

# INTERNET EMOTIONAL DISCOURSE ANALYSIS AND INNOVATION – A MIX SEMIOTIC AND TEXT-MINING METHOD

Jérôme Guibourgé<sup>1</sup> and Philippe Bonny<sup>2</sup>

## ABSTRACT

Blogs, micro-blogs, discussion forums, social networks, wikis, and so on, constitute a rich, huge, recent and varied source of information for innovation. People on the Internet express themselves freely (“open discourse”) and through their posts and comments reveal emotions in their daily experience with products or services. This article presents a method which combines Semiotic and Text-Mining approaches aiming at extracting interesting avenues for innovation from these emotional elements. As meaning is neither necessarily nor directly given in these posts and comments, Semiotics is helpful, because it seeks to understand how and what meaning is generated by an utterance. As emotion is quite tricky to identify and to analyze on a large scale, Text-Mining is helpful too, because it gives robust means for mainly automatizing these points. An experimental application of this method is shown through a concrete case.

## INTRODUCTION

Discourses on Internet, in blogs, micro-blogs, forums, social networks or wikis, open up an access on product final user experiences in their daily life. Contrarily with other means for conducting a product market research, Internet allows an un-steered approach (no interaction of the observer with the user), rich (quite all the subjects are covered), immediate (internet access) and continue (« blogging every day »).

Innovation can then be filled out by the emotional relation that users experience in the “true life” with products or services designed by a company. One often says « customers invent the life going with new launched products or services ». It is precisely in this constant inventiveness revealed in blogs, using powerful text-mining tools, where we can find some gold nuggets for innovation, in the heart of the customer intimacy.

Upstream, discerning company values and specifying possible related strategic framework for innovations allow to better manage the coherence of the whole discourses of the mark (of which products or services form part), their general cohesion and congruence with the targeted recipients.

Downstream, semiotic analysis (or meaning generation analysis), aims at translating repeated meaning effects for design project researches.

How to capture and analyze the emotional discourses related to products or services? How to extract potential leads for innovation and how to build corresponding innovation strategies? This is all the purpose of this paper.

Firstly, an overall description of our mix semiotic x text-mining method for innovation is shown. Secondly, an application to a concrete pilot case dealing with Car Baby-Seat innovation is described. Finally, as a conclusion, some general findings regarding this method are proposed.

## METHOD

### General scheme

The general diagram shown bellow (Figure 1) describes our method. It combines semiotics and text-mining in order to design a strategy that allows brands to use better its values, improve the

---

<sup>1</sup> Consultant in planning strategic, semiotician, PhD in semiotics he gives lectures in ESC and universities.

<sup>2</sup> Co\_founder and CEO of the company Inevidence, a consulting company offering a range of studies, consultation and training on advanced information processing, based in Paris –France, [www.inevidence.com](http://www.inevidence.com)

management of its sensitive registers, and guide design product innovation.

The main components of this diagram will be detailed hereafter.

