of the video are indexed by internal features? How does the video position the sender relative to the receiver? In a relation of power? Dominance? Intimacy? Formality? What does the video request or demand of the receiver? How? How does the video index the stance of the sender (or any voice it projects) toward the text itself? Toward the receiver? Toward its own representational content?

IV. THE META-MODEL

We propose an informal conceptualization - a meta-model not yet a formal ontology - that provides a core set of basic descriptors that can be used to perform a semiotic annotation according to the above requirements. It has been organized into four main related modules (called boxes): the text, discourse, story, and agent boxes. Figure 3 (left) shows a conceptual schema of the modules and their inter-relationships. In this schema, and in the following ones, we use different graphical representations (i.e., types of arrows) to denote three main relationships: hyponymic (i.e., *sub-class-of* relation), meronymic (i.e., *part-of* relation) and generic etherarchical (e.g., associative) relationships between core concepts.

A. The Text Model

By Text we mean a concrete manifestation of a narrative, i.e., a complex fabric of signs belonging to different semiotic modalities (e.g., moving images, sounds, written words) and conveying narrative content to the user's interpretation.

The text model is a key issue of our conceptualization since it relates *content* with *expression*, according to the schema reported in Figure 3 (right). A Text has a T-Structure composed by T-Segments and relations (T-SegmentRel). Segments have been classified into several classes following MPEG-7 [29]. Relations include spatial and temporal relationships as described by Allen [67] and Galton [68].

A text segment is linked to a discourse segment representing its content and points to a set of sensory qualities and quality relations representing its expression. As sketched in Figure 4 (left), we distinguish between tonal - static, persistent - and rhythmic - dynamic, transient - qualities. Qualities may have associated facets and quantity spaces, i.e., domains of possible values. Color, for example, has hue, saturation and lightness as facets, and values in YCbCr or RGB color spaces. Sensorial qualities can be related together by several kinds of relationships such as contrast, affinity, and completion producing higher aesthetic effects such as salience, separation, connection, balance,

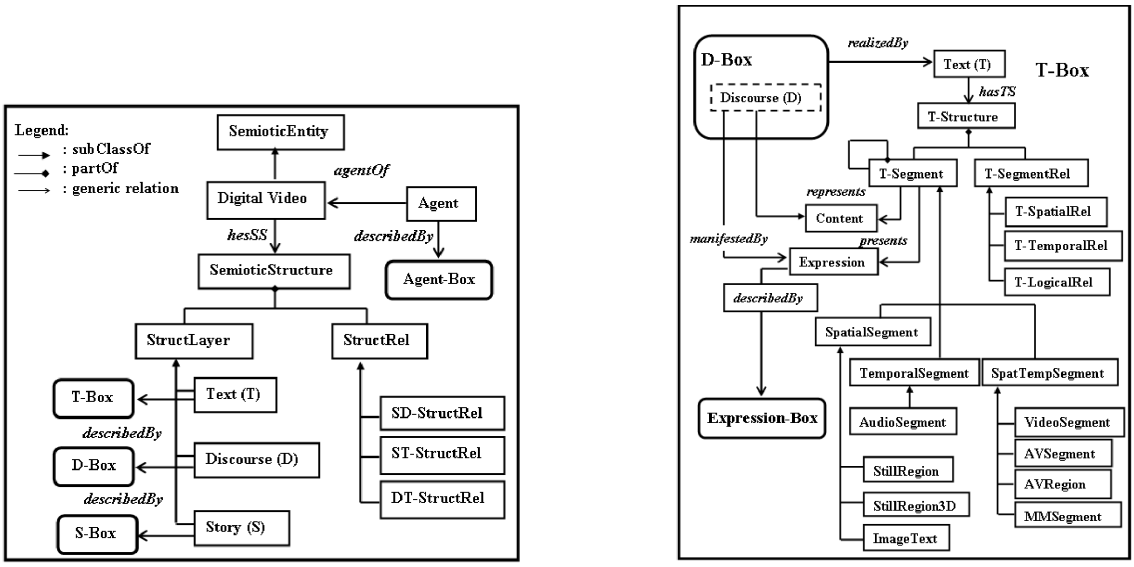

Figure 3. An overview of the proposed meta-model (left); the Text module (right).

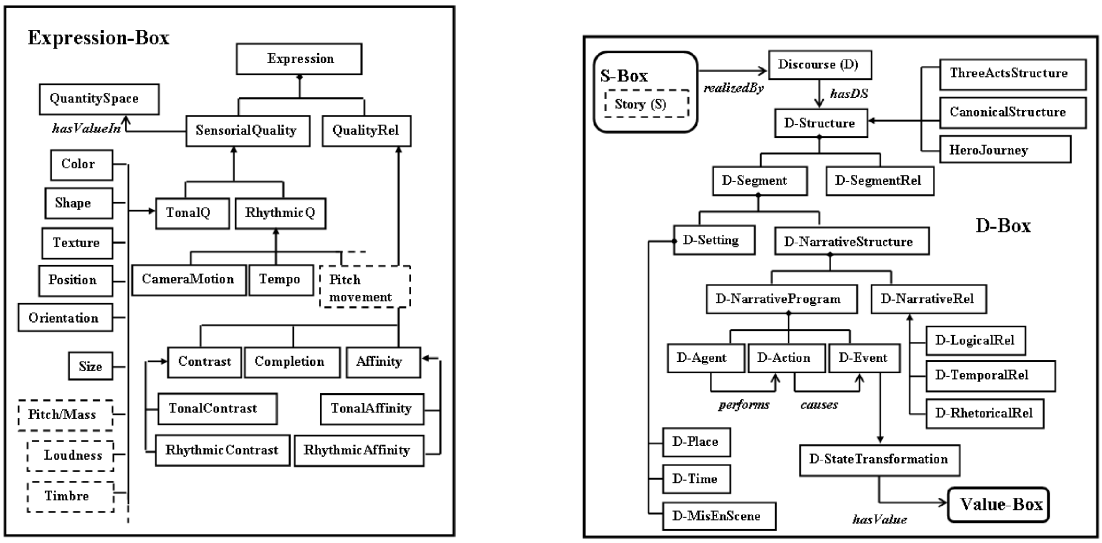

Figure 4. Expression (left) and Discourse (right) modules.

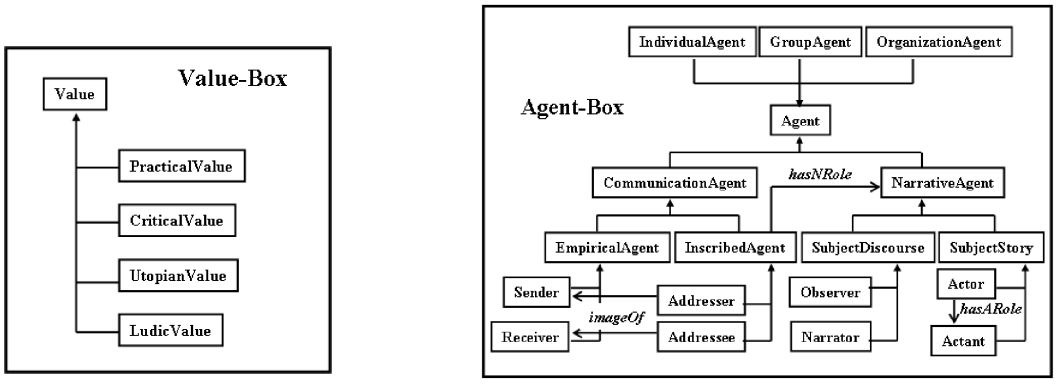

Figure 5. Value module (left) and Agent module (right).

symmetry, order, complexity or affective states (e.g., emotions, feelings, mood or atmospheres).

