Semiotic Annotation of Video Commercials: Why the artifact is the way it is?

Elio Toppano
Dipartimento di Scienze Matematiche, Informatiche e Fisiche (DIMA)
University of Udine
Via delle Scienze, 208 – Udine (Italy)

***Abstract*—Traditionally, semantic annotation of audiovisual texts is used to describe expressive features and content of a product for a more efficient and effective browsing, retrieval, filtering or reuse of the resource. Drawing on semiotic theories, this paper proposes a new concept of annotation – called semiotic annotation – whose goal is to describe the multilayered structure of meanings inscribed within the audiovisual by its author/designer. The advantages of this kind of annotation is discussed with respect to means/ends analysis of video commercials. A case study is then illustrated that exploits a semiotic compliant informal ontology proposed in a previous work to assess the effectiveness of the conceptualization and the annotation method.**

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

I. INTRODUCTION

According to a recent report published by Cisco [1] video is currently representing and will represent in the future - together with gaming - one of most important parts of the global Internet traffic growth. We may expect that video advertising will follow this trend and thus will be a fast-growing opportunity on line and one of the most promising ad formats in the future. Nowadays, web video advertising covers a wide range of products differing in production quality, time length and distribution. The universe of content is broad and varied ranging from professionally produced content, generally, repurposed from Broadcast Video and Cable Networks to clips created and uploaded by everyday people, i.e., user generated content.

Distribution and formats also vary ranging from linear and non-linear in-stream video to in-display video and combinations thereof packaged together in a compelling way [2]. Most ad videos are narratives, that is, they told a micro-story aimed at presenting a product-service or communicating brand identity. This paper is about video annotation [3]. We address this problem following a communication-based design approach [4] according to which video is seen as a *mediator* between the intentions of the designer (i.e., author) and the interpretation of the user.

Intentions (e.g., brand's identity communication, product advertising) are assumed to be inscribed within the artifact through *semantic transformation* [5] and implicitly communicated to the user by the video expression and content. In the following, we will take the perspective of the author/designer of the video rather than the final user. We are interested in how meaning is intentionally constructed and articulated during the design process, how it shapes the audiovisual and how users can infer the designer's intentions – both informative and persuasive - and recognize *that they are users*, i.e., that their experiences with the product have been anticipated. In line with this objective, annotation is conceived of as an activity aimed at describing the *experiential project* envisaged by the author and embodied within the product [6]. It is performed by the author/designer during the development of the audiovisual artifact and requires a set of annotation descriptors (i.e., concepts and relative terminological realizations) that could be used not only to describe how the artifact is made and functions but also why it is the way it is. The paper is organized as follows. The next section elaborates more on the motivations lying behind our work. Section III introduces the concept of *Semiotic Annotation* which is at the core of our approach. Section IV summarizes some basic requirements the design of an ontology supporting the approach should satisfy and suggests a possible solution. Section V describes a method for video annotation and exemplifies it in a specific case. Section VI discusses benefits and limitations of the approach. Finally, Section VII draws some conclusions.

II. MOTIVATIONS

The term *annotation* can be understood in two different ways: i) as an activity (i.e., the process by means of which metadata are attached to other data) and ii) as the result of the activity. [7] proposed a formalization of annotation in terms of a quadruple: the annotated data (i.e., the subject of annotation), the annotating data (i.e., the object of annotation), the annotation relation (i.e., the predicate that defines the type of relationship between annotated and annotating data) and the context in which the annotation is made. Traditionally - see for example Mpeg7 [8] [9] - metadata are used to describe the expressive characteristics (e.g., visual and audio features) or the semantic content of an entire multimedia product or of specific product fragments.