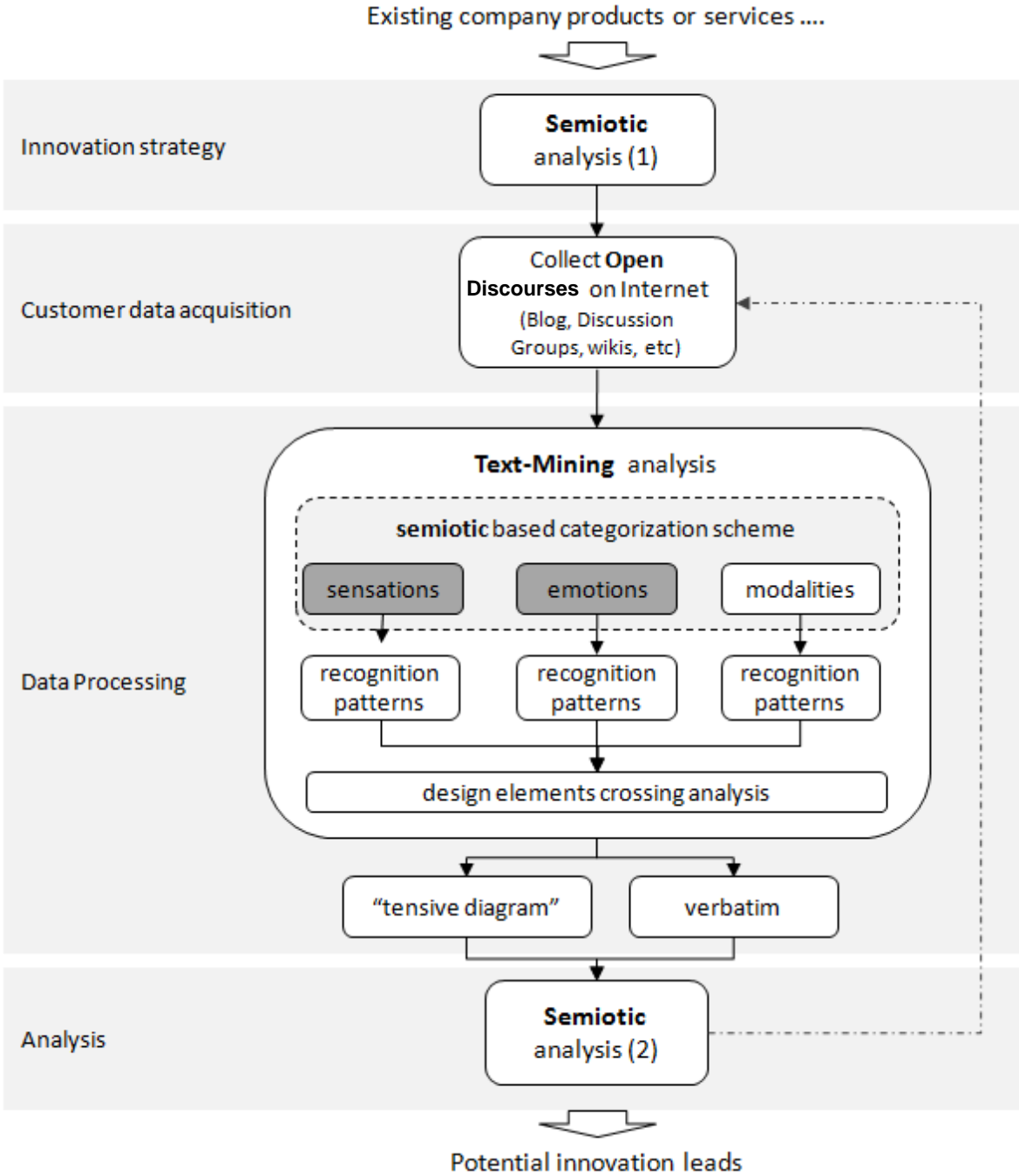


Figure 1. General method diagram

## Semiotic analysis (1): values, strategies and innovation

Semiotics regularly participates in qualitative studies and advertising, but it is also often used retrospectively. Here, we present a semiotic used upstream that allows to develop an innovation product strategy based on company values. Lotman (1998 Lotman) showed that a strategy which intervenes in an already established cultural or economic sphere obeys to a general principle. This one is a dynamics of cultural transformations in the semiosphere (which is a configuration of facts, shapes, figures, signs, symbols, etc.). New or foreign figures (in a large acception) are integrated in the initial culture but change relevant differences and values of this culture. The cultural dynamics is not necessarily strategic but any strategy depends on this dynamics which links an affective intensity (valorization of cultural transformations) with the expansion of cultural figures and facts (number of it, reproduction, translation...). From this dynamics emerge four positions, four developments that we can relate to products:

- (1) **The “brightness of the strange”**: making the utmost of the foreign contribution (innovative products that break with established uses;
- (2) **The “familiarization”**: putting value on the familiarization of new products that have been accepted and assimilated;
- (3) **The “exclusion of all singularities”**: putting value on the exclusion of all singularities that evoke a foreign origin; the product is then categorized as exotic, ethnic...
- (4) **The “universal deployment”**: this also makes the foreign contribution universal (see point 1); this value is, as it were, the sum of part of points one and two – the product offers a new norm.

