A Knowledge-oriented Approach to Semantic Changes

Daniel GALARRETA
CNES
18 avenue Edouard BELIN
31401 Toulouse Cedex 9
335 61 27 32 51
daniel.galarreta@cnes.fr

ABSTRACT

In this paper we present methodological solutions that were proposed for maintaining knowledge attached to long term space missions and to scientific data. It turned out that the question of maintaining the intelligibility of the information processed, i.e. facing the semantic change issue, was closely associated to the knowledge preservation issue as its very motivation. This orientation had a significant impact upon the nature of the research and the type of methodology that was developed at a later stage: the research was multidisciplinary and the methodology evolved toward a multi-viewpoints semiotics. However if the methods we used allow us to detect semantic changes they cannot guarantee to allow us to detect knowledge changes. The reason is that knowledge is a semiotic notion and not a semantic one. This paper is an epistemological contribution to the issue of knowledge changes.

CCS Concepts

• Information systems → Information retrieval → Document representation → Content analysis and feature selection.

Keywords

Semantic changes; knowledge; semiotics; epistemology.

1. INTRODUCTION

The long term space missions (which duration is at least ten years long) set the teams in charge of them a challenge regarding the maintaining of the knowledge attached to them. The knowledge that was produced or called for during design phases are no longer available because the original teams have left and the technological contexts have significantly changed. A possible strategy to face this issue is detecting at an early enough stage the evolutions or the breaks of knowledge that are likely to occur in order to proactively respond to them.

If we accept the idea that the productions of knowledge as well as its mobilization use natural language, the question is then detecting linguistic changes that could correspond to critical knowledge evolutions with respect to the success of the mission.

The Rosetta mission provides a perfect example of such a long term mission. In November 1993, the Rosetta mission was approved as a Cornerstone Mission in ESA's Horizons 2000 Science Programme. Rosetta was launched in 2004 and arrived at Comet 67P/Churyumov-Gerasimenko on 6 August 2014. It was the first mission in history to rendezvous with a comet, escort it as it orbits the Sun, and deploy a lander to its surface. There is more than twenty years between the decision and the landing not to mention the phase of exploitation of collected data!

Exploitation of scientific data long after their production or by a new scientific community is also related to the problem of knowledge variations and their detection: the community designed to exploit given data "may lose its familiarity with some terminology, and the definition of the community may be broadened to include other members with different backgrounds" [1].

The question of semantic change has received an always growing attention. Several approaches have been proposed for instance in linguistic which account for synchronic and diachronic variations [or based upon metaphors: spatio-temporal metaphors of conceptual spaces or physical metaphors of linguistic dynamics [2].

We will first present the studies that were conducted in an R&D program between 2002 and 2012 by the CNES in order bring solutions to the question of detecting knowledge variations. Concerning knowledge related to equipment of space systems in long term missions, the adopted approach was multidisciplinary analyses of the semantic changes that occur in documents available to the teams: statistics, knowledge engineering, logics and linguistics. Concerning knowledge attached to data and scientific communities, we restricted to the detection of knowledge community.

We will stress the fact that despite these approaches provide means for detecting interpretable semantic evolutions they do not bring a general to the detection of knowledge variations.

We will argue that the reason lies in the fact the issue of knowledge variations is semiotic and not semantic. In order to do that it is necessary to have a semiotic definition of knowledge. It is for this reason that we will successively introduce semiotics, Hjelmslev's semiotics and multi-viewpoints semiotics.

2. DETECTING KNOWLEDGE VARIATIONS THROUGH SEMANTIC CHANGES

In this section we present the key ideas and methodologies that were used in an R&D program conducted by the CNES between 2002 and 2012.

2.1 A quality requirement in its own right

In the context of an industrial approach the maintaining of knowledge about a space project, its segments, its instruments, its operations, etc. is a quality requirement in its own right. It pertains to a quality approach. Let us call this *quality factor* a "memory quality".

In order to insure the "memory quality" of a project is satisfied, it is necessary to insure that the memory solutions of the project are consistent with the requirement of intelligibility of the information processed.

These solutions relate elements such as (but not limited to) storage devices, to people equipped with knowledge and immersed within a social and technical environment. This insurance concerns projects and teams organization issues as well as documents and information issues. In the latter case, the question is to re-contextualize the documents and information. Since we admit that the context cannot be *saved* properly, it must be *produced* in a dynamic manner:

- By identifying (when possible) the viewpoints in presence when the documents were produced, and the readers ‘viewpoints
- By building correspondence networks between documents (using statistics tools)
- By re-contextualizing documents that are considered as significant by re-reading or making them commented.

Let us go further in order to implement these suggestions.

2.2 Detecting what could have changed

In order to insure the “memory quality” requirement regarding a given project, it is necessary (but not sufficient) to detect knowledge evolutions soon enough in its life time.

