

From Normal Business to Financial Crisis ... and Back Again. An Illustration of the Benefits of *Cassandra* for Qualitative Analysis

Christophe Lejeune

Key words:

Cassandra;
register; coding;
bounce technique;
free software;
KWIC; notional
family;
interpretation

Abstract: *Cassandra* is a free open source text analysis software tool. It uses semi-automatic coding, based on the identification of markers, grouped into registers, which represent analysis categories. Studying the causes of the financial crisis makes it possible to construct a series of registers (responsibility, negligence, domino effect, change and return to normal). Cross-referencing these registers reveals contrasting views (positive and negative) of the economic, political and financial spheres. These conceptions vary both according to how they explain the phenomenon and the measures that they advocate (ranging from the development of new methods to "getting back to normal"). The example of the financial crisis is an opportunity to discuss the benefits of semi-automatic coding and determine the limitations that are likely to impact negatively on the analysis.

Table of Contents

- [1. Introduction](#)
- [2. The CAQDAS Landscape](#)
- [3. The *Hypertopic* Platform](#)
- [4. Benefits of *Cassandra*'s Semi-Automatic Coding Features](#)
 - [4.1 Markers and concordances](#)
 - [4.2 Registers and the bounce technique](#)
 - [4.3 Semi-automatic coding](#)
- [5. Registers for the Financial Crisis](#)
 - [5.1 Causes, responsibilities and accusations](#)
 - [5.2 Solutions, change and getting back to normal](#)
- [6. The Limitations of Coding by Register](#)
 - [6.1 Homography and polysemy](#)
 - [6.2 Is semi-automatic coding still qualitative?](#)

[Acknowledgments](#)

[Appendix: Registers and Respective Markers for the Financial Crisis](#)

[References](#)

[Author](#)

[Citation](#)

1. Introduction

The use of software is being discussed more and more in the field of qualitative research. However, QDA (Qualitative Data Analysis) software frequently offers the same features. In this paper, I will introduce *Cassandra* as an alternate technique, based on semi-automatic coding, and closely examine its benefits and limitations. [1]

My demonstration rests on some elements of the analysis conducted as part of the KWALON experiment. The organizers of this experiment provided all participants with a large dataset consisting of video documents, press articles and websites (discussion forums) on the 2008 financial crisis. Given that *Cassandra* can only cope with texts, I have focused exclusively on text documents from the KWALON dataset. Readers interested in the analysis of video files are invited to read the interesting and productive paper by Paul DEMPSTER and David WOODS (2011). Being limited to text is the only restriction of the software. [2]

Indeed, unlike other participants (e.g. Susanne FRIESE, 2011, Anne KUCKARTZ & Michael SHARP, 2011), I did not choose to focus on a limited number of selected texts (which is one benefit of *Cassandra*'s design, see Section 4.3). The corpus imported into the software therefore consists of 101 newspaper articles and 67 discussion forums. However, I did not conduct an in-depth analysis of the whole corpus. Instead, I undertook an exploratory analysis. It mainly consisted of a detailed thematic inventarization (creating registers for each topic) coupled with an examination of the orientation of statements about these topics (positive or negative). [3]

My presentation adopts the following structure. The next section provides a brief overview of reflective coding, as the classic feature of qualitative analysis software. In Section 3, I introduce the *Hypertopic* collaborative platform to which *Cassandra* belongs. I will then show, in Section 4, how *Cassandra* supports the identification, display (in concordances) and gathering (in registers) of markers. In this section, I explain why it is possible to speak of a "semi-automatic" and "abductive" approach. Section 5 outlines some of the results of the KWALON dataset analysis. Finally, Section 6 discusses some of the limitations of a semi-automatic approach. [4]

2. The CAQDAS Landscape

Computer Assisted Qualitative Data Analysis Software (CAQDAS) is the collective term for tools, with which most practitioners of qualitative analysis are now familiar. More and more publications are dedicated to these tools (see TESCH, 1990; WEITZMAN & MILES, 1995; POPPING, 1997; KLEIN, 2001 and many others). Even general methodology handbooks look at how they work (FIELDING, 2001; BRYMAN, 2008, pp.564-584; SEALE, 2010). Most CAQDAS is based on the reflective coding of portions of text. [5]

Even if various software tools exist, most of them are based on the same coding device: the software offers a reading window (comparable to those used for word processing), which enables the researcher to highlight (using the mouse) any passages that are of interest and assign them a label (known as a node or code). This basic principle is common to most QDA software, such as *NVivo*, *Atlas.ti*, *MaxQDA* and *WeftQDA*. This software is thus designed as a tool, similar to felt-tip pens (LEJEUNE, 2010a, p.179). It encourages and supports analysis, which is primarily seen as an intellectual task conducted by the researcher. Indeed, in order to select excerpts, decide which are relevant and create appropriate labels, the researcher has to comprehensively read, empathize, reflect and interpret the texts. For the above reasons, I consider this method of coding is "reflective" rather than "free" or "manual." I therefore propose to consider QDA software as *reflective* coding devices. [6]

Of course, QDA software enables some supplementary features such as cross-tab. Codings can then be combined, recorded and manipulated in order to support transversal (GOBIN & DEROUBAIS, 1994, pp.69-70) or experimental (VIRBEL, 1994) reading. This type of reading makes it possible to examine texts differently, develop analysis categories, jump from one passage to another using one of these categories and test links (hypotheses) between these categories and their properties. [7]

Moreover, just about everywhere in the world, this method is considered the most compatible type of IT support for qualitative analysis in the field of human and social sciences (BARRY, 1998; FIELDING & LEE, 1998; GIBBS, 2002). The situation is different for the humanities and language sciences, in which more automated procedures are used. [8]