Strategists encounter these very general tendencies (constraints and paths) and must therefore come up with product strategies (range, trademark) that reflect this. Fontanille (2006 Fontanille) explains that: “The strategic choice, therefore, consists of several degrees of complexity: (1) the simplest is to choose the general position, dictated by the main trends of the moment; (2) a little less simple is the choice of the least position occupied by the competition; (3) a little more difficult: to know better than the others, whatever the targeted position may be, how to resist the deterioration of this position...; (4) the most complicated, and perhaps the most likely impregnable position, consists in anticipating, with a long-term calculation, the strategy’s axiological future.”

In short, even though this article deals with innovation related to emotions, the strategy that takes into account this general principle, means to choose between (1) to resist or to defend an “island” (2) to anticipate or to plan all the phases from the initial strangeness to the end point of the universality of a new product; (3) to play with remanence or to count on the product identity and originality and (4) to adapt continuously the characteristics of the product (in depth monitoring).

That is a broad brush sketch of the axiologies linked to innovation and strategy; we will propose in the APPLICATION section to look at this in more details. In order to observe the values that are in use in our application case dealing with a brand designing and producing products. But before that, let us recapitulate some useful elements underlying our method (to go further about linking marketing and semiotics see 1999 Fraenkel & Legris-Desportes - 2003; Ceriani):

- It’s common knowledge to say Design is not only a question of aesthetics but also of usages; usages presuppose a system of values for products. If we use a product, it’s because we particularly valorize it. The product must represent a value. Between two identical products, in terms of functions, users valorize the one that indeed represent the value. For example, ABS brake system is the same for all cars but for BMW it is “performance” and for Volvo it is “security” (1998; Kapferer).
- Culture determines values, which participate in the identity of the brand and generate the brand’s meaning, and give the product meaning; it also means that values and innovation could be linked and use as inputs for strategy (1990 Floch).
- Values are functioning on an axiological principle that is to say by opposition and reciprocal presupposition (1990 Floch).
- **The semiotic square** is a diagram showing this axiological value system and also the possible trajectories from one value to another.

As, you will on Figure 2 and in the APPLICATION section, we can use for example the Floch square (1990 Floch).

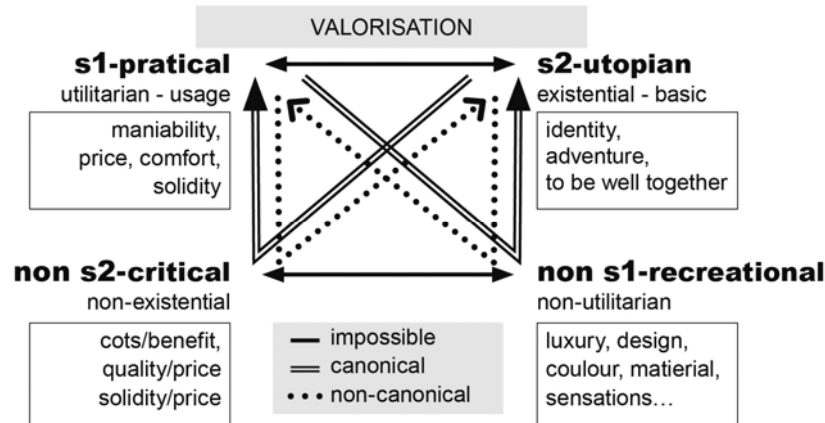


Figure 2. Semiotic square (Floch square)