Contextual information, if present, refers to the people involved in the development of the document (e.g., the scriptwriter, the video-maker, the sound designer), the place and time of its production, its spatial and temporal scope, the target user. Seldom if ever, contextual metadata refer to the design process itself, such as, for example the designer's intentions behind the product, the effects that the designer intends to evoke in the user, the rationale behind specific

 ISBN: 978-1-61208-557-9

functional and expressive design choices. As a consequence, traditional annotation approaches do not allow means/ends analyses [10]. For example, they do not support neither *teleological explanations* (i.e., Why the artifact is the way it is?) nor *causal explanations* (i.e., How a specific communicative goal/intention or impression has been achieved in terms of specific visual and aural choices and compositions). The rationale is that the aim of current annotation approaches is more focused on product filtering, retrieval and reuse of documents than on critical analysis, explanation and evaluation. Following recent developments in the field of Interaction Criticism [11], and Design for Experience [12] [13] we claim that means/ends metadata could add value to the product and could be useful both for the designers and the users. More specifically, designers could exploit this knowledge:

- for highlighting the concerns and design choices made during multimedia development;
- for the analysis and comparison of multimedia products 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;
- for the synthesis of new products because means/ends metadata implicitly codify design knowledge that can be fruitfully extracted and reused in new projects;
- for the evaluation of the internal coherence of products because means/ends metadata explicate the relationships existing between design choices taken at different aggregation and abstraction levels;
- for the "diagnosis and repair" of communication because means/ends metadata allow the identification of symptoms (i.e., discrepancies between the intended meaning of the product and the actual one) and the localization of causes.

These activities are particularly important in some application domains such as transmedia projects, web marketing, brand driving and management where issues related to the differentiation of advertisement products (e.g., web sites, advergames, video clip, etc.), internal product coherence, effective communication of brand identity, time consistency of portfolio products are paramount. The annotation of multimedia product with means/ends metadata could be also useful to the user:

- to make more informed choices, i.e., to better understand if a product is adequate with respect to her values, needs, desires, preferences;
- to better exploit the concept of *genre* in document retrieval. This is because means/ends annotation allows to anchor the genre classification to several internal properties of the product (e.g., content, discourse structure, expression qualities) and their relationships;
- to reconstruct the designer's intentions inscribed within the multimedia product (this is the well-known design stance by Dennet [14]). This is helpful in order to understand a product's *technological mediation*, i.e., the way the product may affect the experience and the actions of the users [15]. This calls for a more responsible ethical attitude by the side of the designers and for a better awareness of the persuasive role of technologies by the side of users;
- to evaluate the authenticity of a product's brand by comparing the brand identity (i.e., the constellation of meanings-values the brand says to adhere to) with the actual meanings embodied and communicated by its marketing portfolio (e.g., video commercial).

III. A SEMIOTIC APPROACH

We address annotation by drawing on results obtained within the fields of semiotics and narratology [16] [17]. As stated by Scolari [17] Semiotics studies objects (texts, discourses) to understand processes (sense production and interpretation). It focuses on the meanings *inscribed* within a product and the *potential experience* that these meanings may trigger or evoke in the final users. It is both empirical and critical. It is based on the analysis of concrete products from a phenomenological perspective and is aimed at re-constructing the experiential project - a reading proposal or contract - that has been implicitly inscribed into the product by the designer-author starting from the product expression (i.e., its sensorial qualities) and explaining how semiotic materials (i.e., written text, images, music) and their combination may support such a project. From Semiotics we borrow the methodology of interpretive multimodal text analysis. Among the various semiotic research traditions that succeeded and stratified in time we are interested in those approaches that consider the meaning as the result of an interpretive process that can be articulated on different conceptual planes or layers. This is because we need a set of conceptualizations that could be used to build the means/ends ladder that we are looking for. Therefore we have taken as a reference the Generative Semiotic of Text by Greimas [18] [19] which we have integrated with some contributions coming from Socio Semiotics [20] and Enunciation Theory [18]. We propose to use the term "*Semiotic Annotation*" instead of the more commonplace term "*Semantic Annotation*" to emphasize this kind of approach. Looking at a video - and more generally at a multimedia product - from a semiotic point of view requires a new perspective on annotation. We consider the video presentation - i.e., the real time succession of multimodal events occurring during the interpretation-execution of digital data and instruction by a HW-SW platform - as a structured whole of *signs* (or semiotic resources) belonging to several *representational modalities* (e.g., written and spoken language, image, music, sound, audiovisual). These signs play the role of annotated data and the meanings referring to configuration of sensorial qualities of signs (expression), and arrangement of semantic entities (narrative content) as annotating data. In other words we annotate the flow of events as occurring during the execution of the presentation with the experiential (e.g., sensorial, narrative and relational-emotional) project envisaged by the designer and inscribed by her into the product [6]. We stress the fact that these meanings are always context sensitive and depend

 ISBN: 978-1-61208-557-9

on the socio-cultural environment of the interpreter. Moreover, the intended meanings could be different from those attributed by the actual user if she does not correspond to the implied user (called the addressee). Only when the actual user projects herself into the implied one (i.e., there is a cooperation between sender and receiver) it can be said that the communication is truly effective.