In France, the situation is significantly different (LEJEUNE, 2010a). Qualitative analysis is actually dominated by a tradition of *quantitative* discourse analysis. As a result, statistical tools for lexical analysis are used more often than QDA software. The software used most frequently in France is *Alceste* (an automatic text categorization tool) while, in many other countries, the most frequently used programs are reflective coding tools such as *NVivo*, *Atlas.ti*, *MaxQDA* (see LEJEUNE, 2010b for details). [9]

This fact motivated me to develop a tool that would offer the best of both worlds: the automatic and the reflective (LEJEUNE, 2010b). This methodological research project therefore aimed to design a software tool, called *Cassandra*, which has the refinement of QDA software, but enables the user to delegate certain tasks to the computer. This is made possible by *Cassandra*, a semi-automatic coding module on the *Hypertopic* platform. [10]

3. The *Hypertopic* Platform

The *Hypertopic* platform is made up of four software packages. In order to explain the respective role of each of these tools, I will start with a description of the *Hypertopic* platform. Users not interested in *Cassandra*'s IT background can skip the current section and jump to the results of my analysis. [11]

Designed by Aurélien BÉNEL, Chao ZHOU and myself, the *Hypertopic* collaborative platform for qualitative data analysis aims to support collaboration between researchers and therefore makes major use of the Internet. In a computer network, some computers host data and software that is shared with other computers. These 'host' computers are called servers, while the computer that uses the shared data or software is known as the client. For instance, some of the various servers available on the Internet host websites. I can visit these websites using a browser, which is a type of client software (like *Firefox* or *Explorer*). The *Hypertopic* platform is designed in the same way. It is made up of two servers (*Argos* and *Cassandra*) and two clients (*LaSuli* and *Porphyry*). Each of these tools plays a dedicated role in helping users to collaborate on qualitative analysis.

1. In the same way as traditional QDA software, the *Argos* server enables reflective coding of texts and images. Being Internet-based, *Argos* allows several researchers to work together on joint or concurrent analyses (ZHOU, LEJEUNE & BÉNEL, 2006).
2. The *Cassandra* server supports the semi-automatic coding of texts (Section 4).
3. Whether manual (with *Argos*) or semi-automatic (with *Cassandra*), coding is conducted using *LaSuli*, a *Firefox* add-on, which creates a margin to the left of the browser window (see BÉNEL & LEJEUNE, 2009 for details).
4. Finally, all coding can be consulted, manipulated and combined thanks to the client *Porphyry* (BÉNEL et al., 2001; ZHOU & BÉNEL, 2008). This researcher portfolio has an interface that lists all the researcher's codes (see Figure 1, top left) and displays the extracts with which they are associated. The size of the coding font is proportional to the number of associated extracts. The interface makes it possible to select several codes at the same time and therefore to conduct set-theoretic intersection operations. For instance, while selecting two registers at the same time (by selecting them while keeping the control key pressed), only excerpts related to both registers are displayed. This feature is similar to a Boolean intersection (GIBBS, 2002, p.145).

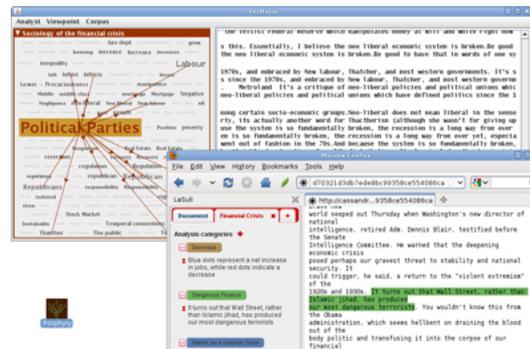


Figure 1: The *Hypertopic* platform, made up of *Porphyry*, *Cassandra*, *LaSuli* and *Argos* (enlarge this figure [here](#)) [12]

On the above screenshot, *Porphyry* (above) includes *Cassandra*'s semi-automatic codings (on the left) while, combined with *Argos*, *LaSuli* Firefox add-ons (below) provide a simple CAQDAS environment. [13]

Far from arguing the supremacy of one strategy over the others, the platform's designers BÉNEL, LEJEUNE and ZHOU (2010) believe, like WELSH (2002, §9), that the various approaches enrich one another. [14]

Within the scope of this paper, my aim is to introduce, document and test the use of semi-automatic coding. The following analysis is therefore based solely on *Cassandra* (and not *Argos*). As the *Cassandra* server can only be used with client software, the following screenshots feature *LaSuli* and *Porphyry*. Nor will I demonstrate the collaborative features of the *Hypertopic* platform, due to having focused on the proof-of-concept of the software's semi-automatic coding. Readers interested in collective analysis can find some illustrations in Fiona WILTSHIER's (2011) paper or in BÉNEL and LEJEUNE (2009). [15]

4. Benefits of *Cassandra*'s Semi-Automatic Coding Features

Cassandra is a free open source qualitative text analysis software tool. Its name comes from Greek mythology. Cassandra, Princess of Troy, had the gift of being able to predict the future. However, after she refused his advances, the god Apollo placed a curse on her so that no one would ever believe her prophesies. Cassandra was therefore condemned to predict misfortunes without being able to prevent them. This was a sad fate for a princess. And a surprising choice for the name of a software tool! [16]

In fact, practitioners of qualitative research often suffer from the *Cassandra* complex, in that they produce interesting, serious and solid results but their colleagues fail to take them seriously. This is why I developed the *Cassandra* software. This tool actually enables researchers to show colleagues or clients the categories, on which the conclusions are based. It makes it possible to easily *replicate* the analysis. The software therefore attempts to overcome the

Cassandra complex affecting qualitative analysis. Rather than expressing a feeling of resignation, it is the name given to an ambitious project. [17]

4.1 Markers and concordances

As already stated, *Cassandra* is a server that enables the semi-automatic coding of texts. In technical terms, it is based on a fairly simple principle.