Although there is a set of potential possibilities and characteristics of positions and trajectories, the value that we are interested in here is the bottom right (recreational or emotional value). We will give some rules of this diagram whose first is that you can't valorize your product both on practical and emotional values. This last is the negation of the practical value. It doesn't mean that a design product based on a sensitive approach is not practical but that its value lies primarily in sensations it gives. For example, to stay on car market, if you are an urban driver of Ferrari F430, Lamborghini Gallardo or a Porche Cayenne it means that you don't bother about practical aspects (speed limit, petrol consumption, size to park...) even if the car is still functional (wheels, engine, seat...). The second rule is that Meaning is generated by differences and all relations (arrows) on this diagram are combinations of these differences. The third rule is that Values are by what we promote our acts but are rarely articulated or directly accessibles; if needed, a survey must be launched to understand how they are expressed. In terms of marketing investigations, generally, Brand uses traditional methods, like focus group for example, but text-mining methods based on Internet open discourses, as proposed here, might be more appropriated for innovation early phases. As such analysis can be launched (and re-launched) rapidly in order to refine the analysis and the related potential innovation leads.

### Open Discourses collection on Internet

Open discourses on Internet, as said, give an un-steered vision of daily user experiences with products or services. Blogs especially, better than Internet groups, give to us some indeed 'slices of life' with sufficient semantic and linguistic material for analysis.

In order to preserve this objective view and construct a « neutral » corpus for analysis, it is necessary to collect a set of blog posts with the use of keywords or expressions related to the concerned product only. It wouldn't be wisely to cross with other keywords for example related to design elements. The acquired corpus and the potential enclosed leads for innovation in that way better reflect reality.

We use classic downloading tools (spider) on GoogleBlog Directory or Technorati results.

### Text-Mining Analysis

As described in Figure 1, Text-Mining analysis is divide in three main parts : firstly, on a semiotic base, define which and how coding categorization must be realized; secondly, build related pattern recognition allowing an automatic coding of relevant discourses dealing with emotions or sensations or modalities; thirdly, launch the generation of "tensive-diagram" linked with design components and verbatim.

### ***Semiotic based categorization scheme***

Semiotics seeks to understand how and what meaning is generated by an utterance (in the broad sense of the term, that is texts, images, objects, sites...) but does not search to understand the psychology or sociology of the target. Coupled with the powerful tools of text-mining it provides an effective method for enriching design hypotheses. The first task is to elaborate a matrix analysis to filter precisely the corpus: feelings, emotions, moods, passions, sensations, senses, sensitivity, sensibility... and so on. In order for designers to directly profit from this process: we not only identify these affective registers in the corpus, but we link them with the categories that have been created by the design research project. The latter depends in part on the object which the designers are working on. We have decided to reduce all terms in three categories: SENSATIONS, EMOTIONS and MODALITIES. The first one will give us direct links with material, forms, colors, etc. The second one will give us general valuation (to like or dislike, to feel good with...) on usage, conformation, contrast and all global perception of the object. Those two categories, SENSATIONS and EMOTIONS, are intensities (valuation) applied on the object functions and characteristics; these determine also the design product operationality. The last one, MODALITIES, is a little different because it allows a better understanding of the modalities of the subject: does he have to do something, can he, does he want to do or even what skills are required?

### ***Pattern recognition***

Emotional discourse automatic coding is carried out by the use of fine-tuned patterns (semantic and psycho-linguistic based) related to sensations and emotions (see Feldman et al 2007 for an overview on Text Mining; see Maurel et al 2006, Godbole et al 2007, Wilson et al 2009 et Prabocoo et al 2009, for recent publications dealing with the extensive area of research on Sentiment Analysis). Due to discourse strong variability with context, a manual tuning of these patterns is necessary to meet specific characteristic of products. An iterative scheme is then conducted until sufficiently low levels of « silence » (missed relevant discourses) and « noise » (wrong elements) are reached. These patterns are divided in the two main categories : EMOTIONS and SENSATIONS, both divided in two sub classes giving the tone POSITIVE or NEGATIVE.

Design components identification can be conducted easier as related linguistic fields are generally less confusing than emotional ones. We can then mainly use keywords and related semantic words to identify quoted design components in a corpus. As said, we use here three main classes: AESTHETIC, FUNCTIONS, and USE as design component elements of products. The first class is including: colors, shapes, materials, size, etc. The second class is divided into two sub classes: operational functions and discourse functions which explain the former and the functionality of the product; and the last class is dealing with actions and operations necessary for users to use the product (grasping, transport, installation, storage ...).