IV. REQUIREMENTS FOR SEMIOTIC ANNOTATION

In this section, we present a list of requirements for the design of a semiotic compliant narrative video annotation. 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 spatio-temporal *aggregation levels* and *conceptual abstractions*. As an instance, it should be possible to look at the video as a spatial configuration of regions within single key frames, or as a temporal sequence of individual shots, scenes, sequences, episodes, etc. Moreover, it should be possible to look at the video in terms of low level features; patterns of features (e.g., visual figures) that support a semantic construct such as an object, event or symbolic association; configurations of objects, subjects and events representing more abstract constructs such as situations, entire discourses and stories;
- to relate together the annotations pertaining to the *same*/*different* aggregation levels or conceptual abstractions by several types of relationships (e.g., spatial, temporal, logical, rhetorical, typological, mereological, causal/teleological relationships);
- to have multiple alternative annotations describing the presentation from different points of view (e.g., multiple coexisting alternative discourses/stories).

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

- to describe basic kinetic and plastic features of visual segments (e.g., shapes, colours, positions, textures, sizes, cinematic movements, visual contrasts, rhythms, etc.) as well as spectro-morphological features of audio segments (e.g., time features such as amplitude, envelope, etc. and spectral ones such as pitch, timbre, harmonicity);
- to represent meta-attributes such as for example aesthetic impressions (e.g., visual balance, order, symmetry) and product character [21];
- to describe representational meanings such as figurative formants in conceptual and narrative images [20]. More specifically to describe narrative structures by specifying all fundamental entities constituting a storyworld such as participants, actions, goals, settings [22]. To represent stories at different abstraction levels and using different dramaturgical schemas (e.g., the canonical scheme of Greimas [18], or the Hero's Journey by Campbell [23])

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

- to describe the interpersonal meaning associated to the presentation. As an instance, it should be possible to annotate the presentation with data regarding the inscribed addresser and addressee (i.e., the simulacra of the empirical sender/ receiver); their relationship, their attitude with reference to the content of the presentation, the intended effect the author wants to evoke in the actual receiver such as affective responses (emotions, mood, feelings);
- to describe the subjects of discourse (i.e., the subjects that are responsible for *how* a story is narrated and expressed by a text) and their relationships both with the characters of the story and the simulacra of the sender/receiver;
- to represent the deep values intended by the author (e.g., brand values) and the way they are inscribed within the video product.

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 storyline, the elements of the story with discourse segments and expressive qualities; expressive qualities with impressions and interpersonal meanings and so forth. We have recently proposed an informal conceptualization - not yet an ontology - that provides a core set of basic descriptors that can be used to perform a semiotic annotation according to the above requirements [24].

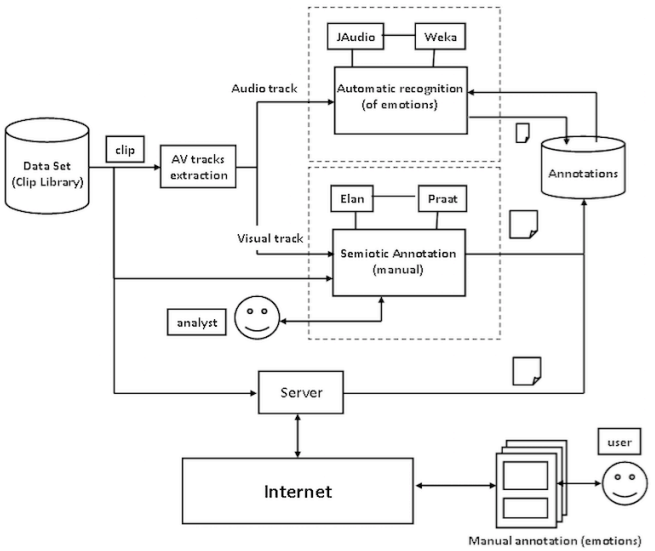

Figure 1. A schematic view of the infrastructure developed for video annotation

Figure 1 shows a conceptual model of an infrastructure we are developing to provide various kinds of automatic and manual annotation services. Currently, semiotic annotation is

performed manually by expert annotators using the EUDICO Linguistic Annotator (*ELAN*) environment [25].

