A Quali-Quantitative Narrative Analysis of the 2012 Fessenheim Nuclear Accident in the French Media

Dario Compagno

A Quali-Quantitative Narrative Analysis of the 2012 Fessenheim Nuclear Accident in the French Media

Dario Compagno[1]

[1]Université Sorbonne Nouvelle – dario.compagno@gmail.com

Abstract

Qualitative methods allow us to work on stories, while quantitative approaches often cannot. But can we work on a story with quali-quantitative instruments? Can we find a story in a corpus just by the use of diagrams, without reading any of its individual documents? We will analyze one series of events that occurred in France in 2012: the accident that took place at the Fessenheim reactor and the subsequent political decision to close the plant. The data have been gathered from the OT-Media database. We will make use of textual statistics and narrative models to understand the constitution and evolution of this series of events. Our aim is to show how the use of textual statistics permits to realize narrative analyses on multimedia corpora.

Keywords: Semiotics, Narrative Analysis, Data Analysis, Mediatization of Events, Multimedia Corpora

1. Quali-Quantitative Narrative Analysis

Semioticians and sociologists have performed many analyses of how the mass media represent events or series of events. Their goals include a segmentation or "anatomy" of a series of events, by identifying its components (Wien and Elmelund-Praestekaer 2009), or a comparison of how different media represent the same event or series of events (Rebillard 2012). A seminal socio-semiotic analysis of this kind was performed by Veron and his collaborators (Veron 1981). They studied how the French media reported the accident that took place in March 1979 at one of the reactors at the Three Mile Island nuclear plant. Veron took into consideration the three main media supports that existed at the time (radio, television and printed press), describing how the logics of narration in the media interacted with the (relatively scarce) information coming from the US. The kind of analysis performed was purely qualitative, based on content analysis, semiotics and sociological models. More recently Rebillard, Fackler and Marty (Rebillard et al. 2012) analyzed the accident that happened at the Fukushima Daiichi reactor in 2011, describing how the topics of discussion were distributed across the different media. Also this analysis required a manual coding of the data into categories. Would it be possible to improve such analyses by the use of textual statistics, and in particular by combining factorial correspondence analysis with topic modeling?

Our aim is to perform a new analysis of a similar event: the accident that happened in France at the Fessenheim reactor in September 2012. It is important to state that the disasters of Three Mile Island and Fukushima had much worse consequences: they have been rated a 5 and a 7 out of 7 on the International Nuclear Event Scale[1], while the Fessenheim accident (the

[1] http://en.wikipedia.org/wiki/International_Nuclear_Event_Scale.

last of a longer series dating back to 1990[2]) will most certainly be ranked at a much lower level. However, we are interested in the way the French media represented the event, independently of its objective magnitude. Our aim is to perform a quali-quantitative analysis that makes use both of statistical tools of investigation and of qualitative models of interpretation. This new kind of narrative analysis presents us with several differences from canonical methods: we will not consider the documents in our corpus as separate texts (each with its own narrative structure), nor we will read all of them before building a global interpretation. We want to *substitute analysis to reading*, and only interpret the diagrams we are capable of producing with data analysis[3].

This move significantly reduces the amount of work needed to produce a manual classification of the documents according to their content. Second, it enhances qualitative analysis with the use of statistical algorithms, in order to produce better justifications for the findings. Third, it has the potential to produce results that are impossible to obtain with purely qualitative analysis (Rastier 2011). Still, we aim to arrive at results of semiotic significance. By this we mean that we do not want to perform a quantitative analysis that ignores the meaning of the news itself, with the risk of "killing meaning by dissection", so to speak. Semiotics was born as the most qualitative of disciplines, given that its foundational principles of analysis prescribe exclusive concentration on the meaning communicated by the various expressive components of written texts or of other multimedia documents and discourage the consideration of the empirical conditions of production and interpretation (Hjelmslev 1969). According to this methodological stance, we want to obtain diagrams that can be interpreted as if they were the original documents themselves; we need graphs that are iconic, motivated representations of the documents (Eco 1976, Dondero and Fontanille 2012), in order to read the graphs as *new expressions for the same contents*[4].

However, in our analysis we will make use of some variables reflecting the empirical conditions of production for the documents (publication dates and media sources). We consider the methodology of discursive analysis developed in the sixties by Foucault (1969) as a point of departure to *interpret corpus variables as discursive variables*. For Foucault, discursive analysis is performed on documents and not directly on the empirical conditions of their production, such as the persons or institutions producing them, the dates and places in which they are produced or interpreted, or the facts or events to which the documents refer. Still, all these dimensions constitute discursive meaning and are therefore object of discursive analysis, as long as they can be accessed from the documents themselves. As a matter of fact, this analysis respects Foucault's indications, given that our variables simply label the documents in the corpus: we interrogate the variable levels only as classifications of words.

---

[2] http://fr.wikipedia.org/wiki/Centrale_nucl%C3%A9aire_de_Fessenheim.

[3] A first approach to quantitative narrative analysis has been developed by Franzosi (2010) and applied to news corpora (Sudhahar et al. 2011). Franzosi uses narrative models (agent-action-patient triplets) as a means *to code* the data; on the other hand, we use unsupervised textual statistics to produce graphs starting from narratively unstructured data, and then a narrative model *to interpret* the final results.

