**Available Online at www.ijcsmc.com**

**International Journal of Computer Science and Mobile Computing**

**A Monthly Journal of Computer Science and Information Technology**

**ISSN 2320–088X**

***IJCSMC, Vol. 3, Issue. 10, October 2014, pg.356 – 366***

**RESEARCH ARTICLE**

# CONTENT AND LEXICAL ANALYSIS: PRACTICAL APPLICATION

**Aakash Goyal[1], Ravinder Yadav[2]**
[1]CSE Department, Dronacharya College of Engineering
Delhi, India
aakashgoyal11@hotmail.co.uk
[2]CSE Department, Dronacharya College of Engineering
Gurgaon, Haryana, India
ravinderyadav2077@gmail.com

***Abstract— This paper aims at exploring the possibilities of applying qualitative content and lexical analysis: starting with general notions in case study research. First , introduction to content and lexical analysis as an interpretation method for qualitative interviews and other data material is given and technical stages of the content analysis are described, pointing out its value as a research instrument. Then case study research as a research strategy within qualitative social research is briefly presented. Then seven levels of lexical analysis are presented in a creative and evolutionary way, considering the use of computer software. The potential contribution of these methods of data analysis will be made clear.***

## 1. INTRODUCTION

*Lexical analysis* or *scanning* is the process where the stream of characters making up the source program is read from left-to-right and grouped into tokens. *Tokens* are sequences of characters with a collective meaning. There are usually only a small number of tokens for a programming language: constants (integer, double, char, string, etc.), operators (arithmetic, relational, logical), punctuation, and reserved words.

while (i > 0)
i = i - 2;
Lexical
Analyzer
error messages
source
language
token
stream
T_WHILE
T_LPAREN
T_IDENTIFIER
T_LESSTHAN

 

T_INTCONSTANT
T_RPAREN
T_IDENTIFIER
T_EQUALS
T_MINUS
T_INTCONSTANT
T_SEMICOLON

The lexical analyzer takes a source program as input, and produces a stream of tokens as output. The lexical analyzer might recognize particular instances of tokens such as: **3** or **255** for an integer constant token **"Fred"** or **"Wilma"** for a string constant token **numTickets** or **queue** for a variable token Such specific instances are called *lexemes*. A lexeme is the actual character sequence forming a token, the token is the general class that a lexeme belongs to. Some tokens have exactly one lexeme (e.g., the > character); for others, there are many lexemes (e.g., integer constants). The scanner is tasked with determining that the input stream can be divided into valid symbols in the source language, but has no smarts about which token should come where. Few errors can be detected at the lexical level alone because the scanner has a very localized view of the source program without any context. The scanner can report about characters that are not valid tokens (e.g., an illegal or unrecognized symbol) and a few other malformed entities (illegal characters within a string constant, unterminated comments, etc.) It does not look for or detect garbled sequences, tokens out of place, undeclared identifiers, misspelled keywords, mismatched types and the like. For example, the following input will not generate any errors in the lexical analysis phase, because the scanner has no concept of the appropriate arrangement of tokens for a declaration. The syntax analyzer will catch this error later in the next phase.

int a double } switch b[2] =;

Furthermore, the scanner has no idea how tokens are grouped. In the above sequence, it returns **b**, **[**, **2**, and **]** as four separate tokens, having no idea they collectively form an array access. The lexical analyzer can be a convenient place to carry out some other chores like stripping out comments and white space between tokens and perhaps even some features like macros and conditional compilation (although often these are handled by some sort of preprocessor which filters the input before the compiler runs).

## 2. A QUALITATIVE PRACTICAL APPLICATION

Document analysis, which includes content analysis and lexical analysis, follow classic methods like the judicial and sociological research. It presents a major common rational characteristic, being more or less intuitive, personal, and subjective. Like others, the historical method has validity problems, such as the authenticity of the text, interpretation validity, and the truthfulness of statements of fact. Among others, all of these analyses, have the defect of a non-systematized work, depending only on the value and competence of the researcher. With the proliferation of the means of communication (TV, radio, etc.), and especially of advertising on the Internet, some things became more obvious: the need to look for meanings, relationships, or even laws to measure influence factors; and the need to foresee certain events.