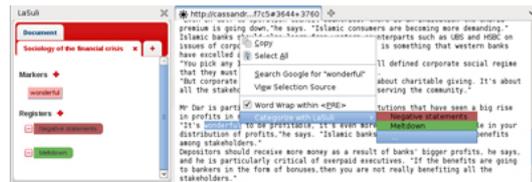


Figure 2: Creating a marker with *LaSuli* (enlarge this figure [here](#)) [18]

When the researcher codes texts, he or she no longer selects long text passages, but a single word (or expression) that is likely to occur several times in the corpus (Figure 2). Whereas, with traditional QDA software, the analysis would highlight a complete sentence, such as "It's wonderful to be equitable in your distribution of profits," he or she now only selects "wonderful." The software now sees "wonderful" as a *marker* and instantly gathers together all the passages that include this word. This type of marker is therefore a word or an expression taken directly from the data. In this way, markers are closer to *in vivo* codes than traditional QDA software codes. Indeed, unlike markers, reflective coding serves to highlight pieces of data and attach a label to them (ultimately distinct from data). Because they are used for indexing, markers cannot differ from the data. They are keywords or expressions from the data. [19]

At this stage, it is important for the analyst to be able to evaluate whether the various passages gathered truly deal with the same phenomenon. This is not always the case, as there will be homographs, ambiguities and polysemies. In order to facilitate this evaluation task, *Cassandra* stacks the various passages by horizontally aligning the various occurrences of the same marker one above the other (in yellow in Figure 3). Known as a *concordance*, this type of presentation dates back to the 13th century (LEJEUNE, 2010a). Hans Peter LUHN (1960) introduced it to information technology under the name "Keyword In Context" (KWIC).

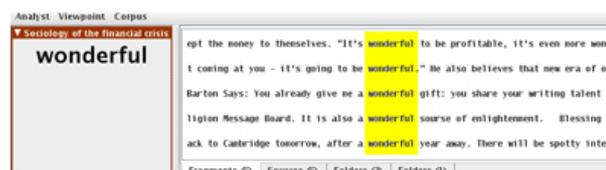


Figure 3: KWIC concordance of "wonderful" in the KWALON corpus (original spelling from the material) (enlarge this figure [here](#)) [20]

Concordances therefore help the researcher to evaluate his/her markers. Moreover, a phenomenon can usually be expressed in various ways. During his/her exploration of texts, the researcher therefore identifies several markers for the same phenomenon. In addition to "wonderful," I therefore created the markers "right," "fair" and "clear," which express a similar (positive) idea. [21]

4.2 Registers and the bounce technique

Once these different markers have been identified and evaluated, the researcher brings them together under one single label. *Register* is the name that I have given to these sets of markers identified by the researcher throughout his/her analysis (LEJEUNE, 2008). In this case, I have combined the above-described markers under the register of positive statements (see Figure 4). Registers are comparable to normal reflective codes, in that they consist of a label that gathers various pieces of data. But there are also differences. First of all, while CAQDAS codes are purely reflective constructions, registers are partly reflective and partly automated. Indeed, the researcher reflectively links markers to a register, but markers automatically match a number of text passages. Moreover, given that a register is a set of markers and that every marker is likely to match a number of passages, a register is likely to be related to more passages than markers. This is not the case with reflective codes, in which every passage is directly (and deliberately) linked to the code. Differences between reflective codes and registers are therefore slight, but they however exist.

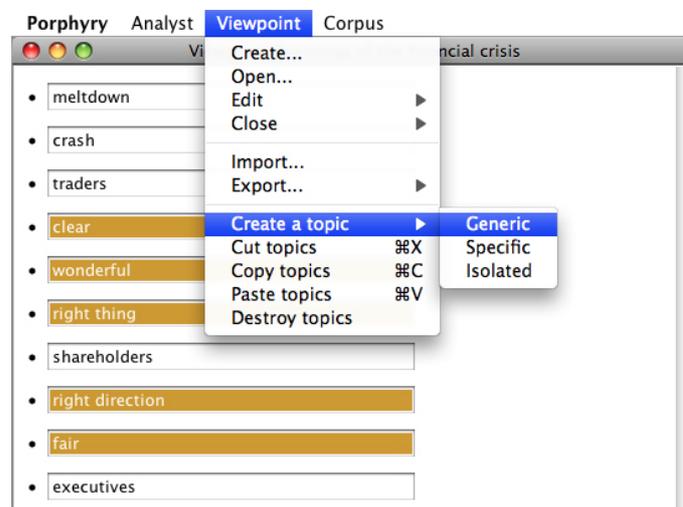


Figure 4: Markers are gathered into a register (screenshot of *Porphyry*) [22]

A register therefore helps to gather together various markers that refer to the same phenomenon. The relevance of each marker is evaluated by means of concordances. In the case of the "positive statements" register, I checked in each case whether the passages effectively corresponded to comparable phenomena. Considering the register of positive statements, this was the case for "fair" and "clear" markers but not for "right," which is a polysemic marker. Far from systematically referring to positive statements, many occurrences of "right"

actually denote a political orientation ("right wing"). By means of the concordance, I diagnosed that "right," however, occurs in non-ambiguous expressions such as "right direction" (which belongs next to "wonderful"). If I had kept the single keyword "right" as a marker for the "positive statements" register, sentences including occurrences of "right wing" would have been coded as "positive statements," which would, of course, have been misleading. The "right" marker was thus removed from the "positive statements" register and replaced with compounds: "right direction" and "right things." In the meantime, "right wing" and "right-wingers" were added as markers of the "political parties" register. Deleting the "right" single marker avoided misleading codes. [23]