A simple idea is to suppose that knowledge which is attached to this project cannot be produced in isolation and involved context knowledge. Thanks to that context knowledge, project knowledge is either understandable or usable. In other words project knowledge presupposed the existence of context knowledge. If later context knowledge evolves significantly, the project knowledge runs the risk of being beyond understanding or useless.

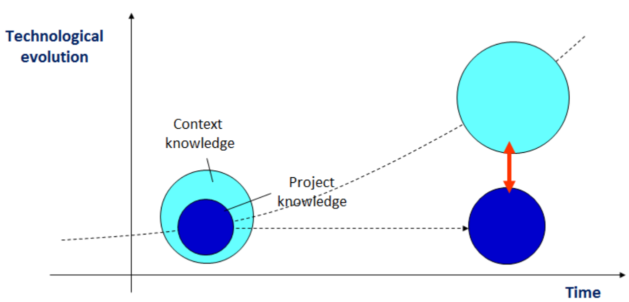

Figure 1. Detecting knowledge evolutions.

Despite its apparent clarity this idea hides several difficulties. First of all, knowledge should be defined in a way which both agrees with common sense and allows measurement of knowledge evolutions.

Another issue is to determine the perimeter of the initial project knowledge and correspondingly of the context knowledge at different steps of the project life.

Regarding the first issue one can decide to use the texts produced during the life time of the project as *representations* of the knowledge attached to the project.

Regarding the distinction between *project knowledge* and *context knowledge*, it is in practice reduced to the documents attached to the project (in our case the initial functional requirements of that piece of equipment) and to reference documents including new versions of these requirements and documents for different phases of the project.

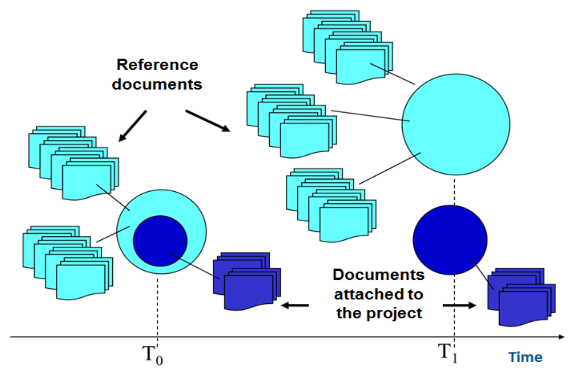

Figure 2. Documents as knowledge supports.

Practically, using these documents, evolutions are to be detected through a blended approach that will take advantage of:

- Statistical distribution of terms within the project documents
- Impact on taxonomies of domains of the project
- Logical relations between “interesting” concepts
- Linguistic cues extracted from documents

2.3 Statistical distribution of terms within the project documents

The documents written in natural language are considered as lists of words limited to content words (such as nouns, verbs, and adjectives) as distinct from function words such as prepositions and articles. The treatment consists in “coding them, in order to obtain quantitative data for computation. Then, statistical methods may be applied, such as graphic representations, exploration and modelling. The essential method used in this treatment is Correspondence Analysis” proposed by JP. Benzecri [3].

The correspondences between the texts categories and their contains are then analyzed. “For example, one can apply this method for the study of the evolution of a set of technical reports through time. Another application is the comparison of the vocabulary of several authors” [3].

“The tools for these analyses are issued from linguistic, mathematical and computational models. They are implemented in several text mining softwares, such as SAS Text Miner, Alceste, Le Sphinx Lexica, SPAD” [3].

2.4 Impact on taxonomies of domains of the project

The idea here is to measure how a corpus of documents fits a taxonomy that represents the knowledge domain. More precisely considering a taxonomy as a tree of terms, we can observe how these terms hook the documents up to that tree – a document being hooked if it contains such a term. Doing so we obtain different fitness measures depending on whether we consider the coverage (in number of documents) (a) of sibling branches of the taxonomy, (b) of the different descendents of a common “forefather term”, (c) of different categories of the taxonomy. Case (a) produces a *contribution* measure, case (b), a *specificity* measure and case (c), a *concentration* measure (see [3]). Measures of this sort can be done with different corpora (corresponding to different periods of time).

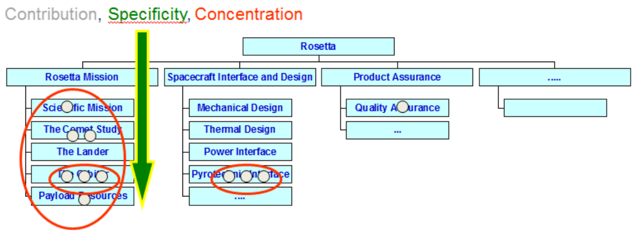

Figure 3. Distributions of documents over a taxonomy.

The detailed presentation of the different metrics is [4]

In that approach we consider that the production a corpus of texts is not a prerequisite condition for the production of taxonomies or ontologies. We rather posit that such taxonomies already exist even in a rough form. By using comparison tools (between texts and taxonomy, between texts using taxonomy or between parts of taxonomy using texts), we can simultaneously improve the taxonomy quality and the characterization of the corpuses of texts.