From an experiential point of view, the text module is intended to capture the sensorial experience the user has when viewing/hearing a narrative video. This experience is strongly related to the visual and auditory qualities of semiotic materials. Plastic forms, in particular, represent an autonomous level of signification of the video that can be used to make sense [58].

B. The Discourse Model

Discourse (also "syuzhet") represents the content of a narrative, i.e., *what* is narrated by the text (e.g., the events of a story) and *how* it has been done (e.g., plot, rhetorical structure, stylistic choices). As shown in Figure 4 (right), a Discourse has a D-Structure composed of D-Segments and relations (D-SegmentRel). A D-Segment consists of D-Setting and D-NarrativeStructure, the former specifying place, time and mis-en-scene; the latter specifying a structured set of narrative programs. A *narrative program* is defined in terms of D-Agent performing a D-Action and causing a D-Event intended here as a kind of state transformation. Narrative programs are related by logical, temporal or rhetorical relationships [69]. The concept of narrative program is a key concept of the proposed conceptualization. It is borrowed from Greimas [52][61]. It allows the annotator to represent "who does what": how actions are distributed among agents (see Figure 5, right) as well as the effects produced by the actions be they external physical changes, or internal cognitive, or affective transformations of subjects. As an instance, the effect of an action can be the acquisition or loss of a concrete object or person or of a more abstract entity such as freedom, knowledge, happiness. In any case, transformations can be interpreted as conjunction or disjunction processes of a subject with an object charged with values (object of value). Values have been classified according to Floch [55] into four classes namely, practical, critical, utopian and ludic values (Figure 5, left). Practical values refer to utility, usefulness; critical values to convenience, performance, quality; utopian values to identity, reflection, social relations, and ludic values to surprise, madness, astonishment, irony and pleasure including aesthetic pleasure. The selection of specific values in the construction of the story allows the author to realize specific marketing strategies. The discourse model is flexible enough for allowing the modeler to represent single actions, aggregates of actions occurring simultaneously or in sequence, and aggregates of aggregates (an entire story). Actions may be performed by the same agent or by different agents; they may occur in the same setting or in different settings. For narrative texts, several discourse structures have been proposed in literature such as [62][63]. Each of them decomposes a story in phases that are bounded by specific kinds of events and imposes a set of specific constraints on the narrative structures constituting each phase. As an instance, each dramatic arc of a Dramatic Arc model is strictly related to specific rhetorical relationships as discussed in [70]. Further constraints exist between discourse and text. Major events of a dramatic discourse structure such as the inciting incident, climax, resolution, etc. should be expressed, at the textual level, by appropriate sensorial qualities. Dramatic tension evolution as well as affective events (i.e., internal state transformations of main characters) should be sensed first, through appropriate visual images or melodic contour, then conceptualized. This imposes an internal coherence between narrative content and expression that is at the core of the semantic ladder we want to establish between different forms of meaning articulation.

From an experiential point of view, discourse is intended to capture cognitive and affective aspects of the interaction. This experience is strongly related to figurative features, narrative structures and their spatial and temporal configuration. Figurative forms (formants), in particular, represent another level of signification of the video that is superimposed to plastic qualities [58].

C. The Story Model

Most of the literature [71] understands "story" (also "fabula" or "histoire") to be the events that constitute the content of a narrative. Since the story is embedded within discourse we do not specify a new model for it but use the conceptualization provided by MPEG-7 with minimal variations [29]. This conceptualization represents basic components of stories namely, existents (Object DS, AgentObject DS), events (Event DS), abstract concepts and states (Concept DS, SemanticState Ds), and settings (SemanticPlace DS, SemanticTime DS).

D. The Agent Model

The agent model takes inspiration from Enunciation Theory [64]. This theory suggests that every communicative artifact contains, inscribed within it, an image or simulacrum (i.e., a constructed representation) of the actual sender and receiver. These images are called the *addresser* and *addressee* respectively. They are embodied in the artifact in the sense that they are analytically available to the critic by means of a close analysis of the artifact itself.

The agent model is aimed at representing the addresser e addressee and their relationships with the subjects of the story and discourse as they are prefigured by the product. More specifically, agents - individuals, groups or organizations - have been classified into two main classes: communication agents and narrative agents (see Figure 5, right). The former class includes the actual sender/receiver (called empirical agents) and their simulacra, the addresser and the addressee. The latter comprises the subjects of the story (e.g., actors and actants) and the subjects of discourse (e.g., observers and narrators). An *observer* is an agent responsible for physical focalization. It establishes the spatial position of the viewer with respect to the story world, for example, by selecting, at the expressive level, specific shot sizes, camera angles, lighting conditions. A *narrator* is an agent responsible for the cognitive and affective focalization. Actually, the viewer/spectator is invited not only to perceive what is told by the video from a spatial position but, more importantly, to interpret what is

happening from a specific conceptual point of view (cognitive perspective taking) and to emphasize with some characters of the story (i.e., to understand and share their perceptual, cognitive, and affective status). Notice, that the D-Agent concept belonging to the discourse box can be equated to the NarrativeAgent of the Agent-Box (e.g., to an Actor or Narrator) thus realizing a connection between these two conceptualizations. The agent box is intended to capture relational experience, that is, possible relationships (e.g., social distance, power relationships, engagement) between the sender/receiver of the advertising message and narrative/discourse agents arising from their images within the text. As an instance, a company (a sender) may be associated with a visual or auditory segment that represents the company visual or auditory logo within the text. The logo plays the role of the addresser. An actor of the story may represent the user (addressee) playing a specific actantial role (e.g., the hero of the story). An observer may adopt the physical position of an actor of the story thus showing the story world through the eyes of that actor. Analogously, an actor may be associated with a narrator (a storyteller) and so forth. The "distance" between the actual receiver (the viewer) and the actors of the story is a function of two main factors: i) the distance existing between the receiver and the observer, and ii) the distance between the observer and the actors. The former can be reduced, for example, by letting the viewer play the role of an observer, i.e., by giving him/her the control of the camera such as in interactive videos; the latter by letting the observer represents the story world and events from the vantage point of an actor of the story. Seeing events from the point of view of a story's character makes the viewer aware of the character's perspective and his or her interpretation of events, and moreover, of the character's motives in relation to events and other characters. By adopting a character's perspective the viewer can understand and relive the character's emotions. This is essentially empathy, a viewer's mirroring of a character's emotional experience. Relational experience is strongly related to processes of narrative engagement such as cognitive perspective taking, empathy, presence, flow, and involvement [11]. Narrative experiences that are more engaging should result in more enjoyment, i.e., fun and pleasure. Therefore, it is important to represent these features in order to be able to compare products and evaluate their respective hedonic effectiveness.

V. THE CASE STUDY

We illustrate an example of manual annotation of a narrative commercial video clip. The aim is evaluating the feasibility of semiotic analysis and annotation. To this end, we start by illustrating the annotation tool we have chosen, then the procedure we followed and obtained results.

A. The annotation tool

A critical comparison of annotation tools has been presented in [72]. Among them, the EUDICO Linguistic Annotator (ELAN) shows several advantages including its relatively shallow learning curve and user-friendly interface [73].