I refer to this strategy as the bounce technique. It serves to create a temporary marker (e.g. "right") as a proxy, which makes it possible to gather together all the matching excerpts. After a close examination of the resulting concordance, the temporary marker is deleted and replaced with a more complex (and therefore more specific) expression (such as "right thing" and "right direction"). The bounce technique is a useful (albeit casual) trick of the trade for qualitative researchers using semi-automatic coding procedures. [24]

Given that no QDA software combines the use of registers and concordances in this way, this represents a new technique for qualitative social science research. However, just like concordances, registers do not represent a revolutionary technological innovation. In fact, Hans Peter LUHN (1957, p.314) proposed this type of semi-automatic coding under the name *notional family*. By using registers (a sort of notional family) and KWIC (for concordances), *Cassandra* therefore makes double use of the work completed by this German researcher from IBM. [25]

4.3 Semi-automatic coding

When I refer to registers or notional families as *semi-automatic* coding, this means that the registers fall mid-way between automatic categorization software (such as *Alceste*) and traditional QDA software (such as *NVivo*). The challenge is to get the best out of both approaches. [26]

On the one hand, registers benefit from the automation facilitated by computers. Each time a marker is selected, the software automatically identifies all the passages, in which the word (or expression) occurs. This is precisely the advantage of automation. Many passages in a wide range of texts can be linked to a register (through markers) and quickly evaluated by the researcher (using concordances). This means that I no longer have to examine a sample from the KWALON dataset and allows me to look at it as a whole (i.e., [a total of 168 text sources](#)). [27]

On the other hand, creating a register necessitates careful reflection on the part of the researcher. First of all, he or she identifies the markers that are able to capture text fragments, which are relevant to a studied phenomenon, by means of concordances. He or she then quickly evaluates whether the fragments collected by the software effectively correspond to the relevant phenomenon. He

or she also identifies the markers that "go together" and creates a *notional family*. Finally, he or she labels the relevant register with a sensitizing concept (BLUMER, 1969, p.148). The selection, evaluation and assembly of markers, and labeling of registers depend on the sensitivity, intelligence and reflection of the researcher; they cannot be automated. A human brain is therefore needed for such a task. [28]

This compromise between automation and reflection aims to assign to the computer and the human being what each does best. The aim is not so much to gain time or analyze more text, but to create an innovative, non-linear and transversal journey through empirical material. Having created *Cassandra*, I am convinced that this source of assistance supports qualitative research. More precisely, registers facilitate the *abductive* generation of analysis categories from the material. Charles Sanders PEIRCE originally devised the term *abduction* as a mode of reasoning that is essentially different from deduction and induction. In qualitative research, abductive reasoning starts with data and develops theories in a bottom-up, progressive and iterative way (ATKINSON & DELAMONT, 2005, p.833; CHARMAZ, 2006, p.103; REICHERTZ, 2010). The interactive movements between texts, markers, concordances of excerpts and registers are especially well suited to an abductive approach. [29]

However, software is not confined to its original uses. I have actually seen *Cassandra* put to a wide range of uses (LEJEUNE, 2010c). Contrary to my aims, registers can therefore be used as part of a *deductive* approach, as predefined categories "tacked" onto a corpus of texts. This working method does not correspond to my approach, but represents an epistemologically valid approach, which can be seen, for example, in quantitative content analysis dictionaries (KRIPPENDORFF, 2004, pp.283-289; see also BÉNEL et al., 2010, pp.49-52, for a discussion on the difference between deductive dictionaries and abductive registers). As FIELDING and LEE (1998) warned, however, these deductive uses lead to the risk of a certain level of superficiality. [30]

The intermediate position of registers therefore offers advantages but also includes risks. These risks may explain why so few software tools offer this type of functionality, although semi-automatic coding has existed since the 1950s. Only the *autocoding* provided by *Atlas.ti* represents a similar tool. When asked about this, Susanne FRIESE confirmed the essential problem of semi-automatic coding. As an *Atlas.ti* training provider, she actually avoids explaining too soon how auto-coding works as, after being attracted by the rapidity of this tool, learners later turn their backs on fine (and reflective) coding of empirical material (Susanne FRIESE, personal communication, in 4/2010). [31]

5. Registers for the Financial Crisis

As an initial method of tackling the corpus, which had to be exploratory, I created the "financial crisis" marker. This enabled me to see, at a glance, the concordance (KWIC) of all occurrences of this expression. I then realized that to describe this event as a crisis was, obviously, already an interpretation. So that I could capture all the points of view expressed in the material, it was therefore important not to assume that the event, on which the text is based, is a crisis. I therefore took stock of the way, in which this event was described and split the marker into two registers, the first of which consists of markers that refer to the event itself: meltdown, crisis, turmoil, crash, failure (see [Appendix](#) for a full list of markers for each register). The second register consists of financial markers: Stock Market, the Exchange, Wall Street (the case—and capitalization—used in these markers originate from corpus words). [32]

Making a list of things that go together and things that do not is a very good exploratory technique for delving into a text corpus. This is why the researcher creates a wide range of registers during the early stages of exploring the corpus. This "swarming" is obtained in various ways. One productive strategy is quite similar to the bounce technique (introduced in Section 4.2). In this case, the proxy of the bounce technique is not limited to a sole marker, but can be a register as a whole. By means of concordances, the researcher scrutinizes all the topics around one register, which is used as a proxy. He or she creates a register for each identified topic and focuses his/her attention on this new register, which is used, in turn, as a proxy to identify and construct the next registers. In the KWALON material, I scrutinized the elements surrounding the first two registers ("meltdown" and "stock market") and, following this snowball effect, I developed registers identifying all actors involved in the events: workers, members of the public, trade unionists, political parties and white collar workers. This exploration soon prompted me to develop registers about the causes and responsibilities that actors attributed to the crisis. [33]