Even if this method allows synchronic comparisons beteween corpuse, it allows also diachronic comparisons. The question is to track knowledge evolutions in a time series of corpuses on the same subject about the project [4].

2.5 Logical relations between "interesting" concepts

In this approach documents are represented by n-dimension vectors (where n is the number of significant words in a document).[5]. We produce association rules between significant words. An association rule links a set of words with a single word: W => w. W is the set of words and w is a word. [5]. The rule has to be read as follows: when the words belonging to W appear in a document, then the word w is also mentioned" [5]. Here words and their relations are considered as descriptions of knowledge.

We look for differences that might not be obvious between documents or sets of documents, using association rules. In other words, the extraction rules process should show: (a) documents that seem very close but produce different rules, (b) association rules that have been extracted from documents that are different. In fact, we want to exclude the rules that have been produced by considering documents that are very close"[5].

Then considering successive generations of documents, we put the emphasis on the concepts (words and relations) that are specific to each generation.

2.6 Linguistic cues extracted from documents

The proposed approach uses linguistic knowledge to identify variations of functioning between two bodies of text (two corpora) written in different periods. These variations are mainly lexical in nature. The idea is to retrieve cues that can be applied systematically in order to retrieve changes. Three methods of analysing will be explored.

Formal cues

Such as: disappearance of terms, appearance of terms, lengthening of terms, abbreviations (e.g. Attitude and Control Orbit System => AOCS), ellipsis (e.g. earth observation satellite => observation satellite), etc.

Semantic cues

This way of looking concerns terms that are correlated by means of conceptual relations: hierarchical, causative, part-whole, etc. There exist *patterns* to systematically retrieve these relations. […] The goal is to retrieve variations in the way terms are related by systematically applying these patterns. Relations can appear or disappear, or a terms can be related to a different terms with respect to the previous text.

Statistical cues

One could say that the distribution of a word (the set of contexts in which it appears) could help to outline the meaning of a word [6] […] temporal variations in distribution could be identified, to which one could associate a variation in the semantic coverage [3].

2.7 Detecting critical knowledge evolutions by using the linguistic expressions of risks

Since a loss of knowledge in the case of a long term mission, may affect the project achievement, we must be able to make a diagnosis about the "memory quality" taking into account the criticality of the components that constitute that project.

There are at least two reasons for that:

- Estimation of criticality of the components (and of their sub-components) provide a prioritization for applying the different analyses we described above;
- A risk is akin to knowledge, more precisely, is a piece of knowledge. The justification of that position rests upon the semiotic approach we adopted concerning, *information*, *knowledge* and *data*. (See section 3.3 below in this text). Therefore risks are markers of textual components (concepts, relations) the intelligibility of which are worth maintaining.

The importance of criticality leads us to use a linguistic analysis in order to meet two objectives. This is first to uncover patterns of expression of risk in documents and then use them in order to control the evolution of knowledge in the documents. Indeed, once the modes of expression of risk have been described, they can be used to locate objects considered at risk from the start of a project. These "risky" items then will be monitored closely throughout the project to ensure that criticality is not forgotten. Furthermore, the systematic implementation of the terms of risk throughout the project will enable us to quickly identify new objects that are associated with a risk.

2.7.1 *Building a grammar of risks*

We admit that the expressions of patterns of risk are not infinite and that we can produce a description the lexical and syntactic rules that govern them and therefore a grammar of risks is conceivable.

The schema we adopt is the following:

- Detect in texts denotated « objects » (i.e. extra linguistic objects) for which risks are considered at the beginning of the project.
- Trace these objects during the project life in order to ensure that their criticity is not forgotten.

For instance identification of linguistic structures signifying risks and allowing to identify names of risky objects can uses pattern such as:

<Name_1> detect <Name_2> (on <Name_3>)

<Name_1>:?,

<Name_2>: default, failure,

<Name_3>: sensor

2.8 Scientific data: Maintaining access to data archives for new communities

In the case of a scientific experiment, the knowledge which is necessary in order to reuse scientific data is the knowledge which let the data "do the talking". However this knowledge can evolve as time goes by since the way to be interested in may change: the experimenters at the start are no longer there or the target community of users changed [1].

2.8.1 Community mining

This approach tries to use the links between web pages to identify and characterize the virtual communities on the Web.

Two major trends can be identified: the analysis of links and identification of communities.

The link analysis is based on the intuition that the pages are gaining more authority by being referenced by pages that have themselves a great authority. This recursive definition is based upon Pages Rank algorithms that are at the source of the classification of results by the search engines on the web. An extension is the characterization of communities through pages which are authoritative but also through pages with more links to authoritative pages (Hubs) (see [7]).