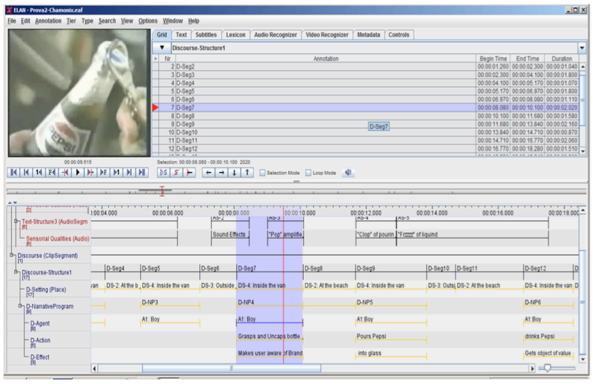

Figure 6. A screenshot of the ELAN annotation tool.

Figure 6 shows a screen shot of the ELAN interface. Annotations in ELAN can be grouped into multiple layers (called tiers) that are part of tier hierarchies. Annotation values are Unicode characters, and the annotation document is saved in an XML format based on the ELAN XML Schema. The tool can be easily connected with the *Praat* software for the analysis of the audio component of the video in the temporal and spectral domains [74].

ELAN allows the annotator to define a vocabulary of descriptors at the beginning of the process; a more recent version of the environment, called ONTO-ELAN, is capable of importing an ontology to be used in the analysis and annotation [75]. A limitation of the tool is that is not possible to annotate single regions within frames.

B. Annotation of a Pepsi Cola clip

The clip produced in the late 1980's is based on the body copy "Pepsi Cola. The choice of a new generation". It lasts 29.4 s at a frame rate of 30fps [76]. The story can be summarized as follows:

A delivery van of Pepsi–Cola reaches a crowded beach. The young driver gets out, opens the side door and switches an amplifier on; two loudspeakers emerge from the roof. The boy brings a bottle of Pepsi close to the microphone; uncaps it; pours the liquid into a glass and drinks emitting an "Ahhhh" of pleasure. People attracted by the puffing of gas and boy's expression rush to the van to quench their thirst (and buy the product!).

A characteristic of this clip that makes it a candidate for semiotic annotation is that it stages a narrative telling "the process itself of advertising and persuasion". The process includes three steps (see Figure 7): 1) insert the potential viewer into a familiar situation (the delivery van of the Pepsi Cola reaches the crowded beach); 2) draw viewer's attention on a positive, euphoric experience of consumption (loudspeakers attract people; the experience of the boy drinking Pepsi is communicated both visually and auditory); 3) activate into the viewer a desire to live a similar

experience assuming a purchase behavior (people rush to the van to buy the Pepsi).

The procedure followed for the analysis and annotation of the clip consists of the following basic stages:

Stage-0. *Annotation of the whole video with its genre and purpose.* The whole clip is represented by an alignable annotation tier linked to a single segment (ClipSegment) representing the root of a hierarchical multi-tiers decomposition. Alignable tiers in ELAN are directly linked to the time axis of the audio-visual and can be further segmented. This tier is annotated with the multimedia genre (video commercial) and intention/purpose: "To advertise Pepsi Cola; to represent a persuasive process".

Stage-1. *Textual decomposition of the video: identification of the visual and auditory textual structures in terms of audio and video segments and their relationships*. The ClipSegment is represented by several text structures (T-Structure): some of them are associated to the visual modality, others to the aural one. A structure (T-Structure1) decomposes the ClipSegment into a sequence of T-Segments representing individual visual shots. A further text structure (T-Structure2) annotates special transitional effects/edits like fades, dissolves, overlayed text. Another structure (T-Structure3) decomposes the ClipSegment based on continuous sequences (T-Segments) of homogeneous sound objects (called Audio Segments, AS). The sequences include silence, speech, environmental sound, and effects. In more complex cases it may be necessary to devote a separate text structure to each constituent of a complex audio sandwich. Notice how the sonic effect "Frzzzz" of the liquid while the boy uncaps the bottle covers several adjacent shots of the video, i.e., the two structures are not aligned in time.

Figure 7. Main steps of the process of advertising and persuasion: 1) insert the viewer into a familiar situation; 2) draw attention on a euphoric experience of consumption; 3) activate a purchase behavior.

Stage-2. *Textual annotation: association of expressive descriptors to the structures found in the previous stage.* During this stage, a set of referring annotation tiers are introduced and associated to previous visual and aural structures to annotate single shots, transitions, and sound objects with tonal and rhythmic sensorial qualities according to the conceptualization shown in Figure 4 (left).

Stage-3. *Discoursive decomposition of the video: identification of the discourse structure of the video in terms of discourse segments and their relationships.* The ClipSegment is represented by one or more discourse structures (D-Structure), based on scene analysis. A scene (D-Segment) is defined as a - not necessarily continuous - sequence of frames representing a narrative situation characterized by a stable setting (i.e., place, time and mise-en-scene). In the case under consideration, we use a single discourse structure (D-Structure1), which is decomposed into 17 D-Segments. Scene boundaries correspond to changes in settings from outside to inside the Pepsi Cola van and vice-versa.

Stage-4. *Narrative segmentation: each discourse segment is further analyzed in terms of a setting and a narrative structure.* Each scene (D-Segment) is annotated by a narrative structure composed of narrative programs and their logical and temporal relationships.

Stage-5. *Annotation of narrative programs.* A set of referring annotation tiers are introduced and associated with previous narrative structures to annotate single narrative programs. For each narrative program, a set of tiers is used to separately describe the main components of the program namely the actor, the action, and the event. The event is further elaborated in terms of state transformation and value. In the example under consideration, D-Segment7 and D-Segment9 (a scene inside the van) is annotated by a narrative structure composed by the temporal sequence of two narrative programs. The first program (D-NP4) refers to the boy (D-Agent) grasping the bottle of Pepsi (D-Action) thus making the user aware of the brand (D-Event). The second narrative program (D-NP5) refers again to the boy (D-Agent) who uncaps the bottle and pours drinks content (D-Action) thus getting the object of value, i.e., the product (D-Event).

Stage-6. *Relational analysis and annotation.* The root segment (ClipSegment) is analyzed in order to identify the markers of addresser and addressee. In the Pepsi Cola clip, the bottle including logo and trademark represents the addresser (i.e., the brand Pepsi Cola). The boy and the people approaching the van represent the addressee. Three types of relationships are shown: i) between the viewer (represented by the boy) and the product/brand ii) between the viewer (represented by people in the beach) and the boy that is consuming the product and iii) between the viewer (represented by the real user) and the people on the beach who are experiencing a growing desire to drink a Pepsi. A set of further tiers have been introduced and linked to the

ClipSegment to implicitly represent relational analysis by annotating actors' gazes, kind of shot, vertical and horizontal camera angle. As already said, these features are related to engagement, social distance, power and involvement relationships, respectively [65]. In the same way, the tone of voice in speech, sound perspective, volume, can be used to represent various degrees of intimacy or distance between the characters of the story (and indirectly the brand) and the user.