A modern method of documents analysis was born from the necessity to accomplish studies in new directions: **content analysis!**

Qualitative Data Analysis.

In Section 2, general notions of content analysis are presented, and its application and use are emphasized in Section 3. In section 4, the technical stages of content analysis are presented, also highlighting its value as a research instrument. The seven levels of lexical analysis are presented in a creative and evolutionary way in Section 5, considering the use of computer software. Having been inspired by qualitative data collected in research, some of the analysis techniques are illustrated in Section 6. The usefulness of, and interest in, content analysis in the context of qualitative research is presented in Section 7

## 3. General Notions about Content Analysis

In this section, the definition of content analysis is described, as well as its characteristics, its sources, types, and trends.

**1.1 - Definitions of Content Analysis**

The resources that the social sciences offer for our reflection are essentially composed of oral and written communications. Oral communications include text of speeches, annotations and reports of interviews, and conversations. Written communications include official texts, newspapers articles, letters, everything in the literary field, even history and politics. During the course of research, we gather data to understand, to explain opinions, conduct, or actions. These data are almost always of verbal origin. The actions, when we apprehended them came from a context of words. Speeches and addresses are documented in words.

 

From our point of view, it is very important that the researcher in social sciences be able to analyze these data in a scientific way and not be satisfied to have a casual impression. The novelty of the modern procedure of content analysis consists in substituting the impression (a personal opinion, almost a guess), for more standardized procedures, quantification, in every case transforming raw documents into data for scientific analysis. For such task, the text will be decomposed; that is, it will be studied as a function of the words that it contains or the ideas it represents, the latter being chosen for reason of its relationship with the objective of our research (for example, "How many times did a political candidate use the term *'social justice'* during the electoral campaign?").

Berelson (apud Grawitz, 1993, p.534) defines content analysis as: "… a research technique for the *objective, systematic and quantitative* description of the obvious content of the communications, having for objective to interpret them". In a general way, Grawitz (1976) complements this definition saying that we should consider dispensable the term "quantitative". We would add to her comments that this "quantification" may be an easy way to compute and that we should seriously consider this possibility. We need to spend some time exploring the contents exactly to better understand what is "not so obvious".

**1.2 - Qualities and Sources of Use**

Grawitz (1993, p.534) describes the characteristics which Berelson attributes to "objective, systematic, and quantitative" as following:

· **Objective:** The analysis should proceed according to the pre-established rules, obeying guidelines sufficiently clear and precise in a way that allows different analysts, working on the same content, to obtain the same results. This means, even if the analysts are in different places, at different times, and on different research teams, they agree on the aspects to analyze, the categories to use, and the operational definition of each category.

**Systematic:** The whole content should be ordered and integrated in the chosen categories as a function of the objective pursued. Elements of information related to the objective should not be overlooked.

**Quantitative:** It is to evidence the significant elements, to calculate its frequency, etc. This condition is dispensable; since when using qualitative data analysis certain analysts look more for themes than for exact measures of importance.

Content analysis can be a good technique to use in all information systems research, which is documented in written texts (official documents, books, newspapers, personal documents), in voice or image recordings (radio, television, etc.); or in other activities, that could be decomposed, such as a management meeting, or the use of any professional's time. All material especially generated for psyco-social research (group discussions, depth interviews, and meeting reports, etc.) can also be subjected to content analysis. We have not yet found the limits to the application of content analysis and related software. The Internet provides a wealth of "free" data for researchers and curious data analysts to conduct a variety of investigations, which could generate unique and powerful information and could even lead to useful and rich conclusions.

**1.3 - Types of Content Analysis**

Since different sources of data are available, we can use a lot of models to explore, process and analyze data. Grawitz (1993, p.536-538) develops three different approaches: exploration versus verification; quantitative versus qualitative; and direct versus indirect.