The identification of communities exploits the links starting from a few initial pages then identifies within the graph of links a subgraph that points out more to himself than to the outside (Maxflow / Mincut algorithm used by the FLG method (see [8]).

The various approaches described above have been implemented and used in different contexts.

The first task was to use the Hits algorithm with requests targeting domains we chose (e.g. " virtual observatory "). The best "authorities" and the best "hubs " were identified , then an adjacency matrix between sites was produced in order to perform a classifications of these sites.

The second task was to identify communities by FLG algorithm. Taking a given scientific community and its website as a reference, we can use one of its pages on a given topic and sets of sites that are authorities on other topics to estimate the weight that of this community for each topic by evaluating inclusions of the identified communities.

2.9 The Results and their discussion

2.9.1 The results

Concerning long term space projects we chose the DORIS (Doppler Orbitography and Radiopositioning Integrated by Satellite) system. It is a French satellite system used for the determination of satellite orbits (e.g. TOPEX/Poseidon) and for positioning. It was first used in 1992 on TOPEX/Poseidon Satellite (1992-2004), on SPOT satellites, SPOT 2 (1990-2009), SPOT 3 (1993-1995), SPOT 4 (1998- 2012), SPOT 5 (2002-2015), on Envisat Satellite (2002-2013) and on Jason satellites, Jason 1 (2002-2013), Jason 2 (2008-2015)

We chose a particular element of this system (the DORIS beacon) and considered the knowledge attached to it and its evolution. We restricted our analysis to three generations of beacons, which corresponded to successive technological evolutions (miniaturization of electronic components which help to reduce its weight and its volume). The first was developed in 1984, the second from 1996 to 1999 and the third from 2000 onward.

The methods that we have sketched out were applied to theses corpuses. It is not possible to present here all the results obtained. See ref CNES Let us consider to examples.

In the case of logical relations between "interesting" concepts, we were able to detect specific rules for the1st and 2nd generation and for the 3rd generation of documents respectively. Note that these two networks have no common terms.

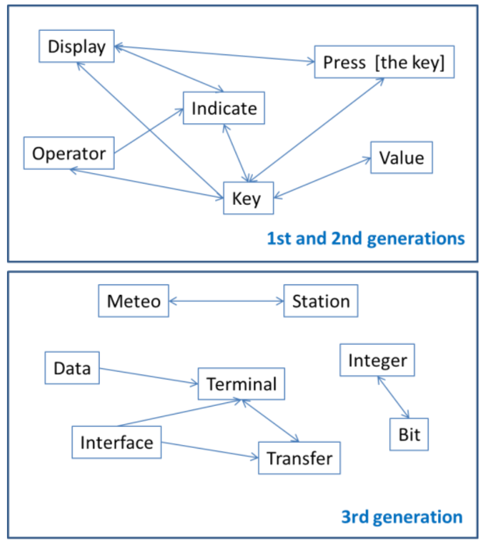

Fig. 4: Specific rules for 1st and 2nd generation and for the 3rd generation of documents respectively

In the case of the impact on taxonomies of domains of the project (fig. 5) :

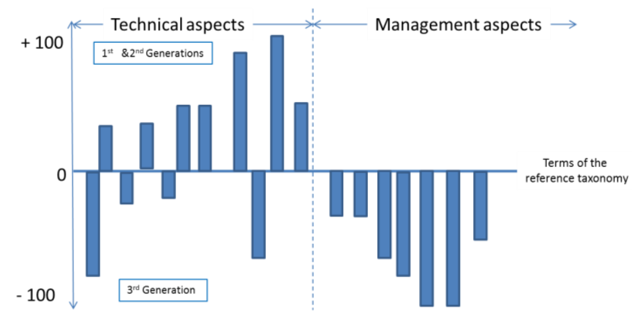

Fig.5. : degree of attachment of terms of the reference taxonomy

We can observe that the Management aspects in documents are only present in the 3rd generation.

In the case of the community mining, communities were computed starting from the web Site of the CDPP (Centre de Données de la Physique du Plasma - https://cdpp-archive.cnes.fr/) and more precisely from the CDPP page describing the " magnetosphere " and finally from a set of sites that are authorities in the field of " plasma space ." The goal was to estimate the weight that the CDPP in either field by evaluating inclusions of the identified communities.

For instance, using the FLG algorithm, we build the community stemming from the web site of the CDPP, then from a page of this site describing the "magnetosphere". Then a community is built from the set of sites which have the best authority on the theme of "space plasma". We then estimate the weight that the CDPP has in one or the other domain by evaluating the inclusions of the communities thus identified.

2.9.2 Do we solve the initial question?

What is worth underlining here is that from the very beginning of our research the question of maintaining the intelligibility of the information processed, that is facing the semantic change issue, was closely associated to the knowledge preservation issue as its initial motivation. In other words we posited that *semantic changes* are a means to detect *knowledge variations*.

But is this need to be carefully examined.

We will use the results obtained in the case of logical relations between "interesting" concepts, to sustain our reasoning.