Several temporal relationships among annotations belonging to different tiers are implicitly described through the relations existing between their corresponding tiers. As an instance, all referring tiers associated with the same alignable tier inherit its time decomposition. As a consequence, their annotations are automatically time aligned. Figure 8 summarizes the resulting decomposition and annotation structures. The figure also shows the articulation of meaning across the various tiers of the annotation hierarchy. Compositional meaning is represented by the decomposition of the video in terms of shots and sound objects and their associated visual and auditory qualities. Representational meaning is represented by scenes and their associated narrative structures (i.e., settings, narrative programs, relationships between narrative programs). Finally, interpersonal meaning is represented by relational annotations (e.g., social distance, engagement, simulacra) associated to video shots.

VI. DISCUSSION

In this section we view semiotic annotation in a broader context by relating it to research work made within Research through Design (RtD) [77][78], Philosophy of Technology [79] and Interface and Interaction Criticism [80][81][82][83]. The aim is to highlight connections with these fields and potential contributions.

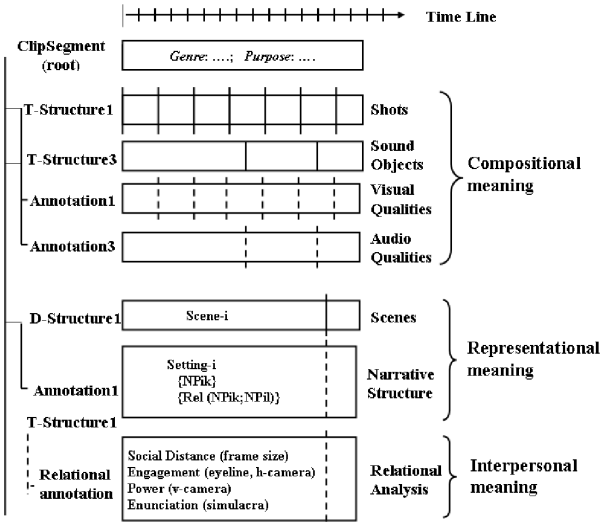

Figure 8. The hierarchy of decomposition tiers used in the Pepsi Cola example and their associated meanings.

A. Semiotic annotation and design knowledge

Semiotic annotation represents a kind of *intermediate level knowledge* [84] (see Figure 9). It is *in-between* the annotated object (the video) and a conceptualization of the object (the meta-model or ontology of the video). On the one hand, the annotation, using the concepts of the meta-model as descriptors $\{d_{ij}\}$, explicates the conceptual model adopted by the designer and, indirectly, the aspects that are implicitly deemed important and relevant for the design of that artifact according to that meta-model. On the other hand, the annotation, by indexing specific parts of the concrete artifact, shows how the descriptors - and thus the concepts of the meta-model - have been instantiated $\{v_{ij}\}$ in *that* particular artifact. As a consequence, the annotation reveals the specific point in the *design space* (i.e., the space of all possible alternative instantiations associated to the adopted conceptualization) occupied by that artifact $\{(d_{ij}=v_{ij})\}$. In this sense, we can say that the design knowledge embedded in the video is unfolded by the annotation that can be seen as a particular type of interpretation of the object made according to a meta-model (e.g., an ontology) of the object itself.

The availability of design knowledge provides several benefits for the designers and the users as well. It allows explaining the way a specific video works from a communicative point of view: how meaning is constructed - in *that* video - by the interplay of several elements located at different levels of the means-end semiotic ladder. For designers, in particular, the annotation affords extraction of design knowledge in order to reuse it, evaluate its internal coherence or take inspiration from it in developing new products.

They can exploit the annotation to compare two or more videos during the phase of competing analysis in order to understand *why* they are designed the way they are and *how* they differ from one to another. They may search for redundancies and variations; or aggregate videos on the base of similarities in the way they function (i.e., how they instantiate the meta-model) with the goal of constructing

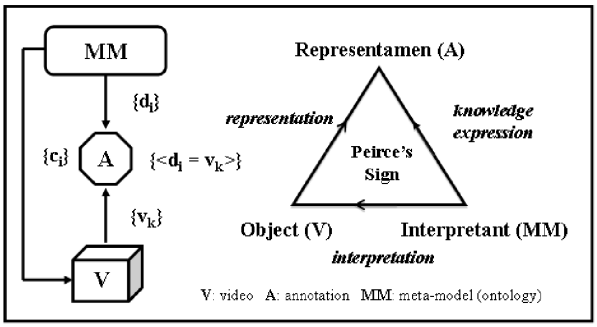

Figure 9. Annotation A is in-between the video V and the metamodel MM. The metamodel provides concepts {ci} and descriptors {di} for interpretation and annotation. The video provides specific values {vi} to instantiate descriptors. The model echoes Peirce's concept of sign [38].

portfolios [85]. Consider, as an instance, the following production scenario.

Scenario-1. Samantha is a video producer. She has to design and develop a narrative video commercial for a client. The main requirement is that the video advertises a new line of products and communicates the core set of values that constitute the brand identity of the client's firm in a way that is innovative with respect to the videos of previous ad campaigns. Different issues are involved in this kind of design problem. First, Samantha has to know what does it means to design a narrative text and the various ways a product can be included in a narrative (e.g., as a prop in the setting, as the main character of the story, as an helper/instrument that can be used to achieve some abstract object of value or as the object of value by itself). Second, she must be able to understand how existing videos communicate the brand identity through their expressive and narrative features. More specifically, how values have been inscribed into the videos. To address these issues it is important to make explicit the brand design language used in the various products.

Therefore, Samantha downloads from the company net the videos belonging to previous campaigns and analyses their associated semiotic annotations in order to understand how meanings have been articulated and distributed across the various semantic layers constituting the artifacts. In particular, she wants to identify which design variables have been consistently used to communicate brand personality and values in the various products under analysis. Therefore, she compares the annotations searching for similarities and differences. This comparison is inspirational. Samantha discovers that brand values have been mainly communicated by the audio tracks, e.g., by a leitmotif that recurs in all the videos with different orchestrations and sound design [13]. So she decides to design a new video where the same values are communicated by images (story) instead of by the music.

In this scenario, semiotic annotation has been exploited to generate insights particularly useful for addressing a design problem.

Semiotic annotation can be useful also for generic users in order to make more informed choices, i.e., to better understand if a product is adequate with respect to their values, needs, desires, preferences. Moreover, the annotation can be an auxiliary to educate to critical analysis and evaluation, contributing to cultivate/enhance users' perceptual sensitivity (making people better at learning to see and hear) and interpretative skills. Consider the following use scenario

Scenario-2. George teaches a course of Multimedia Design at his University. During his lectures, he exemplifies theoretical claims using existing videos. The search and selection of appropriate videos is a difficult and demanding task since they must be carefully analyzed before the lectures in order to understand if they are (or not) suitable to the intended ends. Again design knowledge could be very useful to understand how they function, and why they function in the way they do. The next lecture will be devoted to introduce and discuss various types of narrative structures. So George exploits semiotic annotation to search the web for commercial videos that represent an instance of the Hero's Journey or Greimas' canonical model. The annotation is then used to select the segments of the video that represent specific phases of each considered model to be shown to the students. A critical discussion is then started to assess the degree of fit among the visual and auditory modalities and their coherence with respect to the representational, interpersonal and compositional meanings expressed by each segment.

In this scenario, semiotic annotation has been exploited to search for appropriate videos with respect to their discursive structure and to support discussion and critical judgments. Students can, therefore, improve their understanding and interpretative skills. At the end, they can be helped to better master their craft in designing mediated experiences.

