

# Situational analysis of the communication flow in audiovisual media

Carlos Aguilar<sup>a</sup>, Lydia Sánchez<sup>b</sup> and Manuel Campos<sup>c</sup>

<sup>a)</sup> *Universidad de Barcelona. Facultat de Formació del Professorat. Campus Mundet. Edifici Llevant, 1<sup>a</sup> P.; Passeig de la Vall d'Hebron, 171. 08035 Barcelona. [Caraguilar@gmail.com](mailto:Caraguilar@gmail.com); <http://www.ub.es/devp>*

<sup>b)</sup> *Universidad de Barcelona. Facultat de Formació del Professorat. Campus Mundet. Edifici Llevant, 1<sup>a</sup> P.; Passeig de la Vall d'Hebron, 171. 08035 Barcelona. [lsanchezg@ub.edu](mailto:lsanchezg@ub.edu); <http://www.ub.es/devp>*

<sup>c)</sup> *Universidad de Barcelona. Facultat de Filosofia. Departament de Lògica, Història i Filosofia de la Ciència. C/ Montalegre, 6. 08001 Barcelona. [mcamposh@ub.edu](mailto:mcamposh@ub.edu)*

**Abstract:** This paper presents a mathematical framework for the study of the information contained in audiovisual contents based on the development by Keith Devlin of Situation Theory. In order to obtain this framework, we present accounts of the processes carried out by agents from the reception of audiovisual contents to the extraction of information, in accordance with the definition by Israel and Perry. We finally justify why these accounts concerning the extraction of information from audiovisual content can be included as part of a mathematical formulation of Situation Theory

**Keywords:** Situation Theory, audiovisual media, semantic information, meaning

**Acknowledgements:** Bitrum project managers

**In the proposed framework** for the study of information in audiovisual contents, we need three differentiated theoretical blocks, each sustained by different mathematical formulations.

The first block refers to the set of physical laws that govern the process of emission and transmission of contents. In this first part, the formulation of the mathematical theory of communication of Shannon (MTC) is used as reference frame.

The second analyzes the process of individualization of content realized by the receiving agents. Assuming the postulates of Situation Theory, all agents (human or machines) have a limited amount of sensors that determine their capacity of individualization of the items contained in contents. We consider this process under the point of view of automated recognition systems.

The third justifies the application of Situation Theory and the principles of the mathematical logic formulated by Keith Devlin

(Devlin, 1991) to the study of the informative content included in audiovisual contents. We focus on how it is possible to use the items obtained in the individualization processes as the basic units of information, and set out to determine how elements of the ontology of Situation Theory can be obtained from them. After that we explore how to initiate the process of construction of meaning, in the sense expressed by Israel and Perry (1989). Once these three types of theoretical analyses are presented, we show that it is possible to unify them within the situational framework.

## 1. Extending Situation Theory to audiovisual content analysis

Once the theoretical framework for the transmission of information process is proposed, we consider the extraction, storage, arrangement and recovery of the elements that compose the audiovisual contents. The disciplines that take care of these processes

are Multimedia Content Analysis and Multimedia Information Retrieval (K-Space D3.1, 2006) (Lew et al. 2006) (Hare et al., 2006.) The set of processes and techniques used by these disciplines concentrate on the extraction of significant objects at different levels: from the lowest, concerning the physical components of the content, to the highest, concerning semantic states and objects. All these techniques are based on a wide variety of mathematical treatments oriented to obtain the most precise results possible.

### 1.1. The semantic gap

Once the content transmission system is framed theoretically within the MTC, and the appropriate methods for the identification of meaningful elements selected, the extraction of information is the culminating point in the pyramid of events whose final obstacle is the so-called *semantic gap*. In Hare (2006) we can find a compilation of procedures used to overcome this gap, which help to make this concept more precise. Thus, Smeulders (2000 p. 1349-1380) tells us that the semantic gap is described as

“... *the lack of coincidence between the information that can be extracted from the visual data and the interpretation that the same data have for a user in a certain situation.*”

The obtaining of significant information is, however, not a new problem, and it is not associated solely to the present formulation. Even if, because of the novel character of the study field, the problem appears to have a special relevance, the question of assigning meaning to a data set, as well as its variability depending on the reference situation, is as old, one could say, as communication. Facing the semantic gap forces us to clarify some concepts such as those of *data*, *information* or *knowledge*. To do that, we resort to their mathematical representation. In order to deal with the challenge of the extraction of knowledge from the significant elements of audiovisual contents, we use the mathematical formulation of Situation Theory. By means of this formulation, we identify the elements extracted through multimedia analysis processes as elements of situation

ontology. This allows us to give theoretical support to the process of extraction of information.