We have noted that there is an evolution of the specific rules between the 1st and 2nd generation and for the 3rd generation of documents. We observed for instance, in the case of the 1st and 2nd generation, that the presence of the terms *Operator*, *Display*, *Indicate*, *Press [the key]*, *Value*, are necessary conditions for the presence of the term *Key*. In other words, the existence of an interface of the terminal utilized to configure the beacon and using keys implies the existence of these terms. The specific rules of the 3rd generation support the existence of a purely digital interface.

Let us be more precise. In the case of the 1st and 2nd generation, the graph of relations is interpretable because of the existence of an adequate viewpoint which produces the above interpretation when confronted to the corresponding *graph*. The *possibility of producing* a meaning *from this graph* is not a *semantic* issue but a *semiotic* one. We will return to that point. The meanings themselves definitely pertain to the semantic analysis. But what relations do they maintain with reality?

Does the interpretation of the graph of the 1st and 2nd generations as terminal of programming which is used to configure the beacon by monitoring a microprocessor that controls all the functions, (fig. 6) correspond to a real operator actually pressing real keys? Or are we just conceiving a fictive terminal of programming for a fictive character virtually acting on virtual keys configuring a non-yet existing beacon (fig 7.)?

Fig. 6. Actual terminal of programming used to configure a DORIS beacon

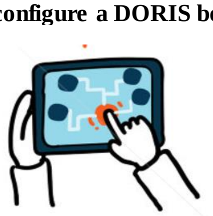

Fig. 7. Fictive terminal of programming used to configure a DORIS beacon

The difference between these two interpretations corresponds to the difference between knowledge and non-knowledge.

Let us try to analyze even further this difference. For clarity we use the figures 6 and 7 as representations of a terminal of programming of a Doris Beacon:

- Figure 6 could be interpreted as the representation of a box the dimensions of which are approximately 60 cm wide, 40 cm deep and 40 cm high with two black handles on its front side. This interpretation is both different and compatible with the interpretation of figure 6 as a view of a terminal of programming which is used to configure the beacon by monitoring a microprocessor that controls all the functions.
- Figure 7 could be interpreted as a view of a person's hand squashing bugs on a tablet. This view is different and non-compatible with the interpretation of figure 7 as a view of a terminal of programming used to configure a DORIS beacon.

In order that an interpretation of a representation of something corresponds to knowledge of something, it is not sufficient that this representation can be interpreted adequately with respect to a reference viewpoint. It is necessary that this representation can be interpreted by with respect to another viewpoint producing another interpretation which is both different and compatible with the previous one.

We observe that we have evolved from semantic issues to epistemological and semiotics issues that we should now examine.

3. A SEMIOTIC DESCRIPTION OF KNOWLEDGE

3.1 What is semiotics?

Roughly speaking Semiotics can be defined as a theory of signs or as a theory of language.

Considered as a theory of signs, it is akin to the Aristotle conception of language: "Spoken words are the symbols of mental experience and written words are the symbols of spoken words. Just as all men have not the same writing, so all men have not the same speech sounds, but the mental experiences, which these directly symbolize, are the same for all, as also are those things of which our experiences are the images"(see [9]. pp. 77-78 (I, 16a, 3-8)).

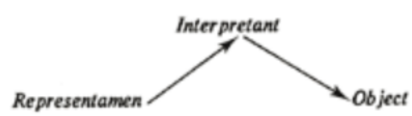

Figure 8. Triadic model of a sign.

C.S. Peirce (1839 -1914) is the most distinguished representative of this trend. He developed a semiotic theory that is at once general, triadic and pragmatic. It is *general* since it takes into consideration emotional, practical and intellectual experience; it includes all of the components of semiotics; and it broadens the concept of the sign. It is *triadic* in that it is founded upon three philosophical categories: firstness, secondness and thirdness; brings three terms into relation: the sign or representamen, the object and the interpretant. It is *pragmatic* since it takes into consideration the context in which signs are produced and interpreted; and defines the sign by its effect on the interpreter [10].

There are other triadic theories of signs such as Charles Morris's (1901-1979) "who drew on Bloomfield's linguistic behaviourism, followed Peirce's semiotics: to determine the meaning of any sign 'we have . . . simply to determine what habits it produces' " (see [11]).

Considered as a theory of language, it is akin to the linguistic theory of F. De Saussure (1857-1913). As Norris expressed the Saussurian approach "open the way to a structuralist account that left no room for naive (pre-scientific) ideas about the one-to-one 'correspondence' between words and ideas or words and objects". [12]. The Saussurian linguistic is "based upon cardinal oppositions between langue and parole, synchrony and diachrony, the paradigmatic and the syntagmatic, and the orders of signifier and signified" (see [9]). It is at least for this reason that the sign is said dyadic.