### ***Tensive-diagram and verbatim generation***

A launch of our text-mining tool on a corpus (here blog posts) will automatically code any relevant part of discourse according to the chosen and tuned patterns. Defining a measure unit (here a post), we can simply build some statistical occurrence and co-occurrence of any of them in the corpus (Verbatim can then be extracted). It is then possible to generate a tensive-diagram such as Intensity/Extensity one, aiming at positioning any quoted design element in term of both:

- INTENSITY: in measuring the emotional density linked with the component
- EXTENSITY: in measuring the number of blog posts quoting the design component

Based on these elements, a 2D diagram can be drawn, where each design elements are positioned with their respective associated emotional density and extensity in blog discourses (see Figure 5, for our application case).

## **Semiotic analysis (2): identification of potential leads for innovation**

We have the general frame for values, strategies and innovation, the method to analyze data, but we need to link them with design research project. We need to transform, organize and put all these information in the same direction, to give to the brand a better coherence, i.e. a better chance of success; it means that design, users, strategy, value and innovation will be “aligned”. See in the application part of this article how all of this work.

## APPLICATION

### **Baby Car Seat Innovation**

The case study is the Brand, named X here, designing and producing baby car seat; the recommended strategy according to the study of its values and its willingness to innovate directing the company to choose a sensitive approach to the design of its products. The strategy is to navigate between planning and adjusting “in real time”; we used text-mining to build semantic categories related to sensitive and convert the data in terms of design. The goal is for the baby car seat manufacturer to design new product.

### **Semiotic Analysis (1)**

After consulting Brand X managers and a certain number of related strategic documents, we realized that Brand X built the attractiveness of its products on a range of utilitarian values, while wanting strategically to orientate them towards existential values.

#### ***The meaning trajectory: innovating in sensitive approach***

For its current innovations then, Brand X needs to develop a «sensitive approach» and use an “emotional valorization” for its products. We can imagine a certain number of questions linked to these innovations: Should all the products be developed in relation to this value? Can one deploy the three values that concern the brand (practical, ‘sensible’ and existential) while preferring the ‘sensible’ approach today? How does one ‘nourish’ the designers with these values? Etc. For each of these questions it is possible to carry out a ‘cahier des charges’ for the design, to carry out a survey of the product’s end users in order to determine the elements that make up the strategy. The theme of this conference being innovation and “emotional valorization” we focus here on design hypotheses.

#### ***Brand X values (or by what a subject promotes his acts)***

Brand uses mainly the practical value but sometime uses existential values for its products. Its goal is to migrate all of its products on the latter. If we use this square in order to situate Brand X, we get the following diagram:

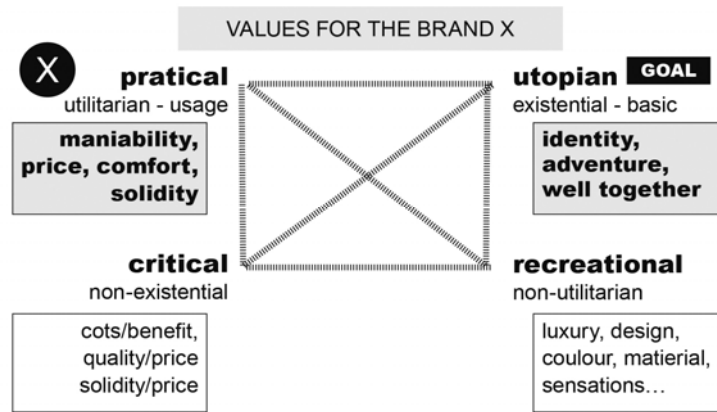


Figure 3. Semiotic square for the brand

**Brand X: values and trajectory**

Brand X conceives and produces products based on practical values; but some of its products are innovations which allow the brand to migrate towards its final goal: an existential valorization. However, as we have seen in the preceding square, the brand cannot go directly from a practical value to an existential value without running the risk of damaging or even breaking the contract of confidence with the receivers; it has to pass via or to promote a sensible valorization:

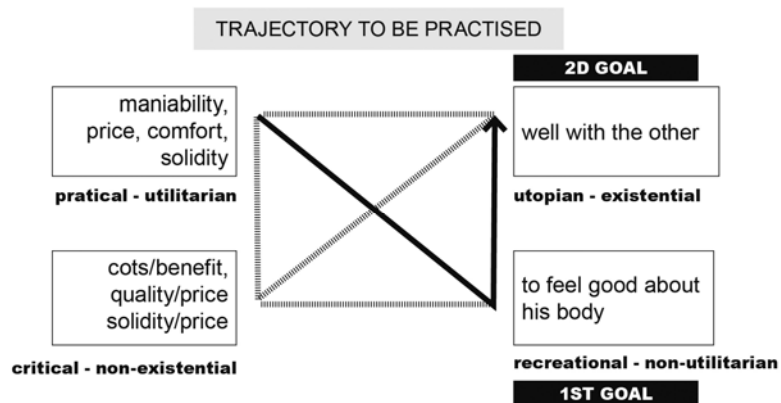


Figure 4. Recommended trajectory for Brand X

**Open discourses collection (Blogs)**

We built a corpus from verbatim collected in blogs and dealing with « car baby seat » (in French). The main characteristics are shown in the table below.

	Statistic (number of blog posts)
Total Number of blog posts	1 066
Number of emotional elements	395
Number of design component quotation	293

Table 1. Main corpus (blogs) characteristics

According to table 1 a rate of around 40% of emotional based posts around the product is quite high (in comparison with other studies) and reveals that some aspects experienced by user in daily life are of importance.

A slightly lower rate of 30% for design element quotation reveals that around 10% of emotional posts just mention in passing the product, and does not concern directly the product.

## Text-Mining Analysis

After tuning our text-mining tool and related filtering and pattern recognition means, we launched the text-mining analysis on the above described corpus. Each post was then assigned with both a code for the experienced emotion and for the concerned design element. Then a simple coding analysis yielded to the following diagram (see Figure 5).

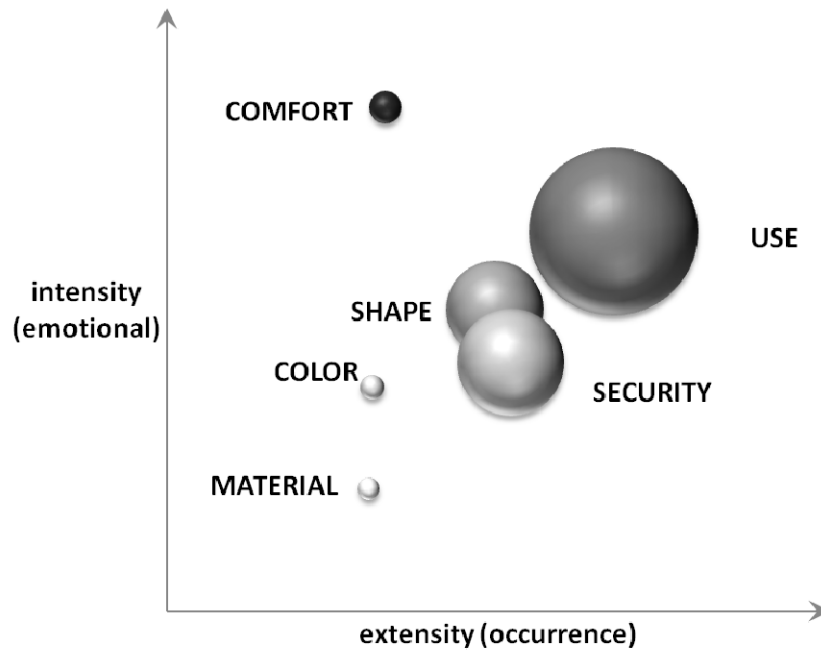


Figure 5 Tensive diagram for design components measured in blog posts

In that diagram, the larger the ball is (or more the ball is on the right) the bigger the occurrence of the corresponding design element is quoted in blogs. The darker the ball is (or more the ball is high on the diagram), the greater the corresponding emotion experienced by users (bloggers) is.