[4] Our aim is to use data analysis so as to transform the expressions in our corpus into new ones, while preserving the semiotic function, defined as a function from a domain called "expressive plane" to an image called "content plane" (Eco 1976). Therefore we aim to a regulated semiotic production obtaining diagrams that are related to their contents in a way that is comparable to the way the original expressions were related to their contents, to the point that we may say that the original and the new expressions are related to the same contents.

2. Data Gathering and Processing

Our corpus has been extracted from the OT-Media database[5]. The OT-Media database collects and indexes documents issued from several French media sources: TV channels, radio stations and Internet websites[6]. OT-Media documents are made of a textual component (copies of website contents, automatic transcriptions of radio news broadcastings, automatic transcriptions of TV news broadcastings extracted from teleprompters), images (screen-shots taken from TV programs, images extracted from web pages) and meta-data. The textual component of each document is analyzed by linguistic algorithms in order to extract a smaller number of significant words, that are collected as "noun phrase" meta-data[7]. Thanks to this data enrichment we are able to work on documents coming from different sources (news agency dispatches, television broadcasts, …). Each OT-Media document is indexed according to some variables including its publication date, the name of the source who published it (e.g., TF1, Le Monde, Mediapart), the kind of support to which the source belongs (TV, radio, web), the name of notable persons, places and institutions named in the document's textual component. We have chosen to constitute our corpus by extracting all the documents in the database that were published in September 2012 by some selected sources[8] and included the word "fessenheim" in their title. We obtained a corpus of 175 documents. Figure 1 shows the distribution of documents in our corpus across the media supports and the publication date variable levels.

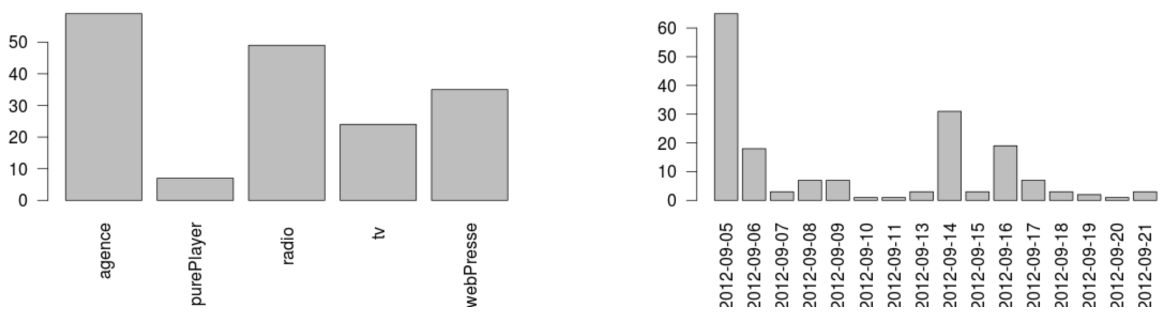

Figure 1: Distribution of Documents across the Media and Publication Date Variables

---

[5] See www.otmedia.fr. For details on the constitution of the OT-Media database, see Viaud et al. (2012).

[6] OT-Media's sources are 1 news agency (AFP), 8 radio stations (BFM, Europe 1, France Culture, France Info, Radio France International, RTL, RMC, France Classique), 11 tv channels (BFM, France 2, TF1, i>TELE, LCI, France 3, France 24, M6, Canal +, France 5, ARTE) and more than 1300 websites (among which 78 news websites and 324 blogs).

[7] The procedure of extraction used rules implemented in Linguistic Object Language (LOL). See (Ma et al. 2011) about the LOL language.

[8] The sources we have selected are the following: Press Agency AFP; BFM, Europe 1, France Info, France Inter, France Internationale, RCM, RTL (radio); ARTE, BFM, France 2, France 3, LCI, TF1 (television); La Croix, Le Figaro, Le Monde, Le Parisien, L'Express, Liberation (web press); Agoravox, Huffington Post French Edition, Mediapart, Rue 89, Slate (pure player). The OT-Media database was accessed in May 2013.

We have developed several R scripts (R Development Core Team 2012) to perform data processing and to interface the XML files extracted from the OT-Media database to the software T-Lab (Lancia 2011)[9]. A first R script makes a query to the OT-Media web interface, downloads the XML files of all the documents retrieved by the query, and builds a spreadsheet selecting some of the meta-data. The meta-data we have selected are: the noun phrases, the publication dates, the sources names (TF1, Radio France International, …), the media type (web press, web radio, blog, ...) and the media support (TV, radio, …) associated with each document. A second R script recodes the media support and media type variables together. At the end of the process each document is associated to a single "media" variable, with the following levels: Agency, TV, Radio, Pure Player, Web Press. A third R script converts the spreadsheet into a text file readable by T-Lab. In our case, T-Lab analyzes the strings of noun phrases associated with each OT-Media document. A fourth R script cleans the text file, by deleting all the occurrences of a list of noun phrases that we have decided to remove from analysis. This selection has been made after several trials with testing corpora extracted from OT-Media. The deleted noun phrases do not refer to the specific content of the document they are associated with, but more generally to the writing procedures of the media source from which they are extracted; therefore, for our ends, they were only a source of noise[10].