We are using this tool to make a series of annotation experiments whose aim is to assess the effectiveness of the conceptualization proposed in [24] and to create a preliminary data base of annotated resources that could be used in the future as a *ground truth* for automatic annotation algorithms. Automatic annotation is currently limited to the emotions expressed by audio tracks. It is performed by using jAudio and specific supervised learning techniques provided by Weka. Non-expert annotation of emotions in videos can be performed manually by users using mobile devices. Users' annotations provide data for the learning stage of the automatic annotation subsystem. In the next section we will illustrate in detail an example of expert annotation using *ELAN*.

V. A CASE STUDY

ELAN is a multimodal annotation tool that is widely used within the domain of Multimodal Discourse Analysis [26]. It enables the annotator to define a reference vocabulary and use it to describe an audiovisual product at different aggregation and abstraction levels (called annotation tiers). An annotation tier can be either alignable or referring. *Alignable tiers* are directly linked to the time axis of an audiovisual clip and can be divided into segments; *referring tiers* contain annotations that are linked to annotations on another tier which is called a parent tier and can be alignable or referring. Thus, tiers form a hierarchy where the root must be an alignable tier. The tool saves annotations in the XML format based on *ELAN* XML Schema. For the analysis of the audio component in the temporal and spectral domains the *Praat* software has been used [27]. *Pratt* results can be imported within *ELAN* and easily integrated with visual annotations.

In order to provide an example of semiotic annotation we will focus here on an advertisement clip by Pepsi-Cola [28]. The clip produced in the late 1980's is based on the bodycopy "Pepsi Cola. The choice of a new generation". 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 of the van. The boy brings a bottle of Pepsi near the microphone, uncaps it, pours liquid into a glass and drinks emitting an "Ahhhh" of pleasure. People attracted by puffing of gas and the boy's expression rush to the van to quench their thirst.

A systematic procedure has been envisaged for the analysis and annotation of video commercials (and more generally of audiovisual products). The procedure consists of the following basic stages:

- Stage-0 (*Whole clip annotation*). The whole clip is represented by an alignable annotation tier linked to a single segment (ClipSegment). This tier represents the root of the hierarchical multi-tiers annotation. In the case of the Pepsi Cola clip the ClipSegment last 29.5s at the frame rate of 30fps. The tier is annotated with the multimedia genre and the intention/goal of the product. In this way the genre and goal are directly linked with other annotations.
- Stage-1 (*Textual decomposition*). The root segment (ClipSegment) is represented by several textual structures (T-Structure). Some structures are associated to the visual representation modality, others to the aural modality. In the considered example, a structure (T-Structure1) is used to decompose ClipSegment into a sequence of T-Segments representing individual shots. By the term shot we intend a series of visual frames produced by the camera in an uninterrupted recording operation. A further textual structure (T-Structure2) is used to annotate special transition edits and effects like fades, dissolves, overlayed text, etc. Finally, another structure (T-Structure3) is used to decompose the ClipSegment on the base of continuous sequences (T-Segments) of homogeneous sound objects. In the Pepsi example these sequences include silence, speech, environmental sound and effects. In more complex examples it could be necessary to devote a separate textual structure to each constituent of a complex audio sandwich e.g., music, effects, speeches, environmental sounds as well as to sound transitions. It should be stressed that visual and aural structures are not necessarily aligned in time. Speech and music for example can continue while the camera switches from one shot to the next one.
- Stage-2 (*Textual annotation*). In 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 such as colour, shape, texture, timbre, pitch, movement, tempo, etc. Further tiers can be used to annotate intended hedonic impressions (e.g., emotions, mood), and meta-attributes (e.g., product character).
- Stage-3 (*Discourse decomposition*). The root ClipSegment is represented by one or more discourse structures (D-Structure). The decomposition is 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 corresponds to changes in settings from outside to inside the Pepsi Cola van and vice-versa.
- Stage-4 (*Narrative structure decomposition*). Each scene (D-Segment) is annotated by a narrative structure composed by narrative programs [18] [19] and their logical and temporal relationships.
- Stage-5 (*Narrative program annotation*). A set of referring annotation tiers are introduced and associated to previous narrative structures to annotate single narrative programs. For each narrative program a set of tiers is used to separately

 ISBN: 978-1-61208-557-9

describe the main components of the program namely the actor (subject of doing), the action, the effect (subject of state, transition) and the object of value. In the example under consideration, D-Segment7 and D-Segment9 (a scene inside 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 user aware of brand (D-effect). 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 object of value, i.e., the product/brand (D-Effect).