### 1.2. Infons

Following Smeulders' characterization of the semantic gap, we use the mathematical development of Situation Theory by Keith Devlin (1994) to face the third part of our job. We present this approach as a possible alternative when it comes to formulating the mechanisms of data processing in audiovisual contents.

Situation Theory is an attempt to provide a mathematical formulation for Barwise's and Perry's (1983) Situation Semantics. In *Logic and Information*, Keith Devlin (1991) supplied Situation Theory with a mathematically coherent apparatus. The basic ontology of Situation Theory is constituted by the entities that a cognitive agent is able to individualize and/or discriminate, given its limitations. This refers us to the results we obtain using the methods of Multimedia Content Analysis and Multimedia Information Retrieval. These systems provide us with *individualizations*, elements that are discerned from others in the context of general content. Among the objects, also known as *uniformities* (or *regularities*), of situation ontology, we find individuals, relations, spatial locations, temporary locations, situations, types and parameters. The framework, relative to the agent, that gathers the ontology is known as “individualization scheme” (it is the appropriate one for the study of the flow of information pertaining to the agent). The information must always concern a particular situation, and it is represented through discreet items known as “infons”. An infon is an object of the form:

$$\langle\langle R, a_1, a_2, \dots, a_n, 0 \rangle\rangle, \langle\langle R, a_1, a_2, \dots, a_n, 1 \rangle\rangle$$

where R is the relation established among  $n$  objects ( $a_1, \dots, a_n$ ), and the last element (1 or 0) is called “polarity”, and shows the veracity:

$$\langle\langle R, a_1, a_2, \dots, a_n, 1 \rangle\rangle$$

or falseness:

$$\langle\langle R, a_1, a_2, \dots, a_n, 0 \rangle\rangle$$

of the instantiation of R in ( $a_1, \dots, a_n$ ).

According to Situation Theory, infons are semantic objects within the mathematical theory; they are not sentences in some language which require an interpretation. They are the minimum units of information. They can concern more than one relation by incorporating the operations of conjunction and disjunction. In the latter case, they are called “compound infons”, as for instance, the one that represents the shout: “Fire!”:

$$\langle\langle shout, \dot{a}, \dot{t}, 1 \rangle\rangle \wedge \langle\langle word, \dot{a}, FIRE, \dot{t}, 1 \rangle\rangle$$

Given a situation  $s$  and an infon  $\sigma$  we write:

$$s \models \sigma$$

in order to indicate that the infon  $\sigma$  is a “fact” for situation  $s$ . In other words, we can say that  $\sigma$  is an informational item that is true in situation  $s$ . Thus, from the situational perspective, information is treated as merchandise. Merchandise that, besides, does not always have to correspond to the value “true”. Thus, for each infon, there is a dual negative, which can be understood as its informational opposite, so that only one of them can be true. The application of this set of laws, derived from the mathematical formulation of Situation Semantics, to the identification of structures and regularities within multimedia contents, constitutes the necessary first step in the proof of the validity of the situational framework for the analysis of contents. The second step involves the recognition of relations among the previously identified elements, and the treatment of these relations according to the information concept as formulated within Situation Theory. Expressed in situational terms, we can consider, for instance, that each one of the regions extracted in the analysis of a photogram can be expressed as a compound infon, in the following manner:

$$\begin{aligned} &\langle\langle Connected, p_1, p_2, p_3, \dots, p_n; 1 \rangle\rangle \wedge \\ &\wedge \langle\langle Pixel, p_1; 1 \rangle\rangle \\ &\wedge \langle\langle Pixel, p_2; 1 \rangle\rangle \wedge \dots \\ &\wedge \langle\langle Pixel, p_n; 1 \rangle\rangle \end{aligned}$$

Where  $p_1, \dots, p_n$  are the points included in the region. In turn, the definition of the infon that represents the *Connected* relationship is formulated in the following manner:

$$\begin{aligned} &\langle\langle Connected, p_1, p_2, p_3, \dots, p_n; 1 \rangle\rangle = \\ &\langle\langle Position_1, p_1; 1 \rangle\rangle \wedge \langle\langle Intensity_1, p_1; 1 \rangle\rangle \wedge \dots \\ &\langle\langle Position_i, p_i; 1 \rangle\rangle \wedge \langle\langle Intensity_i, p_i; 1 \rangle\rangle \wedge \dots \wedge \dots \\ &\langle\langle \Delta Position_{ij}, 1; 1 \rangle\rangle \wedge \\ &\langle\langle \Delta Intensity_{ij}, < threshold; 1 >; 1 \rangle\rangle \wedge \dots \end{aligned}$$

Where all points meet the condition of closeness with at least another point in the region:

$$\langle\langle \Delta Position_{ij}, 1; 1 \rangle\rangle$$

and they also meet the condition of being in the threshold interval:

$$\langle\langle \Delta Intensity_{ij}, < threshold; 1 >; 1 \rangle\rangle$$

By means of this formulation we can characterize the extracted regions as minimal informational elements (infons) for each photogram. The following step in the process of construction of the meaning entails the characterization of situation types from the set of infons obtained; that is, the identification of those subgroups of infons that allow the creation of a typology.