Main findings are:

- (1) Emotion mainly occurs when bloggers speak about COMFORT and USE of their Baby Car Seat;
- (2) COMFORT is less quoted in blog posts but is very often emotion charged, whereas USE is much often quoted in posts with a slightly lower emotion;
- (3) SHAPE and SECURITY are experienced in the same manner, middle emotion and quotation;
- (4) COLOR and MATERIAL, lower emotion and quotation.

### Verbatim samples (translated from French)

In upper case – keywords semantically linked with design elements; in bold some patterns related with emotion

>USE

*"But this day, nothing will stop me, I thus hoist baby quite high to make him pass between the SAFETY BELT and the CAR CEILING what leaves us a very little margin ... we **contort** ourselves, I'm on tiptoe and **phew!** : Alexandre finally sat in his seat ... now he should be BUCKLED : yes but to BUCKLE him, **I have to** get into the car and finally to remove his coat*

*because the belt isn't long enough and me not big enough to do all that from the outside ... in short : awful!"*

*« Even today again, I work up a good sweat each time I must INSTALL THE CAR SEAT, nevertheless I know how it goes”*

*« One hour, what am I saying, Two !!, Yes, to INSTALL this wretched car seat !! I'm not a blonde but even a genius would have cracked !!!*

>COMFORT

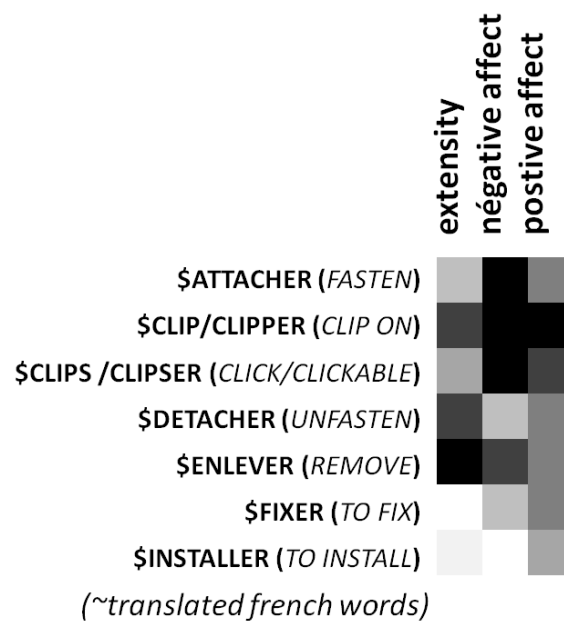
*“A hhhhhh and here is Mattéo with his COMFORTABLE CAR SEAT I think he is very happy and feels like a king in his car »*

*“Do not forget either that a COMFORTABLE SEAT for your baby will avoid you ceaseless tears because your baby is badly installed ! ”*

*«...it's such a pity. While our children are TRUSSED UP in their car seat”*

**Focus Baby Car Seat « USE »**

In that sub-section, we illustrate the use of Text-Mining generated « Heat Diagram » for focusing on emotional expression related to the design component USE (see Figure 6).



**Figure 6. « Heat diagram » on blog posts dealing with « USE »**  
*Darker are cells, lower are the correspond values*

- The shown heat diagram (darker are cells, lower are the correspond values) reveals:
- (1) INSTALLATION of the baby car-seat is clearly the most critical aspect among other component of the USE category;
  - (2) INSTALLATION is highly charged with emotion and especially with a high NEGATIVE affect
  - (3) UNFASTEN, less quoted, is also often charged with NEGATIVE affects.

## Semiotic analysis (2)

We didn't treat here all aspects of the analysis performed, but we extracted here some relevant traits to show the strength of the synergy between text-mining, design and semiotics.