Let us note with Paul Bouissac [13] that despite the fact that Saussure is generally considered as "a major fountainhead of semiotics" such an opinion is based on a short paragraph in the Course in General Linguistics and on a few remarks scattered throughout the book:

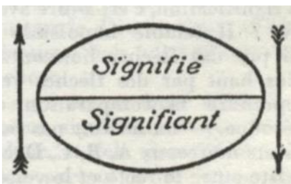

Figure 9. Dyadic model of a sign.

3.1.1 Semantics vs semiotics

A theory of language can help deciding if the traffic lights, the chiming clock, the dial on a phone produce sequence that can be considered as languages, in the same sense as that we can say that English or French for instance are languages. This semiotics should also be able to decide if a programming language is a language in the same sense. It should also explain to what extent we can consider that written English and oral English correspond to one and the same language and not to two different languages.

Obviously semantics is not designed to solve this sort of issues. Semantics refers just to one language at a time. It studies the meaning of the different components that can be produced within a language whichever their size. Depending on the categories of component we decide to adopt, we will define lexical semantics (word/term unit), sentence semantics (sentence), discourse semantics (discourse), textual semantics (text). But semantics does not explain what meaning is. It is concerned by the rules (combinations and requirements) that govern the combination, derivation, equivalence of the meaning of the components of the language we choose to consider.

3.2 Hjelmslev's semiotics

The Danish linguist Louis Hjelmslev (1899-1965), who is generally considered as the continuator of Saussure's work developed a theory of language, namely glossematics, which a classical introduction is provided by "Prolegomena to a theory of language" [14].

The contribution of Hjelmslev to semiotics, is first of all epistemological.

In [15] he summarized the principles of glossematics as follows:

- Recommend an analytical procedure as the only appropriate (also known as deductive, a term that has proven to be ambiguous, and consider the synthesis (...) as presupposing the analysis;
- Emphasize the *form*, which has been neglected in favor of the *substance*;
- Consider the linguistic *form* of the *content*, not just that of the *expression*; and accordingly draw the consequences of these principles,
- Consider that *language*, in the sense commonly adopted by linguists , as a special case of a semiotic system , that is to say a system consisting of different *plans* and within each level, a difference between *form* and *substance* [...] and situate language within the frameworks of a general semiotics (or semiology)

We must insist with Hjelmslev upon this double distinction:

- Form vs substance;
- Content (signified) vs expression (signifier);

Let us take a simple example, a dog.

In an Aristotelian semiotics, we should distinguish between the thing (the actual dog), the sign (the word *dog* in English, *chien* in French, سگ in persian, कुक्कुर in sankrit) and the interpretant (the **concept** of dog).

In Saussure and Hjelmslev's views, we should only distinguish between the *signified* of the word dog and the *signifier* (*dog* in English, *chien* in French, سگ in persian, कुक्कुर in sankrit) without any consideration to any actual physical dog. In other word we must limit our analysis to linguistics and nothing else.

In this sense Saussure and Hjelmslev insisted on the fact that we should consider the "linguistic value" (Saussure) or the "form" (Hjelmlsev) of the signified: the "linguistic value" of the signified of dog, is the *role* that it plays in language, on content plane, when dog is opposed to cat, cow, horse and so on.

Hjelmslev opposed to the *linguistic value* (or *form*) of a signified (or content), its substance (and appreciation level). An English man or woman even would have in view a domesticated animal trained for hunting or watching or may be, used as a companion animal. But other *semantic* definitions are possible quite different from the previous one. In Eskimo society the substance of the content dog is equivalent to working dog used as a sled dog. The Persian would define it as a sacred animal. Hindu people on the opposite would have a pejorative definition of it as a pariah (see [15]).

Concerning the distinction content (signified) vs expression (signifier), Hjelmslev insists on the fact that the relations that organize the elements (forms) on the plane of content are not conform even if the two planes possess equivalent organization. This last point means that there is no formal distinction between the two planes, each of them having an equivalent organization.

But this also implies that a true language cannot be reduced just to one plane! Consequently, from Hjelmslev's point of view, only the forms matter in order to define a language. The meaning of content, i.e. its *substance*, may change, as well as the substance of the expression, without changing the language! Even if this conclusion seems radical, this implies in particular that the meaning of a *word* as well as its evolution cannot be explained by linguistic reasons.

3.3 Multi-viewpoints semiotics

In order to account for the designing activity of complex systems such as space systems we proposed a semiotic methodology which is based upon Hjelmslev's concepts [16] [17].

In particular the above considerations allow us to define a viewpoint as the correspondence that gives a "linguistic value" to a substance:

- Viewpoint: *substance* → linguistic value.
- E.g. Viewpoint of Persian: *sacred animal* → dog (opposed to cat, cow, horse and so on)
- E.g. Viewpoint of Hindus: *pariah animal* → dog (opposed to cat, cow, horse and so on)

We need to distinguish between a viewpoint and a view. This is analogous to the distinction between the function f and a value $f(x)$ of this function for the element x.