B. Semiotic annotation and technological mediation

An important concept arising from the field of Philosophy of Technology is *technological mediation*. As claimed by Verbeek in [86] the use of a particular technology or artifact affects the relation between the user and the world in two ways: by a process of transformation of perception of the world (*mediation of perception*) and by transforming the user's praxis or action in the world (*mediation of action*). The effect of mediation of perception is realized by an amplification or reduction of the experienced aspects of reality while the effect of mediation of action is realized towards an invitation or inhibition to perform certain actions instead of others.

Although all artifacts involved in the semiotic annotation (i.e., the meta-model/ontology, the video and the annotation itself) exert some kind of technological mediation, we are interested, here, in the role played by the ontology.

The ontology realizes a mediation of perception process toward its conceptualization and axioms. The conceptualization brings some aspects of the world into sharp focus at the expense of blurring other aspects. It invites the user to look at the reality through a specific type of glasses that amplify or reduce some experienced aspects of the domain of application. The mediation of action is strictly linked to the ontology's competence questions since they specify the functional aspects of the ontology: its scope, and possible uses. In other words, the ontology invites the user to enquire reality by making certain queries instead of others. As an instance, viewing (and conceptualizing) a video at the computer level as a technological artifact (e.g., a set of digital resources and execution programs; a functional tool) is different from viewing it (also) at the cultural level as a semiotic text (e.g., a cultural interface, a work of rhetoric or a mediator of experience). In the second case, a set of complex aspects related to meaning, ethics, etc. emerge that are usually ignored at the computer level. As an instance, if the artifact is a tool, ethics is irrelevant since the ethical agency is situated in the user. If ontologies can affect people's behavior and relationship with the outside world, the design

or adoption of an existing ontology is an ethic activity and ontology itself is a materialization of values and ethical choices [87].

When ontologies are pushed into artifacts or are embedded in working environments/applications as meta-models they shape them and guide, in this way, the user's experience and expectations. Their mediation effect is thus indirect; it occurs through the artifact in which the ontology is embedded and merges with the mediation effect of the artifact itself. The mediation effect is usually made transparent (i.e., not visible) during use. Transparency of the meta-model during video use is important since it enhances the possibility of uncritical and intense processing that is at the base of transportation effect and narrative persuasion. However, recent studies in the field of Philosophy of Technology and Persuasive Design claim that mediation effects should be made opaque and comprehensible to users [88][89]. A semiotic annotation may support this claim by unfolding contextual information that is information *around* the video (e.g., purpose, assumptions, articulation of meanings, and intended effects). By separating the narrative (video) from the information about how the narrative has been constructed (the annotation) we are able to guarantee both ease of use (transparency of use) and control of mediation effects (opacity of context) that is a kind of semi-opacity of the video artifact. This is useful for the user in order to better understand how the video has been designed to satisfy the author's intended goals, why it functions as it does, what rhetorical mechanisms are at the base of its persuasive and informative functioning, what sort of culture it will encourage or resist. In this way, semiotic annotation may contribute to the diffusion of a critical attitude toward video commercials (and audio-visual products in general) and a greater awareness of the social effects this kind of products may produce. This opens to the last issue we wish to address.

C. *Semiotic annotation and criticism*

Criticism refers to "an expert of a given domain's informed exercise of judgment" [82]. Semiotic annotation has much to offer to the discipline of interface and interaction criticism [80][81]. Quoting Bardzell [81]: "[by Interaction Criticism] we mean rigorous interpretive analysis that explicates how elements of the interface, through their relationships to each other, produce certain meanings, affects, moods, and intuitions in people that interact with them". Then Bardzell moves into the nature of the concept "rigorous": "...we say rigorous to stress that interaction criticism, like the best film and literary criticism transcends anything-goes subjectivism and offers instead systematic, evidence-based analyses of subjective phenomena ..."

It should be evident from the above citation that semiotic analysis and annotation converges and largely overlaps with the notion of criticism [83].

Another issue that is strictly correlated with criticism, concerns the "value" of an artifact intended, here, not in economic terms (e.g., exchange value, use value) but more specifically, as a quality of the artifact that concerns its "inner logic" or the "human good" as discussed by recent theories about aesthetics [90] or cultural quality of new media [91]. We suggest that the annotation can change the value of an artifact. To support this claim we will draw on Danto's concept of *transfiguration* in his Aesthetic Theory [92]. As it is known, the issue Danto wants to address is the relationships between art and reality. In particular, the problem can be stated as follows: how can it happen that two objects that are phenomenologically indistinguishable - e.g., the Brillo Box artwork by Andy Warhol at MoMA and a similar object, the Brillo box containing soap pads at the supermarket - have so different values. The answer provided by Danto is that an object can change its value by means of a transfiguration process that is a particular kind of interpretation - called artistic interpretation - made in a context or atmosphere of art theory (the Art World). The interpretation does not change the physical appearance of the object but its ontological status: it changes the object into a work of art.

We can try to apply these concepts to annotated videos. Here the video is the object and the annotation is a semiotic interpretation playing the role of transfiguration. The interpretation is based on a theoretical background, the ontology that can be mapped to Danto's atmosphere of theory (the context). In our case, the informal ontology illustrated in Section IV takes inspiration from several semiotic theories (e.g., visual semiotics, enunciation theory, narrative semiotics, and social semiotics) as discussed beforehand. These theories provide a theoretical scaffolding for the ontology and the annotation as well. The video *with* the annotation has a value greater than the video alone because the annotation provides a surplus of information that can be used to interpret and understand the video, that is, to attribute some meanings that are not empirically available from the video alone. The value of the video is thus embodied in the product but it is external to it (i.e., it is in the interpretation materialized by the annotation) and cannot be captured by only looking at the tangible object. As a consequence, the semiotic annotation should be considered as a constitutive part of the video. They are strictly correlated and mutually informing: the annotation "illuminates" the artifact giving it value; the artifact provides the ground for and exemplifies the annotation [93].

What is interesting in Danto's theory is that the value is linked to interpretation: not a generic interpretation, however, but a theoretically grounded one. Moreover, Danto, stresses the fact that the interpretation must be *appropriate* to the formal and material characteristics of the objects. In other words, the meanings must be embodied in the form and materiality of the object. The modality of this embodiment affects the quality of the artifact. So, for example, the internal coherence is a kind of inner logic and a dimension of quality because it refers to how meanings are distributed among semiotic materials and how they fit together. This discussion opens up interesting research perspectives and poses critical problems to automatic annotation. If the value rests in an appropriate interpretation then it cannot be captured by automatic procedures that refer only to existence, i.e., to tangible and empirical features. It

must be provided by some human. It is not strictly necessary that the interpretation is the designer's one; obviously, the designer is in a privileged position to provide this kind of knowledge since she is the main source of design choices. However, we can envisage other possibilities - and associated annotations - such as the exploitation of multimedia critics, semioticians, exhibition curators, and commentators. If the interpretation is not the designer's one, it should be, at least, an interpretation that the designers would consider as a possible one. Other important sources of information are represented by the script and the storyboard of the narrative.

VII. CONCLUSIONS

In this paper, we have presented an annotation method for narrative videos, along with a worked-out case study in the field of advertising and brand communication.

The standpoint is that of video production. We are interested in how purpose and meanings are intentionally constructed and articulated during the premeditate and message construction phases; how they are inscribed within the artifact and materialized through semiotic resources; how these meanings can be used to annotate the video for content retrieval, filtering, and content reuse.