**Verification Versus Exploration**

Grawitz (1993, p.536) insists that, in the social sciences, a clear distinction should be made between document analysis having as an objective hypothesis verification and that analysis whose objective is exploration or hypotheses definition. In hypothesis verification, the objective is precise and the results can be quantified. In hypothesis definition, the analysis is less rigorous and systematic, following rules and techniques which cannot be standardized, and it is based intuition and experience. Grawitz observes that when we want to classify certain relevant elements in a text or document in order to quantify them, we must know whether we are trying to elaborate or to follow our intuition to the detriment of any and all systematization. When there is a great mass of data, and we do not have hypotheses or clear research questions to serve as a guideline in the exploration of the text, the analysis should be conducted knowing that maybe we will be leaving essential (and unexpected) elements outside the scope of the study.

**Quantitative Analysis Versus Qualitative Analysis**

According to George (apud Grawitz, 1976, p. 592), an important difference is that quantitative analysis try to accumulate the frequency of themes, words, or symbols, while qualitative analysis is based on the presence or absence of a given characteristic. Other qualitative-quantitative dichotomies: impressions versus systematization, hypothesis instead of verification, and flexibility as oppossed to rigidity. What is important should be clear in each type of analyses. The "number of times" a data element occurs is what counts for the quantitative analysis, while a novelty, interest, theme, or subjective attribute is the object of qualitative analysis. We always have a dilemma: to adopt a few representative categories, or to regroup deliberately the data in a small number of categories, but sacrificing information.

 

## 4. Content Analysis Application and Use

We can apply this kind of data analysis technique in several ways, which can involve different kinds of source, storage or media: oral communications, like TV, radio, discussions or debates, meetings, etc., or written communications, newspapers, books, papers, interviews, questionnaires, etc. Lets try to better understand this communication factor and also the field of application regarding this technique.

Concerning any communication, we will be in presence of an originator or emitter that throws a message possessing content and form, seeking to reach an objective, and addressed to one or several receivers. Who speaks? To say what? To whom? How? With what result? This is a simple outline proposed by Lasswell (apud Grawitz, 1976, p. 598) for the communication, which seems useful for the content analysis, the subject proposals covering again the group of problems.

**Who speaks?** ... or the study of the emitter (originator), where two situations are possible: the emitter reacts more or less to an incentive controlled by the observer, as in a guided or driven interview; or, more frequently, the analyst does not have any relationship with the emitter, and he looks at that based on the content of the emitted informations. Maybe we could want to identify the author or his characteristics, starting from the text-source.

**To say what?** It is to study the characteristics of the message content, to recognize the nuances and different directions of the content, to compare themes, speeches, slogans, or even the evolution of the author's texts. Frequently, it is about to compare materials of different texts. The content analysis contributes strongly to the adapted elaboration of an interview, facilitating the code of the subjects of open-ended questions, or the reports of a meeting or group discussion, facilitating a systematization.

**To whom?**... The receiver is determined, to who goes the message; could be interesting to identify even what is latent and not only clearly explicit (the study of a message of such leader is addressed to his own organization, but in fact he wants to pass a message to the competitors).

**How?** It is the study of the way which the text looks for producing an impression or message. **It is essentially something qualitative and subjective, that we take a look under a quantitative point of view.** The focus are the elements that contributes to produce this impression: choice of words, repetitions, composition of the sentence, etc. The categories will always be qualitative, but the analyst will quantify the associated data.

**With what result?** The point is to know the effect of the message on the receiver, or we want, with the right elements and in the best way, to influence a given public in a certain sense; or still, whether the strategies of the originator are clear, to foresee what he looks for and his implicit objectives. That is, sometimes it is explanatory or elucidative, some times is more preditive (or indicative).