- Stage-6 (*Relational analysis and annotation*). The root segment (ClipSegment) is analyzed in order to identify the markers of addresser and addressee. As an example, in the Pepsi Cola clip, the bottle of Pepsi including the logo and trademark represents the addresser (i.e., the brand Pepsi). The people approaching the van to buy the product is a representation (a surrogate) of the addressee. A set of further tiers have been introduced and linked to the ClipSegment to represent interpersonal metafunctions [20] expressed by visual and aural features. According to social semiotics, a character's gaze, size of shot, vertical and horizontal camera angle, are related to engagement, social distance, power and involvement relationships respectively. In the same way, tone of voice in speech, sound perspective, volume, can be used to evoke various degrees of intimacy or distance between the characters of the story (and indirectly the brand) and the user. The clip aims at establishing both empathy and trust between users/consumers and actors. Empathy can occur between the boy and the user, which is urged to share with crowds the sensation of freshness. The user is also invited to trust that the experience ensured by brand Pepsi – the addresser – is authentic; that drink (and indirectly the brand) is indeed an object of value in that context, so worthwhile purchasing.

Several temporal relationships among annotations belonging to different tiers are implicitly described through the relations existing between their corresponding tiers. For example, all referring tiers associated to the same alignable tier inherit its time decomposition. As a consequence their annotations are automatically time aligned. Figure 2 shows a screen shot of ELAN illustrating a subset of the tiers used to annotate the Pepsi Cola clip.

VI. DISCUSSION

It is important to recognize that semiotic annotation is model based: it exploits a meta-model (i.e., an informal ontology) of the narrative video genre. The meta-model [24] makes explicit different assumptions, conceptualizations and theories shared within the semiotic field. One assumption is that the realization of a commercial video amounts to the construction of meaning and that the meaning rests on the relationships existing between the text, discourse and story layers rather than on the single elements of the video.

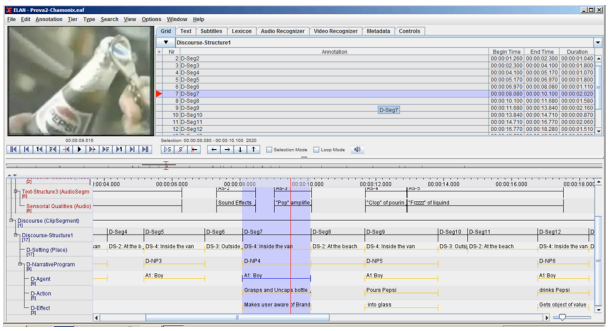

Figure 2. A screenshot of *ELAN* representing a subset of annotation tiers associated to the Pepsi Cola clip.

These relationships provide the video with the kind of unity, internal coherence and sense it shows. As a consequence, as discussed in the Section IV, the meta-model is multi-layered and relational. In this context, the annotation represents a kind of *intermediate level knowledge* [29], connecting the abstract concepts of the ontology, that are used, as descriptors, in the annotation, to the specific values these concepts take in the concrete video under consideration. Under this perspective, the annotation unfolds the design knowledge that is embodied in the artifact. It shows how the meta-model has been instantiated by the author of the video in the concrete artifact.