Finally, the corpus was analyzed with T-Lab. The are two main processing operations performed by T-Lab when importing a corpus: a segmentation of all the documents (strings of noun phrases, in our case) in shorter segments of comparable length ("elementary contexts", EC from now on) and a lemmatization of all the words in the corpus; a selection of lemmas (we will call this selection "keywords" from now on) to be used for the analyses (Lancia 2011, pp. 198, 204). We have chosen to produce the longest possible elementary contexts, so that in the majority of cases a single EC coincides with all the noun phrases associated with a single document[11].

3. Data Analysis

Narrative analysis looks for patterns and regularities in those cultural objects usually called *stories* (*récits, racconti*) regardless of their medium: stories can be found in books, movies, songs. Narrative analysis can also be performed on other kinds of cultural objects, as long as these cultural objects make sense in the same way stories do. We are going to model the series

---

[9] All the R scripts, T-Lab reports and data (raw and processed spreadsheets, corpus txt files) are available for consultation by contacting the author.

[10] This is the full list of removed terms: "documentaires reportages images privés montages caméras factuelles alternances français vidéos transcriptions anglais intelligences plateaux sources citations interviews retours commentaires suites déclarations audio chaînes posés a-t impressions textes articles auteurs files internet journals médias noms photos presse sites télévisions".

[11] We have chosen to work with approximately 1000 keywords selected using a TF-IDF method among all the noun phrases. We have chosen paragraphs as elementary contexts; this option consists in segmenting the original strings of noun phrases into EC separated by marked punctuation (".", "?", "!") each with 2000 words of maximum length. In our corpus this means that every document corresponds to one EC, unless the string of noun phrases associated with a document is longer than 2000 words; in that case the original document will correspond to more than one EC. We have kept the default values of all other T-Lab import options.

of events reported by the media as a story “told” in a corpus of heterogeneous documents. Our analysis will start by identifying a few periods in which we can segment the series of events. Then we will look for the main themes and trends shaping the series. Finally we will combine the two approaches and give an interpretation of the series of events according to a narrative model.

3.1. Identifying Key Dates and Periods

First, we characterize the dates of publication in our corpus, according to the keywords used in the documents published on each date. To obtain this, we use FCA on the Publication Date variable, producing a system of representation for the variable levels (Lebart et al. 1998). What we wish to obtain is proof of a semantic evolution in time, that is, we expect to find a systematic temporal change in the way the media speak about the series of events (Figure 2).

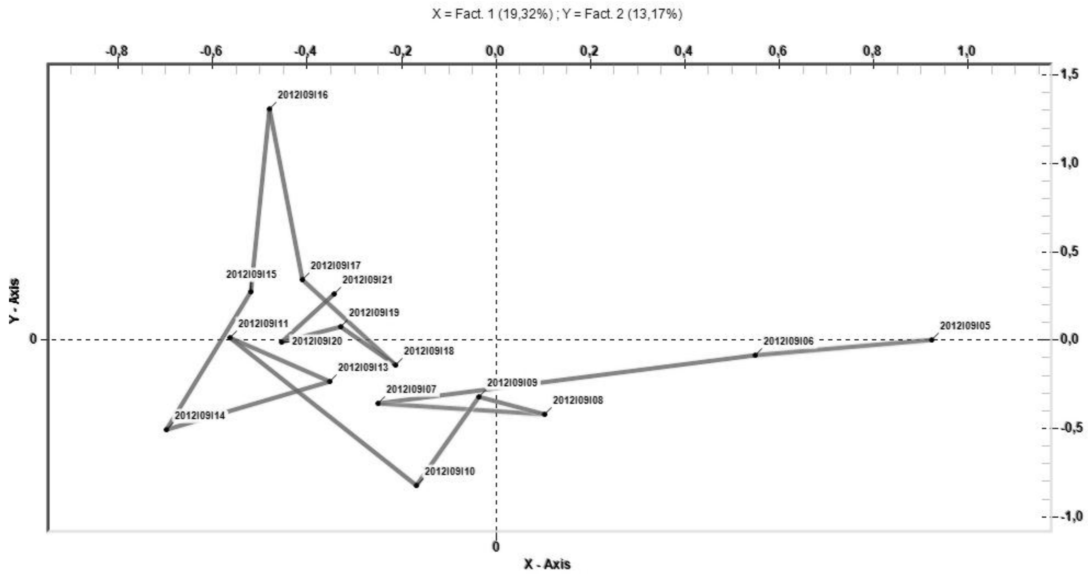

Figure 2: FCA Positioning of the Publication Date Variable Levels

We can observe a clear evolution in the language used by the media to represent the series of events. The first salient point is the beginning of the line (September 5th), on the extreme right-hand side of the plot. This is the start of the event, the “unexpected fact” (as Ch.S. Peirce would say) that is responsible for attracting the attention of the media. The second salient point is September 14th, on the extreme left of the diagram: after that date, the line stops evolving leftwards and begins to climb up. The third salient point is on the extreme top of the diagram (September 16th), where the line inverts its direction. We should also remark the point closest to the bottom of the chart, and more precisely the segment 8th-10th September[12]. FCA manages to find the factors that better differentiate the variable levels by

[12] If we read the absolute and relative contributions of these points to the second factor (Lebart et al. 1998, p. 58), we notice that the 8th and 9th of September are more important than the 10th:

	8th	9th	10th
Absolute Contributions	0.0580	0.1107	0.0070
Relative Contributions	0.0407	0.0056	0.0031

which a corpus is categorized, explaining at best the variance of the data. This means that in an FCA diagram the points farthest from the center are those better differentiated by the axes used to produce the diagram. Therefore, in our case the 5th, the 8th-10th, the 14th and the 16th of September are the dates in which a relatively well-differentiated language was used. Let us recall the main events that happened on the four dates highlighted by FCA (Figure 3).