The multiplication of the studies about communications, with the search of identification and understanding processes (besides the analysis of the explicit or "ordinary"), they took to consider in the analysis, besides the subjects above, the influence of the environment, even on the emitter as on the receiver, to measure the effect of a private communication. The content analysis can lift or same to solve great problems in the sociology area and social psychology (Grawitz, 1976, p. 603). As well as in different areas of management, as we can see in many of the applied references at the end of this paper.

## 5. Content Analysis Technical Stages and Value

As in almost every research, the first step is the idea itself, as well as its objective. The means of obtaining an answer is to ask a question, as in an interview. As the questionnaire allows us to guide an interview, the research analyst will select the categories to guide the analysis based on the data in the documents. Grawitz (1993, p.543-558) describes the main technical stages of content analysis, pointing out that we first need to choose categories, and then we need to start the main analysis, and evaluate content analysis as a research method (Sections 4.1, 4.2 and 4.3, respectively).

**3.1 - Choosing the Categories**

**Definition, Creation or Formulation:** The categories are the basis on which the content will be classified and eventually quantified. In the case of a survey, the categories can already be foreseen in code form. The content analysis should usually allow the appearing of variables and influence factors that we ignored in the beginning of the work. The categories, when we do not have a precise idea, should be defined based on the content. In an exploratory survey, the categories constitute themselves the object or context of analysis allowing us to select the data in the text; while in search of hypothesis verification, we usually have pre-defined categories. Grawitz (1976, p. 605) advises that we should "prolong the period of tests and establish several categorization systems or code before adopting an outline (a structure or system) of categories".

**Characteristics:** The choice of categories is the most important step in the content analysis; the categories are the connection between the research objectives and the results of the content analysis. The value of the analysis is contingent upon the legitimacy of the categories. The objectives of the research will guide the selection of what will be quantified. The categories should be exhaustive, mutually exclusive, objective, and pertinent. Exhaustive means the categories should include as much as possible of the text, if we put all categories together, we can make a judgement regarding the whole text, or we can have an idea concerning the whole text. Mutually exclusive means the same elements cannot belong to several categories. Objective means that the categories will have such clear characteristics that different analysts will obtain the same results with the same text. Pertinent means the categories relate the content of the text to the objectives of the research.

 

**Excesses to avoid:** One excess to avoid is the imposition an overly rigid outline for the analysis from the start that will prevent the comprehension of the complexity of the content. A second excess to avoid is preparation or definition of the outline in a superficial way, focusing on the explicit and ignoring the latent content. A third excess to avoid is choosing too detailed or too numerous categories, almost reproducing the text in order not to lose anything. The fourth excess to avoid is establishing categories too broad to let us sufficiently distinguish the elements. The goal is to approach an optimum number of categories or an optimum percent of the whole.

**Standardization:** It may be helpful to adopt a certain number of categories applicable to several cases; for example, age, gender, profession, religion, nationality, highest level of education. However, there is no algorithm for choosing categories that is appropriate for all cases. Researchers in some academic disciplines have adopted standards; such as, the marketing research under guidance from the American Marketing Association (AMA) and the European Association with the same purpose (ESOMAR). It is unlikely that two different analysts using the same content with the same starting point will obtain the same result. To be satisfactory, the outline of analysis should follow a general conceptual scheme accepted by peers. A standard illudes us because the specific content changes with each new research study.

**3.2 - Problems with Quantification**

Ambiguity, complexity, and variety of sources are problematic for content analysis.

**The sample problem:** These problems do not exist in the cases we studied. However, when the object of analysis is a text in its totality, or a series of speeches, when it is something as vast as several years of a newspaper, or as varied as advertising, we should define the research by choosing a certain type, a certain topic, or a certain section (only the editorials of a newspaper, for example). Or, then it cannot, instead of limiting the theme, to limit a sample that is representative of the whole. Rules to select a sample: (1) Define the universe to which the generalization is to be applied; (2) Make sure that each unit of this universe has a well-known probability of being contained in the sample; (3) Select a sample that is independent of all correlation among the units of the universe; and (4) Select a large enough sample to minimize sampling error.