- *e.g.* dog is a view for *sacred animal* in a *Persian's viewpoint*
- e.g. dog is a view for *pariah animal* in a *Hindu's viewpoint.*

In Hjelmslev's terminology, a view is the *manifestation* of the form in the substance: the substance is the *manifestant* and the form is the *manifestatum.*

The viewpoint can be defined with respect to the plane of content, but since the equivalence of the two planes this qualification is not essential (for a justification of that see [15])

We define the *confrontation of two viewpoints* by the semiotic function (or semiosis) between the two planes.

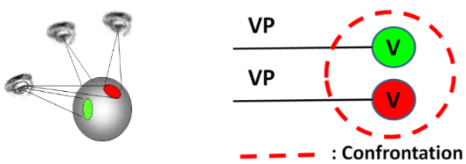

Figure 10. Confrontation of viewpoints.

Two features of that semiotics I am presenting need to be underlined :

F1: A view cannot exist apart from a confrontation of viewpoints (justification is to be found in [15])

F2: A viewpoint can only be analyzed within a confrontation of viewpoints. (This results from the definition of a viewpoint and from F1).

What is empirically observed is the interaction of viewpoints corresponding to the different stakeholders of a project involved in the design of an object. Several cases need to be considered:

Confrontation of viewpoints (VPs): the views produced by each VPs make sense to the others, but are not compatible.

Correlation of viewpoints: After a negotiation process, all VP produce mutually acceptable views.

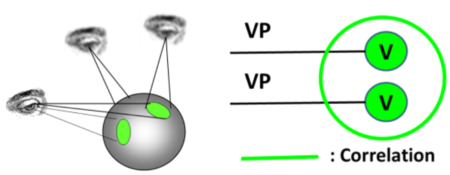

Figure 11. Correlation of viewpoints.

Two cases to be considered (empirically observed):

- All of the considered VPs can evolve during interaction,
- Only one Vp is considered. The interactions with the other VPs are non-evolving processes. These VPs are put between brackets.

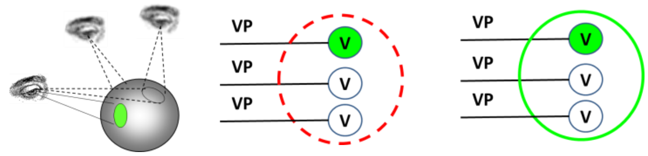

Figure 12. Viewpoints put between brackets.

The products of these cases can be used as definitions of *information*, *knowledge* and *data*. The justifications of these definitions as well as the introduction of this semiotic approach can be found for instance in [17].

Table 1. Types of views according to viewpoints interactions

	Confrontation of viewpoints	Correlation of viewpoints
All of the considered VPs can evolve during interaction	information	knowledge
Only one Vp is considered. These other VPs are put between brackets	Schematization Data	Modelling Data

Let us point out that *scientific data* correspond to *Modelling Data*, whereas *Schematization Data* correspond to the production of a code of ***expression*** (e.g. a *conceptual graph*) – and not of a content. (See [17])

4. DISCUSSION OF PREVIOUS RESULTS (CONTINUATION)

What we observe from the previous studies that we have presented in this paper, is that all the different computations that were performed produced different *views* of their objects (time-series of documents). Thanks to the *multi-viewpoints semiotics* we have just sketched, we understand that these different views were respectively produced by the *confrontation* of statistical, logical, taxonomic, linguistic viewpoints with the viewpoint of information technology (and sometime with the viewpoint of the producers of the document). However when inserted within the context of a space project these views remained in general just information. The explanation again is simple: these views do not correspond in general to the *correlation* of all the viewpoints that come under the purview of space technologies and operations. In other words semantic changes if observed could not in general signify more than anecdotal information.

What we have pointed out is that it is necessary to distinguish between the different planes (content and expression) that intervene in language production. However in order to account for knowledge production this distinction should be extended to viewpoints that play alternatively the role of plane of content or plane of expressions to each other. To be more specific, the substance of the expression (identified to a viewpoint) cannot be considered as innocuous with respect to any other viewpoints that play the role of a plane of content. All the viewpoints at stake deserve to be considered.

5. CONCLUSIONS

The contribution of the paper to the question of semantic changes is clearly epistemological.

In this work we presented the studies we conducted in order to bring a concrete solution to the issue first of long term space missions such as Rosetta, then to the question of data with respect to their usability by their original producers and by other members with different backgrounds. Even if the different approaches that were developed on these occasions produced high quality scientific results revealing what had changed in the case of Doris beacons for instance, they did not revealed which knowledge had changed among the teams. We did not present these results in details but just sketched them out. We preferred to insist on the importance the expression (the graphic language to express it naively) of the result that are offered to more or less arbitrary interpretations.