Dates (September)	Main Events Happening
5th	Accident at the Fessenheim plant.
8th	CFDT accuses the French media of producing a defamatory campaign. Minister Batho announces that the plant will be closed soon.
14th	Hollande officially confirms the closure of the plant.
16th	French power group EDF is said to have asked 2 billion € as compensation for closing the plant.

Figure 3: Main Events Regarding the Fessenheim Plant Happened in September 2012

After having observed a chronological evolution in the language used by the media, and having identified a few key dates, we would like to produce a segmentation of the series of events in a small number of significant periods. These periods should separate the key dates and, taken together, they should also cover the entire duration of the series of events. To this end we calculate a few clusters of keywords that could correspond to the main narrative phases in which the series of event is structured. We produce a (hierarchical clustering) classification starting from the system of representation produced by our precedent FCA on the Publication Date variable: as such, the distance between the keywords is still given by a calculation that takes into account only the recurrence of these keywords in different dates. The clusters are therefore “made of time”, so to speak; they are intrinsically temporal. As for every clustering classification, their existence is also purely differential: they do not simply “exist” as blocks of meaning, the one independent from the other; on the contrary, they have to be considered as a system in order to give a shape to the variance in our corpus. This is exactly what we need to perform a narrative structural analysis: four phases that have a meaning - and even an existence - only if considered as parts of a whole.

In figure 4 we can see the positioning of the 4 clusters produced by our analysis. The clusters are numerated in reverse chronological order: the keywords within the first and largest cluster (in red) are related mostly to the documents published during the last dates of the series; those within the second cluster (in pink) to the dates 10th-14th of September; those within the third cluster (the smaller one in cyan) to the dates 7th-9th; and the keywords within the fourth cluster (the larger one in cyan) are related to the documents published on the very first dates of the series. We have found what will become the four main phases in our narrative modeling of the series of events. We have a starting point (September 5th), a struggle (8th), a main turn (14th), and a conclusion (16th). We have produced 4 clusters of keywords according to the dates on which they were published. Let us see more closely how each of these clusters is constituted. Figure 5 plots a spline diagram of the proportional distribution of each cluster across the Publication Date variable levels.

We see that each cluster is related to multiple dates, and also that the clusters overlap. Still, we notice a preferential association of each cluster with a single date and the period surrounding it, a sort of specific “sphere of action” (as V. Propp would say) for each cluster. Cluster 4 is strongly related to September 5th and 6th, but its keywords continue to be used all along the entire series of events (it is likely that further discussions continued to mention the initial accident). Cluster 3 peaks on September 8th, and then decays very quickly. Cluster 2 is more relevant between the 10th and the 14th of September. Cluster 1 grows in importance with time and has its highest peak on September 16th. We are plotting a proportional distribution, meaning that we are directly comparing dates in which many documents were published (as the 14th and 16th of September) to dates for which we have only one or two documents (10th and 11th). This permits us to stretch the segmentation all along the entire series of events.