The annotation method exploits a semiotic meta-model (an informal ontology) of the video genre. The meta-model articulates the semantic content of the video according to a set of interrelated layers each addressing a specific kind of meaning: structural decomposition, visual and auditory expression, narrative content (i.e., discourse and story), values and interpersonal relationships. A core conceptualization has been provided to represent each layer as well as the relationships existing among layers in a domain-independent way. These relationships are responsible for the global unity and consistency of the artifact. They form a semiotic ladder covering the conceptual gap existing between the shallow and concrete audio-visual qualities and the deep and abstract constructs related to discourse, story, and axiological values.

We highlight some features of the meta-model that on the base of our actual knowledge seem novel. First, the meta-model exhibits a strong separation of concerns that results in a modular conceptual architecture. However, what is important are the links existing between the various conceptualizations since these links are at the base of the articulation of meaning inside the product and of its internal coherence. They reflect design knowledge. Second, the introduction of a discourse model allows the annotator to explicitly aggregate basic story elements in a meaningful way in order to reflect important aspects of narrative products such as their rhetorical and dramaturgical structures that are ignored by traditional metadata standards. Moreover, the use of the concept of narrative program as the building block of discourse forces the annotator to focus simultaneously on several interrelated concepts - such as agent, action, state transition (event), object of value - that are strictly related to narrativity. Third, the introduction of a rich classification of agents inspired by Enunciation Theory provides the vocabulary for describing various roles involved in the development, communication, and use of an audio-visual product as well as their interrelationships.

The role of the meta-model is to guide the annotator during semiotic analysis by focusing her attention on specific features and relationships. Furthermore, it provides the vocabulary and concepts used as descriptors during the annotation process. Semiotic annotation is different from pure keyword or concept annotation. The task is not simply to attach subjective comments, notes, pre-existing opinions or remarks to audio-visual segments but to unfold the generative process of sense-making inscribed within the product.

We have briefly discussed the potential contribution of this kind of annotation for research in the fields of Research Through Design and Technology Mediation. The annotation, by unfolding the design knowledge inscribed within the product, is proposed as a viable means for communicating design thinking in a descriptive yet generative and inspirational fashion. It supports moving from theory/ideal (the meta-model) and practice/concrete (the artifact) and vice versa. Moreover, it makes visible the design decisions that were taken during the construction of the message opening the door to the assessment of technological mediation. This prompts designers and users to put particular attention to this issue.

Much of the job of semiotic annotation resonates with criticism. At the heart of criticism is the attempt to explain abstract meanings and impressions by referring to the properties and forms of the artifact or the way the artifact has been produced. This is the case, for example, of past research in computational media aesthetics [94] where film grammar (i.e., cinematic techniques used during production) is used to explain high order qualities of video resources such as, for example, rhythm and pace or tempo. Our approach is different in two main aspects. First, we use a semiotic model of the video genre to guide analysis and annotation. Second, while in computational media aesthetics design and production knowledge remains hidden in the algorithm for the analysis and annotation, in our proposal design knowledge is explicated in the annotation itself with the benefits we have already explained.

Semiotic annotation is inevitably complex if we try to capture the whole articulation and richness of intended meanings inscribed within a communicative artifact. Doing it well requires expertise. Automatic tools can be used to support low-level analysis of expressive qualities such as shot detection, dominant color identification, spectro-morphological analysis of sound objects, basic video statistics, etc. However, for the more abstract levels, the human intervention is still needed. We do not claim or expect that the average video producer will be able to follow the procedure without prior training. To what extent semiotic annotations can be replicated? If two or more analysts use the meta-model described in Section IV to annotate the same product are their results consistent? This is an important issue that requires further research.

Manual annotation is time-consuming but the case study we have worked out and our past experience with students showed that, for video commercials, it is a feasible approach

due to the limited time extension of these kinds of products. The effort, in this case, is largely rewarded by the benefits connected with the unfolding of new design knowledge as discussed beforehand. Moreover, it should be noted that completeness of analysis is not always necessary. In many cases, only those segments of a video artifact that are deemed interesting and relevant for the purpose of the analysis are annotated. For longer texts such as films and documentaries, the manual approach is surely unfeasible without appropriate supporting tools. This is a direction of possible future research work, together with the construction of a formal ontology based on the proposed meta-model and its integration with existing top-level ontologies such as the OIO design pattern and DOLCE.

Semiotic annotation can be very useful during product use in order to compare actual interpretations with the intended interpretation embodied in the artifact. Intended meanings (or experience) is the golden standard in order to evaluate the effectiveness of communication. Moreover, the means/ends structure of the annotation can be exploited to make a kind of "communication diagnosis", i.e., to link symptoms (e.g., discrepancies between a user's interpretation and the author's intended meaning) with possible causes (e.g., the structural segments of the video that could be responsible for the observed symptoms). This is another future direction of research.

A final remark regards the scope of applicability of semiotic annotation. Although semiotic theories can be fruitfully applied for the analysis of a wide range of genres of texts (and recently to physical artifacts as well) we consider persuasive discourses (such as video commercials, advertising images, learning objects and advergames) the most interesting fields of application. Quoting De Sousa: " ... semiotic methods are often looked upon with skepticism and rarely taken into consideration regardless their usefulness to address interpretative analysis in a rigorous and systematic way..." [95]. We hope that this situation could change in the future and more semiotic aware models and tools could be proposed for a more effective content analysis and annotation. This research aims at being a step toward this end.

ACKNOWLEDGMENT

The authors thanks the reviewers for their constructive and thoughtful critiques.

REFERENCES

[1] E. Toppano, "Semiotic annotation of video commercials: why the artifact is the way it is?," Proc. WEB 2017: The Fifth International Conference on Building and Exploring Web Based Environments, IARIA, Barcelona, Spain, pp. 45-51, 2017.

[2] J. L. Aaker, "Dimensions of brand personality," Journal of Marketing Research, Vol. 34, No. 3, pp.347-356, 1997.

[3] F. R. Esch, "Brand identity: the guiding star for successful brands," in Schmitt, B.H., Rogers, D.L. (Editors). Handbook on Brand and Experience Management. Edward Elgar Publishing Limited, UK, Cheltenham, Chapter 4, pp. 58-76, 2008.

[4] B. H. Schmitt and D. L. Rogers, Handbook on Brand and Experience Mangement, Edward E. Northampton, MA, USA, 2008.

[5] IAB, "A digital video advertising overview," Interactive Advertising Bureau, 2008.

[6] IAB, "Digital video in-stream ad format guidelines," Interactive Advertising Bureau, 2016.

[7] J. Bruner, "Life as narrative," Social Research, Vol.71, No. 3, pp. 691-710, 2004.

[8] M. C. Green and T. C. Brock, "The role of transportation in the persuasiveness of public narratives," Journal of Personality and Social Psychology, Vol. 79, No. 5, pp. 701-721, 2000.

[9] H. Bilandzic and R. Busselle, "Narrative persuasion," in The SAGE Handbook of Persuasion, J. P. Dillard and L. Shen (Eds.), Sage Publications, Chapter 13, pp. 200-219, 2013.

[10] R. Busselle and H. Bilandzic, "Measuring narrative engagement," Media Psychology, 12, pp. 321-347, 2009.

[11] T. M. Karjalainen, "It looks like a Toyota: Educational approaches to designing for visual brand recognition," International Journal of Design, Vol.1, No.1, pp. 67-81. 2007.