5.1 Causes, responsibilities and accusations

I found attempts to explain the crisis both in press articles and discussion forums, which link the crisis to subprimes and hedge funds. I resisted the temptation to create a single register for subprimes and hedge funds for the same analytical reasons that prompted me to create two separate registers for "financial" matters and "crisis" events. I therefore created a real estate and a mortgage register. [34]

The identification of causes (which has been one of the questions suggested by the experiment organizers) also draws upon the identification of responsibilities. Attributing responsibilities boils down to naming culprits: in other words, registers for the causes and responsibilities are found in critical arguments concerning (for example) the trader bonus system, inconsistency of banks, negligence of directors and *laissez-faire* approach of public authorities. In order to identify the criticisms, I created a register of negative statements, which includes markers such as "bad," "dangerous," "toxic," "stupid" and "evil." At the same time, I also

developed a register of positive statements (mentioned in Section 4.2), including markers such as "wonderful," "right direction," "fair" or "clear." [35]

Of course, the critics were not in unanimous agreement and could not agree on responsibilities. Every argument is contradicted by a counter-argument in the data. Some explanations are reduced to unverified allegations or beliefs and therefore discredited. [36]

5.2 Solutions, change and getting back to normal

In various sources, the debate stumbles on the social responsibility of financial players. The conception of actors motivated only by interest (as in the rational actor paradigm) raises the question of the justifications of liberalism. Once again, several views can be observed. For some, the market self-regulates and any intervention (particularly by the state) constitutes disruption. In this perspective, the activities of the market, companies and banks are synonymous with the *common good*, as they create wealth, which benefits the various actors within the system. This type of stance combines the register of positive statements and that of the market. Their rhetoric typically corresponds to the *market polity* identified by BOLTANSKI and THÉVENOT (2006). [37]

For others, the unbridled race for profit can only lead to catastrophe, of which the financial crisis is a sad example. In addition, trader bonuses are out of proportion, and even indecent, compared to the poverty, uncertainty and distress suffered by people who take out loans to buy houses. With *Porphyry*, I intersected passages coded with the financial register and those coded as negative statements. This is done by clicking, with the mouse, on the former ("Stock Market"), keeping the control key pressed and clicking on the label of the latter ("Negative"). The portfolio then displays excerpts related both to stock markets and negative statements at the same time (see the left-hand panel on the Figure 5).

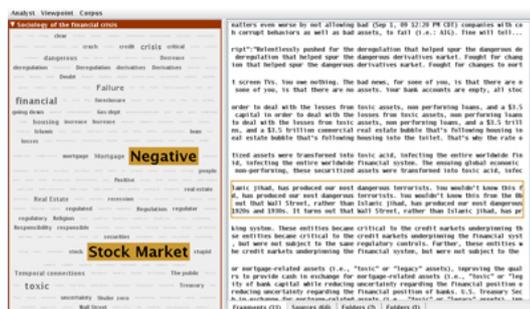


Figure 5: Intersecting two registers ("Negative" and "Stock Market") with *Porphyry* (enlarge this figure [here](#)) [38]

This intersection feature helped me to examine excerpts expressing critical views of the financial world. The criticism may be extreme, when, for example, a stakeholder reports that "Wall street has produced our most dangerous terrorists." In post-9/11 America, this comparison sounds like an unconditional condemnation. As a result, bearing in mind that self-discipline does not seem to

encourage moderation on the part of financial players, regulation by the public authorities is desirable and necessary. [39]

As could have been predicted, actors that consider the marketplace to be a common good and those who are radically critical of the financial world propose very different measures aimed at bringing an end to the crisis. [40]

Comparing the crisis to a procedural error or breakdown, the first view suggests that the system should be re-established as it was before. Its key themes are reconstruction and getting back to normal. The *corrective* measures that it proposes are based more on normal management. [41]

Considering the crisis led to a disaster, the second point of view argues for a change in practices and mentalities. Enthusiastic and proactive, it advises all concerned to learn from their mistakes and make a fresh start. It proposes measures based on *radical* change. [42]

The differences found between the two views are staggering. [43]

6. The Limitations of Coding by Register

Cassandra is an "open machine" (SIMONDON, 1958, p.11), which does not say anything without a human interpreter, who acts as a permanent organizer. Any theories generated stand or fall by the registers that they are based on (BERELSON, 1952, p.147). In turn, the relevance of registers depends on their markers. Ultimately, markers rely on the skills of the researcher. In order to set appropriate markers, the social scientist needs to be inventive and have an intimate knowledge of the field. This background knowledge includes cultural and language peculiarities, that help him/her to interpret markers without losing specific local assumptions (SCHUTZ, 1962). With the example of the "right" marker (see Section 4.2), I have demonstrated the importance of vigilance on the part of the researcher, when evaluating passages that match a marker. The suitability of a marker actually varies between corpuses. It is therefore possible that the "right" marker may be suitable for a corpus that does not deal with politics. But it has not been sufficiently specific for a corpus on the financial crisis. It cannot therefore be decided whether a marker is suitable independently of the corpus. Unless you succumb to superficiality, a marker cannot therefore be defined in advance, without close examination of the matching passages (see Section 4.3). As for many other computer programs, the main limitation of *Cassandra* lies in the conscientious approach required of the researcher conducting the analysis. [44]

6.1 Homography and polysemy

Semi-automatic coding is sensitive to the phenomenon of homography, which refers to the fact that two similar text strings can actually be different words. In some cases, two different words are written the same but belong to different grammatical categories. For instance, *can* is either a modal auxiliary or a noun

referring to a metal container, in which soft drinks are sold. In other cases, the same text string can refer to different meanings (linguists then speak then of polysemy). [45]

The construction of a register for the 2008 "meltdown" provided me with an excellent example of homography. The meltdown register is made up of various markers, such as "broken," "collapse," "downturn" and "recession." While gathering relevant markers, I included "fall" as an appropriate marker for the meltdown register. This inclusion was motivated by excerpts similar to the following.