If we take the field MATERIAL that has the smallest presence in terms of intensity (emotional) and "extensity" (number of occurrences; see fig. 5); in other words, the hypothesis to explore would be for users that "material" is not significant for uses and sensitivity; but designer knows the significant effects of material : it is soft and warm to the skin (fleece), another evokes coolness, allows breathing or even allows a grip without risk of slipping or struck. The mark could then have a strategic advantage (emotional value correlated with innovation in a strategy that is from the "brightness of the strange" to "universal deployment") to design its products taking into account the materials.

The field COMFORT is also to be explored (see figure 5, high intensity and low extensity) for its relation with the category modalities. This combination allows to specify user's needed knowledge and user's needed movement. Emotional and sensational data with modalities (to have to do, to be able to do, to want to do, etc.) can also be distributed according to the user (the baby) and the fitter (the adult) even more, it can be crossed with gender (boy-girl and man-woman). The difficulties encountered by both sides (male, female) can lead to design choices.

If we choose fields USE and SECURITY, the mark must found with the buyer-fitter a trust relationship in which the baby has a specific role. Adults trust in the brand so that it offers the child comfort and safety. If the child is well and secure, adults travel better, but at the same time the seat must also be easy to be fixed by adults. From there, it is possible to reconstruct a practical scene with actors, actions, contract between these actors, and the sanctions for each part (seat easy to fix means positive sanction - child in an uncomfortable situation means negative sanction).

## CONCLUDING REMARKS

The aim of this case study was to show how it is possible to link emotional user's words with a brand design strategy. The method described here and applied on a concrete case shows how text-mining and semiotic can be successfully crossed in the frame of innovation. On the one hand, semiotic is giving a robust theoretical analysis and potential leads for products or services innovation. On the other hand, text-mining on open discourses in blog gives, quite immediately even on a real-time base, concrete related elements from user's « daily life », and can be easily repeated. Crossing these two inputs allows analysts to build innovation strategy with more accuracy as soon as an iterative scheme can be simply set up between text-mining on open-discourse and semiotic. The method can be also successfully applied on other open-discourse sources such as: call center transcripts, open-questionnaires, customer emails, etc. More-over, cross or multilingual, and so related cultural aspects on emotion and sensation can be also carried out by the use of Text-Mining tools. Doing this would enrich on a larger scale the theoretical analysis of semiotic and hence strengthen innovation strategy.

## REFERENCE SECTION

- 1990 Floch Jean-Marie, *sémiotique, marketing et communication*, PUF
- 1991 Greimas A. J. & Fontanille J., *sémiotique des passions*, Seuil
- 1998 Lotman Iouri, *La sémiosphère*, Limoges, PULIM, coll. Nouveaux Actes Sémiotiques, 1998. Traduction d'Anka Ledenko
- 1998 Kapferer J.N., *Les marques capital de l'entreprise*, éditions d'Organisation
- 1999 Fraenkel & Legris-Desportes, *Entreprise et sémiologie*, Dunod
- 2003 Ceriani Giulia, *marketing moving*, L'Harmattan
- 2005 Fontanille Jacques, *Maisonnette & Larose*

- 2006 Fontanille Jacques, la sémiotique stratégique : pour une extension du domaine d'intervention, Université de Limoges, Institut Universitaire De France.
- 2007 Ronen Feldman, James Sanger, The Text Mining HandBook, CambridgeUniversity Press 2007
- 2007 Sigrid Maurel, Paolo Curtoni et Luca Dini, Classification d'opinions par méthodes symbolique, statistique et hybride, AFIA 2007 – DEFT'07
- 2007 Namrata Godbole, Manjunath Srinivasaiah, Steven Skiena, Large-Scale Sentiment Analysis for News and Blogs Int. Conf. on Weblogs and Social Media (ICWSM 2007), Denver CO, March 26-28, 2007
- 2009 Theresa Wilson, Paul Hoffmann, Janice Wiebe, Recognizing Contextual Polarity:An exploration of features for phrase-level sentiment analysis, Computational Linguistics 35(3)
- 2009 Rudy Prabowo, Mike Thelwall, Sentiment Analysis: A Combined Approach, Journal of Informetrics, 3(1), 143-157