### 1.3. Constraints

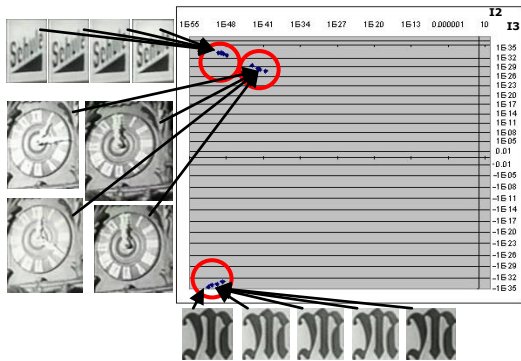
Constraints are abstract bonds between types of situations. They can be of different varieties: natural laws, empirical conventions, logical relations, linguistic rules. Their role in the flow of information is represented in Israel and Perry (1990) through the word “meaning”.

The existence of a constraint between two types of situations,  $T$  and  $T'$ , indicates that any instance of type  $T$  will carry information concerning an instance of type  $T'$ , within the restrictions determined by the global situation that involves both instances. Thus, in order to acquire significant information, successive constraints should be established between any two situation types.

The task of identification, characterization and discrimination of the significant elements of the image or the sound of an instance of audiovisual content implies, generally, the extraction of great amount of descriptors (from a few to thousands) (Lowe, 2004), as well as the use of statistical means and/or discriminative systems, focused on the determination of significant regions within the

characteristics space formed by the set of descriptors.

Akin elements represented in this space tend to occupy connected regions within it.



**Figure 1:** Location of extracted elements in the characteristics space.

These spatial constraints on the characteristics space allow us to identify the existing types. These types of objects are determined on a given situation. Taking  $s$  to be the situation, if  $\dot{x}$  is a parameter and  $\sigma$  a compound infon (generally involving the said parameter), then there is a type (object-type):

$$T = [\dot{x} \mid s \mid \sigma]$$

such that it is the type of the objects  $x$  for which the parameter can be anchored in situation  $s$ , a situation in which the conditions imposed by the infon is obtained.

In order to build the meaning, the agent must be able to establish constraints between each one of the situations types identified in the context of the situation. The representation of the infon that accounts for the constraint between two situation types  $s$  and  $s'$  is the following:

$$\langle\langle \Rightarrow, s, s', 1 \rangle\rangle$$

Notice how, on the basis of the analysis of the items extracted from the content, we can easily establish spatial, temporal or hierarchical constraints between the types, but we can hardly establish complex relations concerning ideas or sensations. Although complex reasoning can be reflected in the flow of images, the set of structures with which a human viewer can compare them is infinitely wider than the one obtained by the extraction of situation types from the content.

A knowledge data base turns, then, to be unavoidable: it will allow us to establish the constraints on the objects that will allow us to have access to a higher level of informative meaning for the showed images.

## 2. Situational review

Seen in perspective, we can say that this framework uses three types of analyses for the treatment of multimedia contents: the MTC for the process of transmission, the different techniques derived from the analysis of multimedia contents, and Situation Theory in the process of obtaining significant information. Is it possible to unify the three types within a single theoretical frame? What we show in the present section is the justification for the inclusion of the MTC and the analysis of multimedia content within the theoretical frame of the mathematical formulation of Situation Theory.

It is not difficult to establish that the procedures for the identification of significant elements within contents can be assimilated to the extraction of infons, and the grouping of these, based on common characteristics, to the extraction of types in Situation Theory. The identification between objects and infons can be non-trivial, but, in any case, it is conceptually simple. To include under Situation Semantics the conclusions obtained from the use of the MTC, we need to resort to the bases of the process of transmission of signals and to the elements that take part in it. In this process, the symbols of the source's alphabet are emitted through a physical and possibly noisy channel, and collected by a receptor as elements of the receiving source alphabet (Shannon, 1948), (Pérez, 2007). Given this background, and from a situational perspective on the process, the elements of the alphabet of the receiving source are characterized as situation types. If it is possible to characterize the different items of the sources' alphabets as elements pertaining to situation types, then we can reinterpret the significant items of the MTC in the following way:

- The mutual information corresponds to the number of constraints that can be established between both alphabets.