[12] E. Toppano and A. Toppano, "Ethos, in sound design for brand advertisement," in Proc. ICMC|SMC, Athens, Greece, pp. 1685-1692, 2014.

[13] K. Krippendorff, "On the essential contexts of artifacts or on the proposition that design is making sense (of things)," in The Idea of Design, V. Margolin and R. Buchanan (Eds.), The MIT Press, pp. 156-184, 1995.

[14] E. T. Kazmierczak, "Design as meaning making: from making things to the design of thinking," Design Issues, Vol. 19, No. 2, pp. 45-59, 2003.

[15] R. Buchanan, "Declaration by design: rhetoric, argument, and demostration," Design Issues, Vol. 2, No. 1, pp. 4-22, 1985.

[16] P. Desmet and P. Hekkert, "Framework of product experience," International Journal of Design, Vol.1, No. 1, pp. 57-66, 2007.

[17] N. Crilly, A. Maier, and P.J. Clarkson, "Representing artifacts as media: modelling the relationship between designer intent and consumer experience," Journal of Design, Vol. 2, No.3, pp.15–27, 2008.

[18] R. Troncy, B. Huet, and S. Schenk, Multimedia Semantics, Metadata, Analysis and Interaction, Wiley, 2011.

[19] F. M. Nack, "Capture and transfer of metadata during video production," CWI Report INS-E0513, 2005.

[20] E. Toppano and V. Roberto, "Semiotic design and analysis of hypermedia," in Proc. 20th ACM Conference on Hypertext and Hypermedia, HT2009, Turin, Italy, 367-368, 2009.

[21] E. Toppano and V. Roberto, "Semiotic-based conceptual modelling of hypermedia," Proc. Image Analysis and Processing, ICIAP2013, Napoli, Italy, Lecture Notes in Computer Science, Vol. 8157, pp. 663-672, 2013.

[22] M. Agosti and N. Ferro, "A formal model of annotations of digital content," ACM Trans. on Information Systems, Vol. 26, No. 1, Article 3, 2007.

[23] A. Hanbury, "A survey of methods for image annotation," Journal of Visual Languages and Computing, Vol. 19, Issue 5, pp. 617-627, 2008.

[24] P. Andrews, I. Zaihrayeu, and J. Pane, "A Classification of semantic annotation systems," Journal of Semantic Web, Vol.3, Issue 3, IOS Press, pp. 223-248, 2012.

[25] E. Oren, K. H. Moller, S. Scerri, S. Handschuh, and M. Sintek, "What are semantic annotations," Technical Report, DERI, Galway, 2006.

[26] W3C, "Open annotation data model," Open Annotation Comunity Group, 2013.

[27] W3C, "Semantic Web," https://www.w3.org/standards/semanticweb/ Accessed: 11 Dec. 2017.

[28] J. M. Martinez, “MPEG-7 Overview,” ISO/IEC JTC1/SC29/WG11 N6828, Palma de Mallorca, Spain, 2004.

[29] A. B. Benitez, J. M. Martinez, H. Rising, and P. Salembier, “Description of a single multimedia document,” in Introduction to MPEG 7: Multimedia Content Description Language,” B.S.Manjunath, P.Salembier, T.Sikora (Eds.), Wiley, pp.111-138, 2002.

[30] S. Dasiopoulou, V. Tzouvaras, I. Kompatsiaris, and M.G. Strintzis, "Enquiring MPEG-7 based multimedia ontologies," Journal of Multimedia Tools and Applications, Vol. 46, Issue 2-3, pp.331-370, 2010.

[31] R. Arnd, R. Troncy, S. Staab, and L. Hardman, "COMM: a core ontology for multimedia annotation," Handbook on Ontologies, Springer, Berlin, Heidelberg, pp. 403-421, 2009.

[32] M. Hausenblas, "Multimedia vocabularies on the Semantic Web," W3C Incubator Group Report, 2007.

[33] P. Schallauer, W. Bailer, R. Troncy, and F. Kaiser, "Multimedia Metadata Standards," in R. Troncy, B. Huet, and S. Schenk (Eds.), Multimedia Semantics, Metadata, Analysis and Interaction, Wiley, pp.129-144, 2011.

[34] L. Hardman, Z. Obrenovic, F. Nack., B. Kerherve B., and K. Piersol, "Canonical Processes of semantically annotated media production," Multimedia Systems, 14(6), 327-340, 2008.

[35] M.L. Ryan, "Narrativity and its modes as culture-transcending analytical categories", Japan Forum, 21(3), BAJS, pp. 307–323, 2009.

[36] S. Kinnebrock and H. Bilandzic, "How to make a story work: introducing the concept of narrativity into narrative persuasion," International Communication Association Conference, ICA-2006, Dresden, 2006. Retrieved from: http://publications.rwth-aachen.de/record/50885/files/3638.pdf Accessed: 11 Dec. 2017.

[37] F. de Saussure, Cours de linguistique generale, Payot, Paris, 1922.

[38] C. S. Peirce, Collected Writings, C. Hartshorne, P. Weiss, and A. W. Burks (Eds.), Harvard University Press, Cambridge, MA, 1931.

[39] T. van Leeuwen, Introducing Social Semiotics, London: Routledge, 2005.

[40] R. Eugeni, "Media experiences and practices of analysis. For a critical pragmatics of media," International Workshop “Practicing Theory”, University of Amsterdam, March 2-4, 2011.

[41] C. S. de Souza, The Semiotic Engineering of Human-Computer Interaction, the MIT Press, Cambridge, MA, USA, 2004.

[42] S. Vihma, "Design semiotics- Institutional experiences and an initiative for a semiotic theory of form," in Design Research Now: Essays and Selected Projects, R. Michel (Ed.), pp. 219-232, 2007.

[43] C. Scolari, "Transmedia storytelling: implicit consumers, narrative worlds, and branding in contemporary media production", International Journal of Communication, Vol.3. pp. 586–606, 2009.

[44] T. Gruber, "A translation approach to portable ontology specification," Knowledge Acquisition, Vol. 5, No. 2, pp. 199-220, 1993.

[45] L. Manovich, The language of new media, Cambridge, MA: the MIT Press, 2001.

[46] S. Grimaldi, S. Fokkinga, and I. Ocnarescu, "Narratives in design: a study of the types, applications, and functions of narratives in design practice," in Proc. DPPI 2013, ACM NY, USA, pp. 201-210, 2013.

[47] M. Davis, " Knowledge representation for video," in, Proc. of Twelfth National Conference on Artificial Intelligence (AAAI-94), Seattle, Washington, AAAI Press, pp. 120-127, 1994.

[48] C. G. M. Snoek and M. Worring, ”Multimodal video indexing: a review of the state-of-the-art,” Multimedia Tools and Applications, Vol. 25, pp. 5-35, 2005.

[49] S. Pfeiffer, R. Lienhart, and W. Effelsberg, "Scene Determination based on video and audio features," Multimedia Tools and Applications, 15, pp.59-81, 2001.

[50] P. von Stackelberg, "Creating transmedia narratives: the structure and design of stories told across multiple media," Master's Thesis, State University of New York, 2011.

[51] V. Lombardo and R. Damiano, "Narrative annotation and editing of video," in Proc. ICIDS'10, Third Joint Conference on Interactive Digital Storytelling, pp.62-73, 2010.

[52] A.J. Greimas and P. Courtès, Semiotics and language. An analytical dictionary. Indiana University Press, Bloomington, IN, 1982.