"However the trend does not end there as preliminary analysis suggests the housing market could **fall** by 25% by mid 2011" (excerpt from the Marketoracle.co.uk website, KWALON dataset).

"Though there is a **fall** in demand for our products, we do not care as we believe it shall pick up soon (excerpt from a BCC discussion forum on the financial crisis" KWALON dataset). [46]

Facing such passages, the researcher must be aware of the different grammatical categories: in the first quotation, "fall" is a verb while, in the second quote, "fall" is a common noun. Anyway, in both passages, "fall" refers to a sudden movement from top to bottom. Both excerpts thus express the same meaning of "fall"; it is right to consider that they go together. Here, the difference of grammatical category does not raise any problems (but might be relevant in other cases). [47]

Indeed, the story of the financial crisis is all about falls, which can often form part of a domino effect. Many commentators express the view that the fall (of house prices, market rates, etc.) will continue. At this stage, it therefore seems appropriate to include the "fall" marker in the "meltdown" register, together with "broken," "collapse," "downturn" and "recession." [48]

In the following passages, however, some occurrences of "fall" refer to something other than a sudden decrease.

"All financial institutions, wherever they're regulated, must stand ready with strong capital reserves to serve as a cushion during times of unexpected market and economic difficulties. This must be combined with adequate loan loss reserves, to cover the expected losses from the growing number of borrowers who likely will default, and necessary liquidity, in case credit markets freeze up as they did last **fall**" (excerpt from the opinion rubrics of the Wall Street Journal, June 2009 the 29th, KWALON dataset).

"The investment climate has been further soured by Mrs. Kirchner's nationalization of pension funds and an airline since the start of the global financial crisis last **fall**" (excerpt from the Wall Street Journal, June 2009, KWALON dataset). [49]

In the two last passages, "fall" refers to a season (autumn 2008) rather than a top-down movement. This could have prompted me to remove the "fall" marker from the meltdown register. Indeed, this case is similar to the situation of the "right" marker. As I discussed in Section 4.2, "right" can be a (positive) evaluation, spatial direction or political orientation. In the case of "right," the solution was to delete the marker and replace it with more specific compounds. The situation is slightly different in this case. Indeed, when reading the relevant extracts, it struck me that the season is actually a metonymic evocation of events that occurred (see CALABRESE, 2010 on metonymic evocation of events). The ambiguity (between the movement and the season) could have been a problem, but this is not the case here. As a result, I have kept the marker unchanged. [50]

This situation is not, however, very typical. Homographic phenomena cannot always be bypassed. Most of the times, homographs are likely not to refer to the same phenomenon. For instance, in the corpus studied, two personalities bear the same surname: Dennis and Tony BLAIR. The two men were assigned very different roles in arguments: the former was the director of national intelligence in OBAMA's administration (he has since resigned) while, of course, the latter was the prime minister of the UK, before the events took place (he retired in 2007). After examining the concordances, it is actually impossible to discriminate once and for all between occurrences of these two examples: they have to be examined on a case-by-case basis. It was not without surprise that I then realized that the former is only mentioned on discussion forums, while the latter only appears in the press. I did not draw any conclusion from this absence. But it served to remind me of one of the limitations of textual analysis tools. Diagnosis always depends on what is said (or not) in the material. Absence in the material does not imply absence from the relevant events. Similarly, a presence in the material does not imply an effective role in the events. Experienced qualitative researchers always bear in mind that such gaps are always likely to occur. Gaps between material and events are sometimes due to a bias in document gathering, but more frequently, they are due to a lack of visibility (for actors) of selected events, objects, causes, arguments or persons. Such gaps belong to the tricky problem of "unobservables." They remind us that materials, such as newspapers and discussion forums, provide clues about social representations, arguments and modalities of interactions, rather than events or external elements. [51]

This limitation does not imply that any interpretation must be limited to explicit pieces of data. In particular, reflective coding from traditional QDA software makes it possible to read between the lines. Reading between the lines is, however, more difficult with semi-automatic coding, because registers need some explicit formulations for a marker to be matched to a passage. This limitation strengthens the argument that reflective coding and semi-automatic coding are complementary, rather than concurrent (WELSH, 2002, §9; BÉNEL et al., 2010). [52]

6.2 Is semi-automatic coding still qualitative?

When I demonstrate *Cassandra*, some colleagues are intrigued by the fact that the font size for the register names in the *Porphyry* portfolio is proportional to the number of associated extracts (see Section 3). With reference to this functionality, they begin to ask themselves whether *Cassandra* is really a qualitative analysis tool. [53]

This is a valid question. The differences in size are actually a result of quantification. As I mentioned at the beginning of this article, the challenge for the software lies in combining two logics. *Cassandra* is thus not simply a code-and-retrieve tool. However, the quantification is neither determining nor probative. Unlike quantitative content analysis, counting is not an end in itself, but a means of supporting the researcher with his or her interpretative task. [54]

The very marginal role of counts does not therefore allow us to describe *Cassandra* as a software tool for mixed methods research. It does not even claim to be partially quantitative. [55]

Ultimately, *Cassandra* does not "process" qualitative data. It assists and supports the researcher as part of a process that constantly renews itself. The tool is neither a guide nor a source of evidence; it is a method that owes its value to the vigilance of the analyst. [56]

Acknowledgments

The author is grateful to Aurélien BÉNEL, Jeanine EVERS, Jean-Pierre HIERNAUX, Nicolas MAZZIOTTA, Katja MRUCK for their comments on the first draft of this article and to David WARD for helping to revise the preliminary version and correct the English text.