- The noise corresponds to the uncertainty associated to the impossibility to establish a unique constraint between the elements of the sources.
- Entropy, which in the MTC is taken as a measurement of the amount of potential information, corresponds to the amount of possible constraints that could be established in a univocal relation of the alphabet of the emitting source and an appropriate alphabet.

It is evident that the account, in situational terms, of the transmission, analysis and identification processes, is extremely tedious, and does not provide any additional value to any of the processes. In spite of this, the formulation of these processes in situational terms allows for the inclusion of the proposed framework in a unique theoretical perspective, the mathematical formulation of Situation Semantics.

### 3. Conclusions and future works

In this work we have presented a framework for the treatment of the different procedures involved in the processing, analysis and extraction of information contained in audiovisual contents. We have presented the theoretical principles on which it is based, and have shown how the use of the mathematical formulation of Situation Semantics realized by Keith Devlin can be

used in this context. We have also shown how one can extend the principles of Situation Semantics to the analysis of images in multimedia contents, demonstrating the utility of the proposed methodology for the study of the processes of transmission of information mediated by multimedia contents. The presented development allows us to conclude, as well, that it is necessary to advance in several directions in order to take advantage of the potential of analysis offered by the mathematical formulation of Situation Theory. It is necessary to improve the process of identification of objects within contents and their representation as infons in Situation Theory. Another aspects requiring improvement are the characterization of the described situations on the basis of the constraints existing between types as well as the relation that these situations may establish with the mathematical procedures used by the Multimedia Content Analysis in the extraction of semantic content. Exploring these possibilities is necessary in order to develop the basic elements of the mathematical formulation of Situation Theory, so that the highest levels of semantic complexity are reached. Finally, we have seen the necessity to choose a suitable reference framework as well as a cluster of reference situations sufficient to allow us to establish a set of significant constraints that allows the characterization of contents in an adequate manner.

## References

- Barwise, J., Perry, J., (1983). *Situations and Attitudes*. Cambridge: MIT-Bradford.
- Devlin, K. (1991). *Logic and Information*. Cambridge, UK: Cambridge University Press.
- Hare, J. S., Sinclair, P. A. S., Lewis, P. H., Martinez, K., Enser, P. G. B. and Sandom, C. J. (2006). Bridging the Semantic Gap in Multimedia Information Retrieval: Top-down and Bottom-up approaches. Mastering the Gap: From Information Extraction to Semantic Representation / 3rd European Semantic Web Conference. [On line] <<http://eprints.ecs.soton.ac.uk/12737/>>
- Israel, D., Perry, J. (1990). What is information? In Philip Hanson (Ed.). *Information, Language and Cognition*. Vancouver: University of British Columbia Press, pp. 1-19.
- K-Space D3.1 (2006). FP6-027026, K-Space D3.1: State of the art on multimedia content analysis [on line]. K-Space 2006. [http://kspace.qmul.net:8080/kspace/deliverables/KS\\_D3.1\\_20060531-MM\\_content\\_analysis.pdf](http://kspace.qmul.net:8080/kspace/deliverables/KS_D3.1_20060531-MM_content_analysis.pdf)
- Lev, S. M., Sebe, N., Djeraba, C, Jain, R. (2006) Content. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1), 1-16.
- Lowe, D., G. (2004). Distinctive Image Features form Sale-Invariant Keypoints. *Int'l J. Computer Vision*, 60(2), 91-110.
- Pérez-Montoro, Mario (2007). *The Phenomenon of Information. A Conceptual Approach to Information Flow*. Lanham (Maryland): Scarecrow Press. ISBN 978-0-8108-5942-5.
- Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379-423, 623-656.
- Smeulders, A., Worring, M., Santini, S., Gupta, A. y Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22(12), 1349-1380.

## Authors

### Carlos Aguilar

is current Ph.D Student for the DEVP department of the School of Education at the Universitat de Barcelona. Currently teaches in the ESCAC (Escola Superior de Cinema i Audiovisuals de Catalunya).

### Lydia Sánchez

is Ph.D. in Philosophy by Stanford University. She is currently teaching Communication in the DEVP department of the School of Education at the Universitat de Barcelona. She has taken part in the doctoral programs \*Comunicació, Art, Educació\* and \*Formació del professorat: pràctica educativa i comunicació\*. She also teaches in the Humanities, Psychology and Documentation departments of the UOC. Her research focuses in the theory and philosophy of communication. She is author of different articles and contributions to books, and co-edited "Industrias de la comunicación audiovisual" (2008).

### Manuel Campos

is Ph.D. in Philosophy by Stanford University. Currently teaching at the Logic Dept. of Universitat de Barcelona. Interested in Philosophy of Language and Communication, and Philosophy of Science.