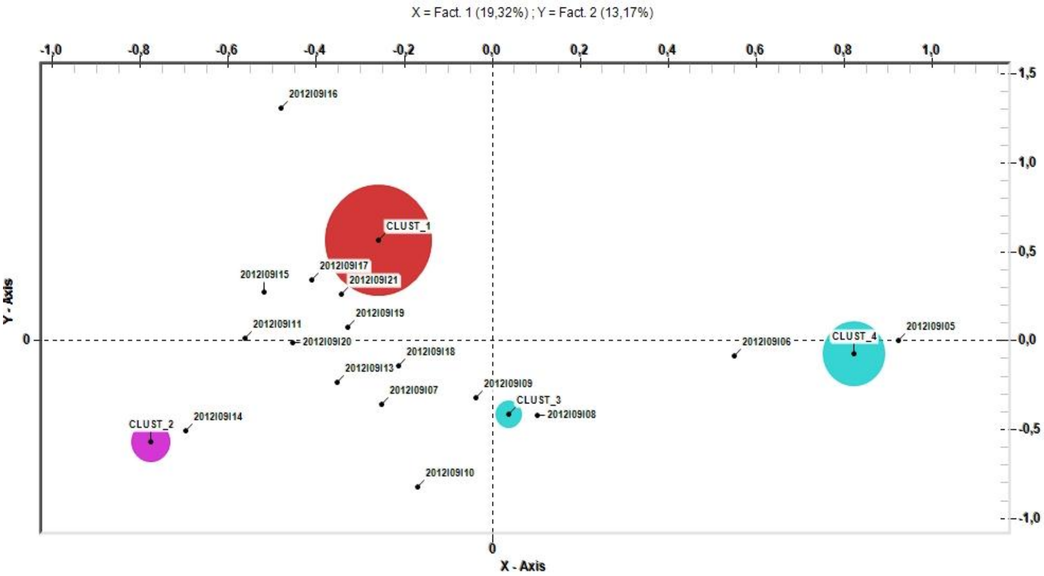

Figure 4: Clusters of Keywords after FCA on the Publication Date Variable

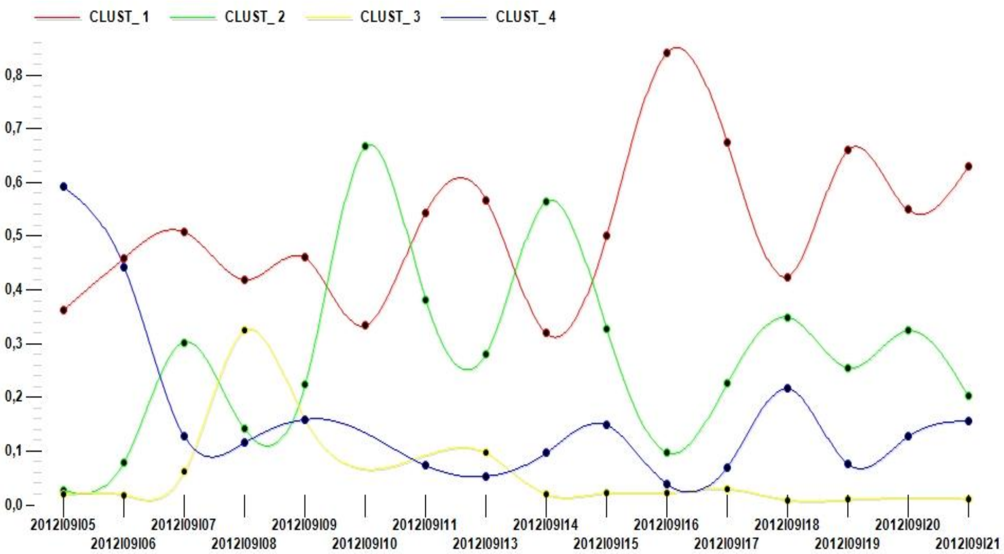

Figure 5: Proportional Distribution of the Clusters of Keywords across the Publication Date Levels

According to the diagrams we have produced so far, it seems that we have an “interesting narrative”, capable of retaining the attention of the audience around a core meaning. It seems to be true that the media tell stories. Media documents make sense when they are taken together within a structural unity that gives a role to each. Still, our ambition is to say something more than generically attributing four broad narrative phases to a series of events. We have the syntagmatic parsing of a tale, and now we would like to go more into details about what has been told in this tale.

3.2. Thematic Classification of Elementary Contexts

The “Modeling of emerging themes” function in T-Lab can help us to obtain a thematic classification of the elementary contexts (Lancia 2012, p. 107). This function uses Latent Dirichlet Allocation (Blei et al. 2003) and Gibbs Sampling to calculate a number of clusters (topics) chosen by the analyst. Figure 6 reports the percentage of EC included in each of the 10 clusters we identified and a brief description that highlights the main events associated with each of them (the cluster names are those automatically generated by T-Lab). Each theme may have been discussed for several days, and some of them overlap.

Cluster ID	% of EC	Description
CAMPAGNE	3.09%	La Conféderation Française Démocratique du Travail accuses the media of having produced a defamatory campaign against the Fessenheim plant.
CENTRALE	23.75%	Information, reactions and comments regarding the Fessenheim plant, its managers and employers.
CONFERENCE	3.33%	During the French national environmental conference, President Hollande declares that the Fessenheim plant will be closed in 2016.
DEGAGEMENT	12.11%	Technical descriptions of the accident happened on the 5th of September 2012.
EMPLOI	10.93%	Comments about the consequences in terms of employments (lost or created) if the Fessenheim plant will be closed.
FERMETURE	11.88%	About the closure of the Fessenheim plant.
GAZ	9.98%	About other forms of energy, and in particular about shale gas.
INCIDENT	8.08%	Comments on the accident happened on the 5th of September 2012.
MINISTRE	7.6%	Political comments (notably of the Ministers Batho and Sapin) about the consequences of the closure.
VIE	9.26%	French power group EDF is said to have asked 2 billion € as compensation for closing the plant.

Figure 6: Description of the Thematic Clusters of EC

We can now use FCA to plot the relative positioning of the clusters on a factorial system of representation. This may help for understanding and interpreting the clusters. Figure 7 shows position and magnitude of each cluster. The colors identify the quadrant: the red clusters (NW quadrant) are more related to the accident that happened on the 5th of September; the blue clusters (NE quadrant) are more related to the discussion about environmental politics in France; the cyan clusters (SE quadrant) are more related to the age of the plant and to the subsequent costs and risks. Figure 8 shows instead the keywords that contributed the most to the constitution of the axes, and plots the position of the clusters in relation to them[13].

It is interesting to see how the clusters are distributed across the media supports. Figure 9 plots this distribution. We can observe a few differences that could lead to the formulation of hypotheses about media coverage and pluralism. However, we should remark that our corpus is relatively small, and that not all the variable levels have the same number of documents associated to them (in particular, the Pure player level is significantly smaller than the other ones, while the Agency level is significantly larger). If we use the AFP press agency dispatches as a reference, it seems that the radio stations covered the topics associated to the

[13] The interpretation of this diagram needs special attention. Lebart, Salem and Berry write: “As a matter of fact, it is not possible to interpret these cross-proximities between a row-point and a column-point, because these two points do not come from the same initial space. Nevertheless, it is possible to interpret the position of a single row-point with respect to the set of column-points or of a single column-point with respect to the set of row-points.” (Lebart et al. 1998, p. 54)

"Centrale" cluster much more than the other media. We have interpreted this cluster as including generic discussions about the plant and its employees. Television over emphasized the "Incident" cluster, that is, non-technical comments about the accident of the 5th September. The rumor about EDF asking for an economical compensation ("Vie" cluster) seems to have been under emphasized by television, radio and pure players. On-line press seems to follow the agenda promoted by AFP very closely. According to our data, the technical discussion about the accident ("Dégagement" cluster) had a predominant role for the pure players agenda.