[53] C. Bianchi, "Semiotic approaches to advertising texts and strategies: narrative, passion, marketing", Semiotica 183, 1/4, pp. 243-271, 2011.

[54] C. Colombo, A. Del Bimbo, and P. Pala, "Retrieval of commercials by semantic content: the semiotic perspective," Multimedia Tools and Applications13, pp. 93-118, 2001.

[55] J. M. Floch, Semiotics, Marketing, and Communication: Beneath the Signs, the Strategies, Palgrave, Hampshire, 2001.

[56] J. A. Bateman, "Towards a grande paradigmatique of film: Christian Metz reloaded," Semiotica, 167-1/4, pp. 13-64, 2007.

[57] C. Tseng, "Analysing characters' interactions in filmic texts: a functional semiotic approach," Social Semiotics, 23:5, pp. 587-605, 2013.

[58] A. J. Greimas, F. Collins, and P. Perron, "Figurative Semiotics and the Semiotics of the Plastic Arts," New Literary History, Vol.20, No. 3, pp. 627-649, 1989.

[59] D. Smalley, "Spectromorphology: explaining sound-shapes", Organized Sound, Vol.2, No.2, Cambridge University Press, pp.107-126, 1997.

[60] G. Kress and T. van Leeuwen, Reading images. The grammar of visual design. Routledge, New York, 2003.

[61] L. Hébert, Tools for text and Image analysis: An introduction to applied semiotics, Département de lettres Université du Québec à Rimouski, Version 3, 2011.

[62] J. Campbell, The hero with a thousand faces, New World Library, 2008.

[63] B. Rolfe, C. M. Jones, and H. Wallace, "Designing dramatic play: story and game structure," Proc. HCI2010, Int. Conf. on Human-Computer Interaction, Dundee: British Computer Society, pp. 448-452, 2010.

[64] E. Benveniste, Problemes de linguistique generale, Gallimard, Paris, 1966.

[65] C. Harrison, "Visual social semiotics: understanding how still images make meaning," Technical Communication, Vol. 50, no. 1, pp. 46-60, 2003.

[66] J. L. Lemke, "Travels in hypermodality," Visual Communication, Vol. 3 (1), pp. 299-325, 2002.

[67] J. F. Allen, "Towards a general theory of action and time," Artificial Intelligence, 23, pp. 123-154, 1984.

[68] A. Galton, "Towards an integrated logic of space, time and motion," in Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI-93), Chambery, France, pp. 1550-1555, 1993.

[69] W. C. Mann and S. A. Thompson, "Rhetorical Structure Theory: toward a functional theory of text organization," *Text*, Vol. 8, No. 3, pp. 243-281,1988.

[70] A. Gaeta, M. Gaeta, G. Guarino, and S. Miranda, "A smart methodology to improve the story building process," Journal of e-learning and knowledge society," Vol. 11, No. 1, pp. 97-124, 2015.

[71] N.J. Lowe, "A cognitive model," in Bernstein, M., and Gerco D. (Eds.) Reading Hypertext, Eastgate Systems, pp. 35-57, 2009.

[72] K. Rohlfing, D. Loehr, S. Duncan, A. Brown, A. Franklin, I. Kimbara, J. T. Milde, F. Parrill, T. Rose, T. Schmidt, H. Sloetjes, A. Thies, and S. Wellinghoff, "Comparison of multimodal annotation tools," Workshop report, Gesprächforschung - Online-Zeitschrift zur Verbalen Interaktion, 7, pp. 99-123, 2006.

[73] B. Hellwing and D. Uytvanck, "EUDICO linguistic annotator (ELAN)," Version 2.0.2, software manual, 2004.

[74] P. Boersmimedia and D. Weenink, "Praat: doing phonetics by computer," available from: http://www.fon.hum.uva.nl/praat/ Accessed: 11 Dec. 2017.

[75] A. Chebotko, Y. Deng, S. Lu, F. Fotouhi, and A. Aristar, "An Ontology-Based Multimedia Annotator for the Semantic Web of Language Engineering," in A. Sheth, M. Lytras (Eds.) Semantic Web-Based Information Systems. State-of-the-Art Applications, Cybertech Publishing, pp. 140-160, 2007.

[76] Pepsi Cola clip [online]. Available from: http://www.youtube.com/watch?v=edfRG9lREHc Accessed: 11 Dec. 2017.

[77] W. Gaver, "What should we expect from research through design?," Proc. CHI'12, ACM Press, pp. 937-946, 2012.

[78] J. Bardzell, S. Bardzell, and L. K. Hansen, "Immodest Proposals: research through design and knowledge," Proc. CHI'15, ACM Press, pp. 2093-2102, 2015.

[79] P. Brey, "Philosophy of Technology after the empirical turn," Techné: Research in Philosophy and Technology, 14(1), pp. 36-48, 2010.

[80] O. W. Bertelsen and S. Pold, "Criticism as an approach to interface aesthetics," in Proc. NordiCHI'04, ACM Press, New York, pp. 23-32, 2004.

[81] J. Bardzell and S. Bardzell, "Interaction criticism: a proposal and framework for a new discipline of HCI," Proc. CHI-08, ACM Press, pp. 2463-2472, 2008.

[82] J. Bardzell, "Interaction criticism: an introduction to the practice," Interacting with Computers, 23, pp. 604-621, 2011.

[83] J. Bardzell, "Interaction criticism and aesthetics," Proc. CHI'09, ACM Press, pp. 2357-2366, 2009.

[84] J. Lowgren, "Annotated portfolios and other forms of intermediate-level knowledge," Interactions, pp. 30-34, 2013.

[85] B. Gaver and J. Bower, "Annotated portfolios," Interactions, pp. 40-49, 2012.

[86] P. P. Verbeek, "Materializing morality: design ethics and technological mediation," Science, Technology & Human Values, Vol. 31, No. 3, pp. 361- 380, 2006.

[87] L. Anticoli and E. Toppano, "Technological mediation of ontologies: the need for tools to help designers in materializing ethics," International Journal of Philosophy Study, Vol.1, Issue 3, pp. 23-31, 2013.

[88] Y. Van Den Eede, "In between us: on the transparency and opacity of technological mediation," Found. Sci. 16, pp.139-159, 2011.

[89] D. Berdichevski and E. Neuenschwander, "Toward an ethics of persuasive technology," Communication of the ACM, Vol. 45, No. 5, pp. 51-58, 1999.

[90] L. Hallnas and J. Redstrom, "From use to presence: on the expressions and aesthetic of everyday computational things," ACM Trans. on Human-Computer Interaction, Vol.9, No 2, pp. 106-124, 2002.

[91] P. Brey, "Theorizing the cultural quality of new media," Techné: Research in Philosophy and Technology, 11(1), pp. 2-18, 2007.

[92] A. C. Danto, The transfiguration of the commonplace. A philosophy of art. Harvard University Press, CA: Massachusetts, 1990.

[93] J. Bowers, "The logic of annotated portfolios: communicating the value of Research Through Design," Proc. DIS 2012, ACM Press, pp. 68-77, 2012.

[94] C. Dorai and S. Venkatesh, "Bridging the semantic gap with computational media aesthetics," IEEE MultiMedia, Vol. 10, Issue 2, pp. 15-17, 2003.

[95] C. S. de Souza, "Viewpoint: semiotic perspectives on interactive languages for life on the screen," Journal of Visual Languages & Computing, Vol. 24, Issue 3, pp. 218-221, 2013.