Appendix: Registers and Respective Markers for the Financial Crisis

| Registers | Markers |
|---------------|--|
| Bonuses | bonus, bonuses, compensation, Golden parachutes, overpaid, perks |
| Causality | because, cause, fault, guilty, thanks to, why |
| Change | fix, fixed, reformed, Reforming, reform |
| Common good | community, distribute, distributed, distributing, distribution, distributions, distributors, equitable |
| Deregulation | deregulation, Deregulation, unregulated |
| Domino effect | domino, dominos, epidemic, spiral |
| Doubt | doubt, uncertainty |

| Registers | Markers |
|----------------------|---|
| Meltdown | apocalypse, broken, carnage, crashes, crash, crisis, collapse, collapses, decrease, Depression, disaster, downturn, Downturn, failure, fallen, fall, fell, flop, meltdown, recession, turmoil |
| Famous economists | Keynes, Malthus, Marx, Ricardo |
| Gov dept | Treasury, work and pensions, government departments |
| Mortgage | subprime, Lehman Brothers, mortgage, loan, credit |
| Negative | stupid, critical, bad, evil, toxic, dangerous, crap, terror, terrorist, terrorists, terrorism |
| Negligence | abuse, abusive, con artist, deceit, fraud, manipulation, manipulates, negligence |
| OECD | Economic Co-operation and Development, OECD |
| Pol Parties | leftists, leftist, Labour, New labour, Neo liberal, neo liberal, right wing, Right-wingers, republican, Republicans, republicans, Reaganism, Thatcher, Thatcherism |
| Positive | clear, fair, right thing, right direction, wonderful |
| Precariousness | indigence, inequalities, inequality, poor, poverty, unemployment, Unemployment |
| Proof | evidence, proven, proven |
| Public money | public funds, public finances, public intervention, public spending |
| Real Estate | housing, real estate, Real Estate |
| Recovery | rebuild, restore, restoring, recovery, repair, restoration, turnaround, rebuilding, recover, sunshine, recovered, restored, rebirth, Restoring |
| Regulation | regulators, regulated, regulatory, Regulation, regulation, regulator, regulate |
| Religion | God, Islamic, Jesus, sharia |
| Responsibility | accountable, accountability, anticipate, foresee, integrity, reserves, responsible, responsibility |
| Stock Market | assets, bonds, bond, CDS, Derivatives, derivatives, equities, financial, securities, swaps, the Exchange, stock, Wall Street, |
| Sustainable | environmental, Sustainability |
| Temporal connections | before, after |
| The public | consumer, consumers, customer, customers, people, public opinion, taxpayer, taxpayers |
| Trade Unions | trade union, tradeunions, unions |

| Registers | Markers |
|---|--|
| Transparency | oversee, overseen, oversight, transparent, transparency |
| Trust | confidence, trust |
| Under zero | bankruptcy, bankruptcies, debt, in the red, loss, losses |
| Unverified | belief, believable, placebo story |
| White collars | advisors, bankers, businesses, Directors, executives, shareholders, traders |
| Workforce | labor forces, workforce |
| Markers used as proxy in bounce technique | Blair, collateralized, foreclosure, hedge funds, malinvestment, stakeholders, vehicles |

References

- Atkinson, Paul & Delamont, Sara (2005). Analytic perspectives. In [Norman Denzin](#) & Yvonna Lincoln (Ed.), *Handbook of qualitative research* (pp.821-840). London: Sage.
- Barry, Christine (1998). Choosing qualitative data analysis software: Atlas/ti and Nudist compared. *Sociological Research Online*, 3(3), <http://www.socresonline.org.uk/3/3/4.html> [Date of access: November 18, 2010].
- Bénel, Aurélien & Lejeune, Christophe (2009). Humanities 2.0: Document, interpretation and intersubjectivity in the digital age. *International Journal of Web Based Communities*, 5(4), 562-576.
- Bénel, Aurélien; Lejeune, Christophe & Zhou, Chao (2010). Éloge de l'hétérogénéité des structures d'analyse de texte. *Document Numérique*, 13, 41-56.
- Bénel, Aurélien; Egyed-Zsigmond, Előd; Prié, Yannick; Calabretto, Sylvie; Mille Alain; Iacovella, Andréa & Pinon, Jean-Marie (2001). Truth in the digital library: From ontological to hermeneutical systems. In Panos Constantopoulos & Ingeborg Sølvberg (Eds.), *Research and advanced technology for digital libraries. Lecture notes in computer science*, 2163 (pp.366-377). Berlin: Springer.
- Berelson, Bernard (1952). *Content analysis in communication research*. Glencoe: The Free Press.
- Blumer, Herbert (1969). *Symbolic interactionism. Perspective and method*. Berkeley: University of California Press.
- Boltanski, Luc & Thévenot, Laurent (2006). *On justification. Economies of worth*. Princetown: Princetown University Press.
- Bryman, Alan (2008). *Social research methods*. Oxford: Oxford University Press.
- Calabrese, Laura (2010). Le rôle des désignants d'événements historico-médiatiques dans la construction de l'histoire immédiate. Une analyse du discours de la presse écrite. *PdD Thesis, Université Libre de Bruxelles*.
- Charmaz, Kathy (2006). *Constructing grounded theory. A practical guide though qualitative analysis*. London: Sage.
- Dempster, Paul G. & Woods, David K. (2011). The economic crisis though the eyes of Transana. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Art. 16, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1101169> [Accessed: January 30, 2011].
- [Fielding, Nigel](#) (2001). Computer applications in qualitative research. In Paul Atkinson, Amanda Coffey, Sara Delamont, John Lofland & Lyn Lofland (Eds.), *Handbook of ethnography* (pp.453-467). London: Sage.
- Fielding, Nigel & Lee, Raymond (1998). *Computer analysis and qualitative research*. London: Sage.