The availability of design knowledge provides several benefits for the designers and the users as well. It allows to answer several questions about the product. These questions - that are the *competence questions* associated to the ontology - refer, for example, to the way a narrative is decomposed into narrative programs; how a specific narrative program has been translated at the textual level (in terms of visual and auditory qualities); who is involved in the story (as well as in discourse) and which functional role he/she is playing (e.g., subject of action, subject of value, simulacrum of the sender/receiver, narrator, and so forth); how social distance and involvement are evoked through the visual and the auditory features (e.g., by selection of shot size, direction of the eye gaze, horizontal and vertical position of video camera; or by selection of sonic perspective, panorama, pitch distribution, etc.). The annotation can be used for search and retrieval (e.g., search all videos or parts of videos where a specific narrative program is represented or where a specific social distance is employed) but we think that design knowledge inscribed within the annotation is fundamentally useful for explaining the way a specific video functions 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 the design knowledge embodied within the video in order to reuse it,

 ISBN: 978-1-61208-557-9

evaluate its internal coherence or take inspiration from it in developing new products. They can exploit the annotation to compare two or more video of the same author or brand in order to search for redundancies and variations. They can 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 portfolios or exhibitions. This may benefit several target users such as, for example, artists, architects, industrial designers. For generic users, the annotation presents a more fundamental benefit. It is known that every artifact - and technology, in general - plays a *mediation* role: it changes the way users perceive and experience the world as well as the way they act in the world [30]. The mediation effect is usually made transparent, in the sense that is not visible. However, recent studies in the field of Philosophy of Technology claim that such an effect should be made opaque and comprehensible to users [31]. In the case of advertisement products - such as the video commercials - this means to make explicit the rhetorical mechanisms that are at the base of their persuasive and informative functioning. 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 sort of culture it will encourage or resist. Moreover, the disclosure of motivations, methods, and intended outcomes is one of the ethical principles in persuasive design as discussed in [32]. Semiotic annotation may support this principle and contribute to the diffusion of a critical attitude toward multimedia and a greater awareness of the social effects this kind of products may produce.

In the specific case study under consideration, semiotic annotation is useful to explain how the Pepsi Cola video functions. The video tries to persuade to consume the Pepsi Cola by means of a narrative telling us "the process of persuasion of buying/consuming a Pepsi Cola". This process includes the following steps: 1) insert yourself in a familiar situation (the delivery van of the Pepsi Cola reaches the crowded beach); 2) draw attention and represent a positive and euphoric experience of consumption (loudspeakers attract people; the experience of the boy drinking the Pepsi is communicated both visually and auditory); 3) activate in the consumer a desire to have a similar experience through the planning and realization of a purchase behavior (people rush to the van to buy the Pepsi). The persuasive goal is realized through three types of relationships: between the user and the product; between the user and the subject using the product, and, finally, between the user and the people-crowd on the beach that desire the product and activate themself to buy it. It is sufficient to view and hear the protagonist (the boy) uncapping the bottle, drinking the liquid and emitting the "Ahhhh" of pleasure to activate, in the user and in the crowd, a similar experience on the base of a common competence of what does it means to drink a cold beverage. Semiotic annotation allows the annotator to associate the shots of the video to the various phases of the persuasive process; to describe each shots by representing its associated narrative programs and visual and auditory qualities; to describe the technical and discursive mechanisms that are used to address the user and to engage him/her; to evaluate the degree of verisimilitude associated to the video by analyzing the kinds of sound objects that are used and the use of subjective or objective shots; etc.

Semiotic annotation is different from pure keyword or concept annotation. The task is not simply to attach subjective comments, notes, interpretations or remarks to audiovisual segments but to unfold the generative process of sense making inscribed within the product. As a consequence the annotation should be performed by the video author since he/she is in a privileged position to provide valuable knowledge about design decisions. Alternatively, it could be made by other subjects such as critics or commentators, preferably with the help of the author. Anyway, the annotation should be considered as a constitutive part of the video. Through its indexical nature it points to features of the video and connects them to general concepts and issues making them topical for further discussion. It adds value to the product since it supports interpretation, clarification and comprehension.

One critical question regards the complexity of the task. Semiotic annotation requires deep knowledge about semiotic theories and well developed analytical skills. Part of this expert knowledge is embodied in the meta-model [24] that provides the relevant conceptualizations and vocabularies for the description. This is a benefit with respect to more general (i.e., not model-based) approaches. Automatic tools can be used to support low level analysis of expressive qualities such as shot detection, dominant colour identification, spectro-morphological analysis of sound objects, basic video statistics, etc. However, for the more abstract levels, the human intervention is still needed.