We proposed a semiotic framework in order to have a better insight of knowledge (and consequently of information and data). With this view we sketched out a multi-viewpoints semiotics which is based upon glossematics concepts. The necessity to take into account all the viewpoints at stake when considering the question of knowledge evolution points out its relation with cultural issues and the description of any semiotic system. In this respect, the Estonian semiotician Juri Lotman (1922 – 1993) insisted on the necessary interaction between semiotic systems in presence and between languages – that we interpret as viewpoints. "The fundamental questions relating to the description of any semiotic system are, firstly, its relation to the extra-system, to the world which lies beyond its borders and, secondly, its static and dynamic relations. The latter question could be formulated thus: how can a system develop and yet remain true to itself?" [18].

In order to bring an answer to that question, there are at least two conditions according to Lotman: (a) the necessity that more than one language (a minimum of two) is required in order to reflect a given reality; (b) the inevitable fact that the space of reality cannot be represented by a single language but only by an aggregate of languages. Juri Lotman added this warning: "The idea of the possibility for a single ideal language to serve as an optimal mechanism for the representation of reality is an illusion. A minimally functional structure requires the presence of at least two languages and their incapacity, each independently of the other, to embrace the world external to each of them. This incapacity is not a deficiency, but rather a condition of existence, as it dictates the necessity of the other (another person, another language, another culture)" [18].

6. REFERENCES

[1] CCSDS. 2012. Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-M-2 Page 3-6

[2] Wittek, P., Darányi, S., Kontopoulos, E., Moysiadis, T., and Kompatsiaris, I. 2015. Monitoring term drift based on semantic consistency in an evolving vector field. In International Joint Conference on Neural Networks (IJCNN), 2015, 1–8.

[3] Condamines, A., Galarreta, D., Perrussel, L., Rebeyrolle, J., Rothenburger, B., Viguier-Pla, S. 2003. Tools and methods for knowledge evolution measure in space project, Proceedings of IAF-IAA.6.2.7, Bremen, Germany, September 2003.

[4] Rothenburger, B. 2002, A Differential Approach for Knowledge Management, ECAI workshop on Machine Learning and Natural Language Processing for Ontology Engineering, Lyon

[5] Rajman, M. & Besançon, R.1998. *Text Mining – Knowledge Extraction from unstructured textual data*, Proceedings of 6th Conference of International Federation of Classification Societies (IFCS-98), Roma (Italy), July 1998, pp. 473-480

[6] Habert, B. & Zweigenbaum, P. (2002). Contextual acquisition of information categories: what has been done and what can be done automatically ? In B. Nevin (Ed.), Volume 2. Computability of language and computer applications. Amsterdam / Philadelphia: John Benjamins.

[7] Kleinberg, J. 1997. Authoritative Sources in a Hyperlinked Environment Proceedings of the ACM-SIAM Symposium on Discrete Algorithms, 1998, and as IBM Research Report RJ 10076, May 1997.

[8] Flake, G. W., Lawrence, S., Giles, C.L. 2000 Efficient Identification of Web Communities. In the proceedings of the 6th ACM SIGKDD International Conference on Knowledge discovery and data mining, Boston, Massachusetts, USA, ACM. pp. 150-160, 2000.

[9] Aristotle On Interpretation Translated by E. M. Edghill http://classics.mit.edu/Aristotle/interpretation.1.1.html

[10] Everaert-Desmedt, N. 2011. Peirce's Semiotics. In Hébert, L. (Ed.), *Signo* [online], Rimouski (Quebec), http://www.signosemio.com/peirce/semiotics.asp.

[11] Falk, J.S. 2004. Saussure and American linguistics. In Sanders, C. The Cambridge Companion to Saussure. Cambridge University Press. pp. 107-123.

[12] Norris, C. 2004. Saussure, linguistic theory and philosophy of science. In Sanders, C. The Cambridge Companion to Saussure. Cambridge University Press. pp. 219-239.

[13] Bouissac, P. 2004. Saussure's legacy in semiotics. In Sanders, C. The Cambridge Companion to Saussure. Cambridge University Press. pp. 240-260.

[14] Hjelmslev, L. 1961. Prolegomena to a theory of language. Madison: Univ. of Wisconsin Press.

[15] Hjelmslev, L. 1971. La stratification du langage. Essais Linguistiques. Les Editions de Minuit, Paris, 1971. pp.45-77

[16] Galarreta, D., 2010. A Semiotic Approach of contexts for Pervasive systems. 12th International Conference on Informatics and Semiotics in Organisations IFIP WG8.1 Working Conference University of Reading, UK. 19-21 July 2010.

[17] Galarreta, D. *Are things, objects? A semiotic contribution to the Web of Things*. Web of Things, People and Information Systems. 14th International Conference on Informatics and Semiotics in Organisations (ICISO 2013). IFIP WG8.1 Working Conference. March 25-27, 2013 Stockholm, Sweden

[18] Lotman J. 2009. *The explosion and the Culture.* Walter de Gruyter GmbH & Co. KG, D-10785 Berlin.