- Friese, Susanne (2011). Using ATLAS.ti for analyzing the financial crisis data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Art. 39, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1101397> [Accessed: January 30, 2011].
- Gibbs, Graham (2002). *Qualitative data analysis. Explorations with NVivo*. Buckingham: Open University Press.
- Gobin, Corinne & Deroubaix, Jean-Claude (1994), *Quand la commission se présente devant le parlement*. Bruxelles: Recherche européenne en Sciences Humaines.
- Klein, Harald (2001). Overview of text analysis software. *Bulletin de Méthodologie Sociologique*, 70, 53-66.
- Krippendorff, Klaus (2004). *Content analysis. An introduction to its methodology*. Thousand Oaks: Sage.
- Kuckartz, Anne M. & Sharp, Michael J. (2011). Responsibility: A key category for understanding the discourse on the financial crisis—Analyzing the KWALON data set with MAXQDA 10. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Art. 22, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1101222> [Accessed: January 30, 2011].
- Lejeune, Christophe (2008). Au fil de l'interprétation. L'apport des registres aux logiciels d'analyse qualitative. *Revue Suisse de Sociologie*, 34(3), 593-603, <http://hdl.handle.net/2268/6123> [Date of access: November 18, 2010].
- Lejeune, Christophe (2010a). A brief history of software resources for qualitative analysis. In Claire Brossaud & Bernard Reber (Eds), *Digital cognitive technologies. Epistemology and knowledge society* (pp.169-186). Hoboken-London: Wiley-ISTE, <http://hdl.handle.net/2268/2473> [Date of access: November 18, 2010].
- Lejeune, Christophe (2010b). Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative. *Recherches Qualitatives*, Hors série 9, 15-32, <http://hdl.handle.net/2268/61098> [Date of access: November 18, 2010].
- Lejeune, Christophe (2010c). *Cassandra*, un outil pour construire, confronter et expliciter les interprétations. *2e Colloque International Francophone sur les Méthodes Qualitatives*, <http://hdl.handle.net/2268/25352> [Date of access: November 18, 2010].
- Luhn, Hans Peter (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal*, 1(4), 309-317.
- Luhn, Hans Peter (1960). Keyword-in-context index for technical literature. *American Documentation*, 11(4), 288-295.
- Popping, Roel (1997). Computer programs for the analysis of texts and transcripts. In Carl Roberts (Ed.), *Text analysis for the social sciences. Methods for drawing statistical inferences from texts and transcripts* (pp.209-221). New Jersey: Lawrence Erlbaum.
- Reichert, Jo (2010). Abduction: The logic of discovery of grounded theory. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 11(1), Art. 13, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1001135> [Date of access: December 14, 2010].
- Schutz, Alfred (1962), *Collected papers I. The problem of social reality*. Dordrecht: Martinus Nijhoff.
- Seale, Clive (2010). Using computers to analyse qualitative data. In David Silverman (2010), *Doing qualitative research* (pp.251-267). London: Sage.
- Simondon, Gilbert (1958). *Du mode d'existence des objets techniques*. Paris: Aubier.
- Tesch, Renata (1990). *Qualitative research: Analysis types and software tools*. New York: The Falmer Press.
- Virbel, Jacques (1994). Annotation dynamique et lecture expérimentale: vers une nouvelle glose?. *Littérature*, 96, 91-105.
- Weitzman, Eben & Miles, Matthew (1995). *A software source book. Computer programs for qualitative data analysis*. London: Sage.
- Welsh, Elaine (2002). Dealing with data: Using NVivo in the qualitative data analysis process. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 3(2), Art. 26, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0202260> [Date of access: November 18, 2010].
- Wiltshier, Fiona (2011). Researching with NVivo 8. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Art. 23, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1101234> [Accessed: January 30, 2011].

Zhou, Chao & Bénel, Aurélien (2008). From the crowd to communities: New interfaces for social tagging. In Parina Hassanaly, Athissingh Ramrajsingh, Dave Randall, Pascal Salembier and Matthieu Tixier (2008), *Proceedings of the eighth international conference on the design of cooperative systems* (COOP'08) (pp.242-250) Aix en Provence: Institut d'Études Politiques d'Aix en Provence.

Zhou, Chao; Lejeune, Christophe & Bénel, Aurélien (2006), Towards a standard protocol for community-driven organizations of knowledge. In Parisa Ghodous, Rose Dieng-Kuntz and Geilson Loureiro (2006), *Leading the web in concurrent engineering* (pp.438-449). Amsterdam: IOS Press, <http://hdl.handle.net/2268/4186> [Date of access: November 18, 2010].

Author

With a PhD in sociology, *Christophe LEJEUNE* is a scientific expert in the field of qualitative method at the University of Liège. He has completed several studies on the free software movement, in particular the social organization of mediated collectives, the regulation of discussion forums and the expression of trust online. He is also the author of the free *Cassandra* qualitative analysis software described in this article.

Contact:

Christophe Lejeune

Université de Liège
Institut des Sciences Humaines et Sociales (ISHS)
Faculté de Psychologie et des Sciences de l'Éducation (FAPSE)
École de Gestion (HEC)
Bd du Rectorat, 7, B31
B-4000 Liège
Belgium

E-mail: christophe.lejeune@ulg.ac.be

URL:

<http://www.squash.ulg.ac.be/index.en.html>

Citation

Lejeune, Christophe (2011). From Normal Business to Financial Crisis ... and Back Again. An Illustration of the Benefits of *Cassandra* for Qualitative Analysis [56 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Art. 24, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1101247>.

Revised: 2/2011