Manual annotation is time-consuming but our experience showed that, for video commercials, it is a feasible approach due to the limited time extension of these kinds of texts. The effort, in this case, is largely rewarded by the benefits connected with the unfolding of new design knowledge as discussed beforehand. 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. 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. The rationale is that these kinds of texts are *intentionally* developed to affect the experience and behavior of the intended users so they are usually carefully designed to achieve these persuasive goals. Therefore, it is particularly interesting to unfold the design thinking embodied in such types of products.

VII. CONCLUSIONS

This paper focuses on annotation of video commercials viewed as mediators between the intention of the designer/author and the interpretation of the user.

The main contributions can be summarized as follows:

 ISBN: 978-1-61208-557-9

- a new concept of annotation called semiotic annotation has been proposed to support means/ends description of video texts. The concept emphasizes the multilayered and interrelated nature of meanings embedded within the product. More specifically, a (narrative) video is conceived of as a structured system composed by three interrelated layers: story (*what* is depicted in the product), discourse (*how* it is told) and text (how the discourse is manifested through multimodal resources) [33];
- a systematic method has been outlined that can be used to manually annotate commercial video using *ELAN* as the annotation tool and the informal ontology proposed in [24] as the source of descriptors. The empirical annotation work done with the *ELAN* tool has showed the effectiveness of the proposed conceptual framework. A formalization of the conceptualization using *Protégé* is under development. The aim is to build an OWL-2 ontology that can be linked to DOLCE and could support a rich set of competency questions not currently supported by the simple query engine by *ELAN*.

Semiotic methods of analysis and descriptions are currently under utilized in the field of multimedia semantic annotation [10]. This paper strives for being a preliminary step toward a more "semiotic aware" attitude in this field.

REFERENCES

[1] CISCO, "Cisco Visual Networking Index: Forecast and Methodology, 2015-2020," White Paper, Cisco, 2016.

[2] IAB, "Digital Video In-Stream Ad Format Guidelines," Interactive Advertising Bureau, 2016.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] E. Oren, K. H. Moller, S. Scerri, S. Handschuh, and M. Sintek, "What are Semantic Annotations," Technical Report, DERI, Galway, 2006.

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

[9] 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.

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

[11] J. Bardzell, and S. Bardzell, "Interaction criticism: a proposal and framework for a new discipline of HCI," Proc. CHI2008, ACM, New York, USA, pp. 2463-2472, 2008.

[12] P. Wright, J. Wallace, and J. McCarthy, "Aesthetic and experience-centered design," ACM Transactions on Computer-Human Interaction, Vol.15, No.4, Article 18, 2008.

[13] G. Cockton, "When and Why feelings and impressions matter in Interaction Design," Invited keynote address, Proc. Kansei 2009, Warsawa, Poland, 2009.

[14] N. Crilly, "The design stance in user-system interaction," Design Issues, Vol. 27, No. 4, pp. 16-29, 2011.

[15] 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.

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

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

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

[19] 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.

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

[21] L. Janlert, and E. Stolterman, "The character of things," Design Studies, Vol. 18, No. 3, pp. 297-314, 1997.

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

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

[24] 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.

[25] B. Hellwing, and D. Uytvanck, "EUDICO Linguistic Annotator (ELAN)," Version 2.0.2, software manual, 2004.

[26] K. L. O'Halloran, Multimodal discourse analysis. In K. Hyland and B. Paltridge (eds.) Companion to Discourse, London and New York: Continuum, 2011.

[27] P. Boersmimedia a, and D. Weenink, "Praat: doing phonetics by computer," available from: http://www.fon.hum.uva.nl/

[28] Pepsi Cola clip [online]. Available from: http://www.youtube.com/watch?v=edfRG9lREHc Accessed: 2 february 2017.

[29] J. Lowgren, "Annotated Portfolios and other form of Intermediate-Level Knowledge," Interactions, pp. 30-34, 2013.

[30] P. P. Verbeek, "Materializing Morality: Design Ethics and Technological Mediation," Science, Technology & Human Values, Vol. 31, No. 3, pp. 361- 380, 2006.

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

[32] D. Berdichevski, and E. Neuenschwander, "Toward an Ethics of Persuasive Technology," Communication of the ACM, Vol. 45, No. 5, pp. 51-58, 1999.

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

 ISBN: 978-1-61208-557-9