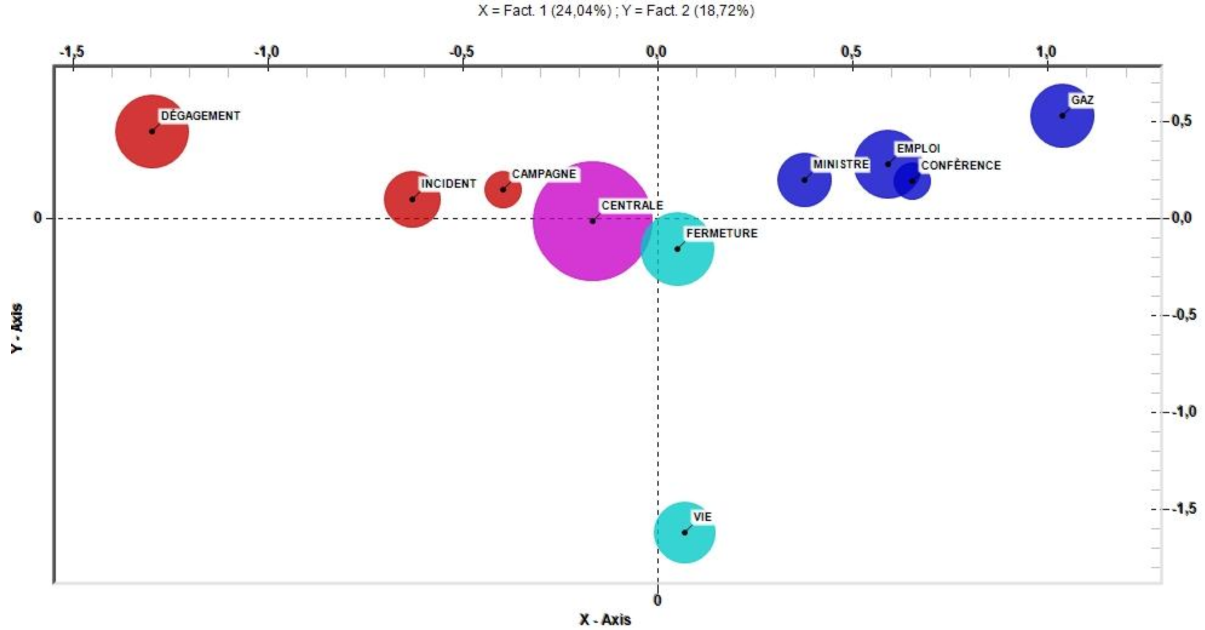

Figure 7: FCA Positioning of the Thematic Clusters of EC

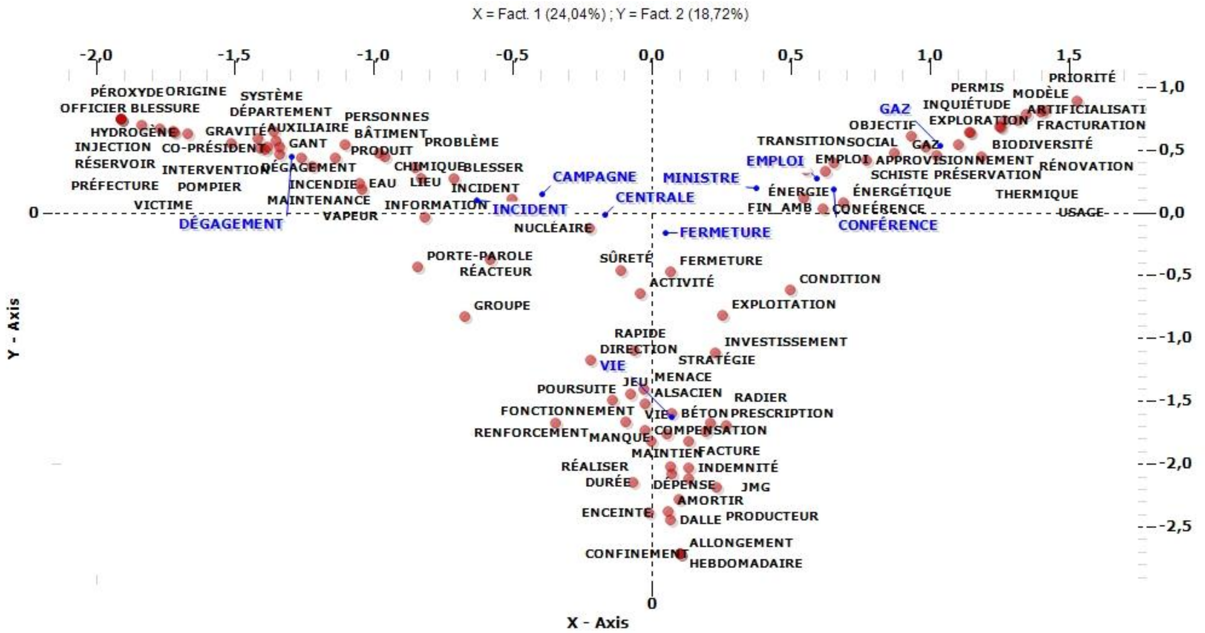

Figure 8: Simultaneous Plotting of the Thematic Clusters of EC and of the Keywords

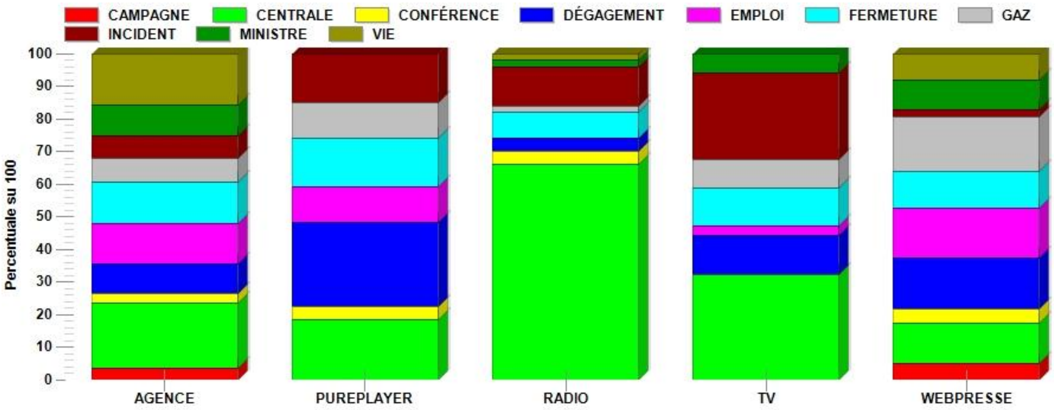

Figure 9: Distribution of the Thematic Clusters of EC across the Media Supports

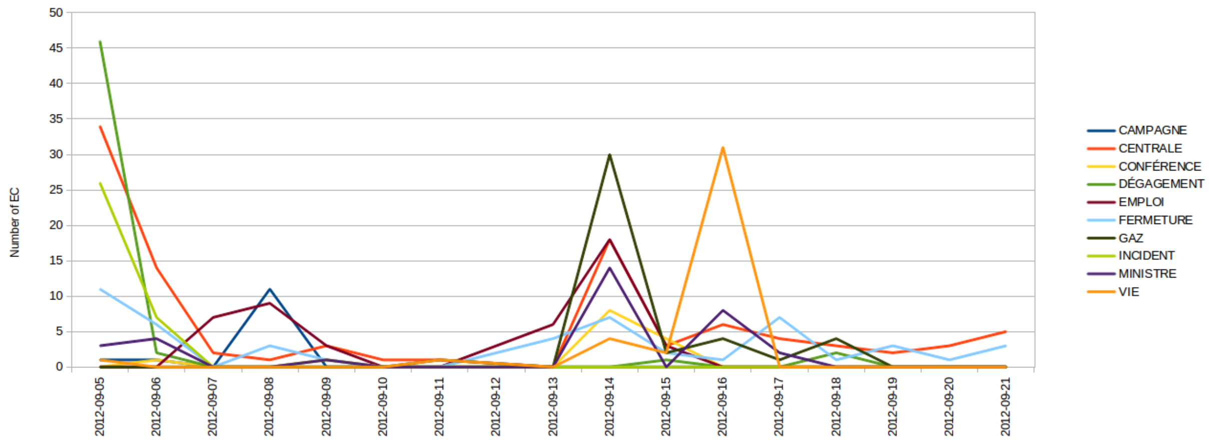

Figure 10: Distribution of the Thematic Clusters of EC across the Publication Date Levels

From our research perspective, however, it is even more interesting to plot the distribution of thematic clusters against time: this will help us to obtain a narrative interpretation of how the media turned the series of events into a story. In 3.1 we produced a syntagmatic segmentation of the series of events into four periods. Now we want to "populate" these abstract narrative phases: we would like to know more about what happened in each of them. In other words, we want to understand what paradigmatic options were chosen at every given syntagmatic phase. In order to do so, we plot the 10 clusters against the Publication Date variable levels (Figure 10).

We can now characterize thematically the documents published on each date. For example, the very first days of the event (September 5th and 6th) are characterized by the themes "Dégagement", "Incident" and "Centrale": the media were reporting about the accident and giving the first general updates on the plant. Starting from September 7th, the media began to talk about the possibility of closing the plant ("Emploi"). This theme has a first peak on September 8th (when Minister Batho declares that the plant will be closed) and a second one on September 14th (when President Hollande confirms this decision). On the 16th the EDF power group is said to have asked for an economical compensation to the French State ("Vie"). After the 17th the media continue the discussion more in general terms ("Fermeture", "Centrale").

4. Narrative Interpretation

Now we have all the information we need to model the series of events related to the accident occurring at the Fessenheim plant on September 5th. We have both a paradigm of alternative themes and a syntagmatic segmentation of the series of events into phases. To model the series as a story we will make reference to Greimas' narrative model (Greimas 1987).

According to Greimas, in any story there is a hero pursuing an objective, and the hero has to complete three tests in order to succeed: a first minor one (qualifying test) in which s/he proves his or her ability thanks to the aid of a helper, a second one (decisive test) in which s/he engages and defeats the principal opponent and a third one (glorifying test) in which the hero manages to be recognized and acclaimed as triumphant. These tests are the key moments of three abstract narrative phases (competence, performance, sanction) in which any story is articulated. There is also a fourth phase (manipulation) at the beginning of every story, in which the hero chooses to undertake his or her quest. There is no test associated with manipulation, but a story is often started by an "inciting incident" (as G. Freytag would say) propelling the hero into action.

What is useful about Greimas' model, is that after identifying a few major phases it allows us to focus on a very limited number of narrative roles (or *actants*). A hero, an opponent, a helper and an objective are all that is needed to have a story. These abstract roles are instantiated in some more concrete figures (characters, places, events) and if we want to understand the core meaning of a story we need to extract its structure of narrative roles.

Hero's Tests	**Thematic Clusters**	**Dates**	**Temporal Clusters**	**Narrative Phases**
Inciting incident	Dégagement, Incident, Centrale	5th	4	Manipulation
		6th		
		7th	3	Competence
Qualifying	Campagne	8th		
		9th		
		10th-13th	2	Performance
Decisive	Gaz, Emploi, Ministre, Fermeture, Conference	14th		
		15th		
Glorifying	Vie	16th	1	Sanction
		17th-21st		

Figure 11: Narrative Interpretation

Figure 11 summarizes the information we have obtained from our data, and interprets them according to Greimas' tests and narrative phases. Each date is associated with a temporal cluster and a narrative phase on the right-hand side, and with the global peak of a thematic cluster and a hero's test on the left-hand side. Data analysis and textual statistics have helped us to identify a narrative structure in a corpus of documents. Greimas' model also allows us to "calculate" the narrative roles of our story. Let us see what story we managed to find in our corpus. It should be clear that the narrative roles we found (determining who is "good" or “bad” in the story) are determined by our analysis and not by subjective speculation. The hero of our story is President Hollande. He is called to action by the accident of September 5th: public opinion demands a resolution after this dangerous event. A first minor dispute between the hero's side and an antagonist happens on September 8th: CFDT accuses the media of building up a defamatory campaign about an issue that does not exist. On the same date, Hollande's helper (Minister Batho) declares that the plant will be closed and that the media are not exaggerating the news. This declaration anticipates Hollandes announcement on September 14th, which constitutes the victory of the hero against his main opponent: EDF. The opponent still tries to “hit back”, on September 16th, when a rumor menaces the request for an economic compensation[14]. However this rumor dies out soon, and Hollande can be recognized as the winner of his quest.

References

Blei D.M., Ng A.Y. and Jordan M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993-1022.

Dondero M.G. and Fontanille J. (2012). *Des images à problèmes. Le sens du visuel à l'épreuve de l'image scientifique*. PULIM.

Eco U. (1976). *A Theory of Semiotics*. Indiana University Press.

Foucault M. (1969). *The Archaeology of Knowledge*. Routledge.

Franzosi R. (2010). *Quantitative Narrative Analysis*. Sage.

Greimas A.J. (1987). *On Meaning*. Minnesota UP.

Hjelmslev L.T. (1969). *Prolegomena to a Theory of Language*. Winsconsin University Press.

Lancia F. (2012). *T-Lab 8.1 User's Manual*. Available at: http://www.mytlab.com/Manual_en.zip.

Lebart L., Salem A. and Berry L. (1998). *Exploring Textual Data*. Kluwer.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: http://www.R-project.org.

Rastier F. (2011). *La Mesure et le Grain. Sémantique de corpus*. Champion.

Rebillard F., editor (2012). *Internet et pluralisme de l'information*. *Réseaux*, 176.

Rebillard F., Fackler D. and Marty E. (2012). L'offre d'informations est-elle plus diversifiée sur le web qu'à la télévision?, *Réseaux*, 176: 141-172.

Sudhahar S., Franzosi R. and Cristianini N. (2011). Automating Quantitative Narrative Analysis of News Data. In Diethe T., Balzacar J., Shawe-Taylor J and Tirnauca C., editors, *JMLR: Workshop and Conference Proceedings*, 17: 63-71.

Veron E. (1981). *Construire l'événement: les médias et l'accident de Three Miles Island*. Minuit.

Viaud M.-L., Rebillard F., Thièvre J. and Castex L. (2012). OT-Media Corpus: périmètre et capture. Technical Report, Agence Nationale de la Recherche (ANR).

Wien C. and Elmelund-Praestekaer C. (2009). An Anatomy of Media Hypes. Developing a Model for the Dynamics and Structure of Intense Media Coverage of Single Issues. *European Journal of Communication*, 24(2): 183-201.

Ma J., Mounier M., Blancafort H., Couto J. and Loupy C. de (2011). LOL: Langage objet dédié à la programmation linguistique. Traitement automatique des langues naturelles. In *Proceedings of TALN 2011*. Available at: http://www.lirmm.fr/TALN2011/DEMOS/DEMO_Ma_Syllabs.pdf.

[14] From a semiotic standpoint, it is inessential to know whether EDF really asked for a compensation. This rumor was part of the series of events, and played a role in the constitution of the story.