**3.3 - The Value of Content Analysis as a Research Instrument**

Content analysis is a refined technique, therefore delicate, and it demands much dedication, patience, and time to satisfy the investigator's curiosity, in addition to intuition, imagination to notice what is important, and creativity to choose the categories. At the same time, the analyst should have discipline and perseverance, rigidity when decomposing content or to count results of analyses. These are only some of the main aspects that Grawitz (1993, p. 553-558) points out, along with fidelity, logical validity, inference, and empirical validity.

**Fidelity:** By definition, content analysis should be objective, and the results should be independent of the measurement instrument, being convenient to minimize the differences in points of view among the analysts. But this is an old problem in the social sciences, the fidelity cannot be faced in the same way according to it is a quantitative analysis of a clear and obvious content or if it is an analysis more qualitative where we look for to identify latent intentions, where it is worth more the presence or absence of an element and not its frequency. In this case, it will be driven to minimize the subject of the fidelity. In the quantitative analysis, the analyst will treat at first the frequency of the elements, not mattering so much with the validity of isolated data.

**Logical validity:** Does the instrument measure what it intends to measure? An analysis is valid when the quantified description that it offers concerning the content (object of the study) is significant for original problem and it reproduces faithfully the reality of the facts that it represents.. Of course, it is essential condition of the representativity of the sample and that the inherent technical conditions to each stage are observed satisfactorily. The more limited or defined the objective, the easier it will be to confirm the validity; and it is easier to demonstrate validity in a quantitative analysis than in a qualitative one.

**The inference:** This point deserves special attention because sometimes one expression has more than one interpretation, even positive and negative interpretations, depending on its context.

**Empirical Validity:** Is the study prediction just or precise? A difficult question to answer. Instead of convictions, prudence and humility are recommended when drafting the conclusions. However, Grawitz believes that the experience and the analyst's training will countersign the value of its analyses.

## 6. Placing in evidence certain potentialities, applications and limits of the textual data analysis tools

Our aim is to evaluate the usefulness of content analysis and lexical analysis. For example, content analysis can be used to analyze in depth each person's or group's specific expression involved in a debate. The work on an organized "lemmatized" text allows us to use grammatical categories -- nouns, verbs, adjectives—to organize the first impression of the content of the text.

The "lemmatizor" or "lemmatizator" is an English translation of the French "lemmatiseur" (used by Professor Jean Moscarola and his research team from Université de Savoie, France) representing a computerized tool that aids the analyst

 

marking in the text nouns, verbs, adjectives, names, etc. in the text. We will use the term text "lemmatized" to describe the text that was already processed with the "lemmatizator". The banality (or triviality) of lexical indicators can be shown in the specificity that guides the selection of the most trivial sentences, or the most significant or original words, from any group or category. With these resources, we can also build *ad hoc* indicators to evaluate the degree of commitment regarding the actors or members of discussion groups.

**Seven Levels of Lexical Analysis:**

**Level 1: The summary lexicon approach.** Reduce the body of the original text to the "high" or top of the lexicon (for example the top 25), which means that examining only the most frequently occurring words we can have us an idea of its content. By lexicon we mean the list of all words of the text used with the number of times that each occur.

Body of the text
Lexicon
Figure 1. The Summary Lexicon Approach

**Level 2: The controlled lexicon approach.** Reduce the text to its lexicon (each word of the text and its frequency), and to control, through lexicon surfing or navigation, the validity and the foundations of the interpretations elaborated, starting from the lexicon. See Figure 2. By "verbatim" we mean that we can, if needed, export or transfer some of the contents to a report or file in order to use them to illustrate something. For example, export to a text file all comments, or opinions, or phrases that contain the word "love". It depends on the analysis unit.

Body of the text
Lexicon
**Lexicon**
**surfing** Verbatim

**Level 3: The Selective Lexicon Approach.** Work on a reduced lexicon after having eliminated or deleted the tool words (definition of those words without a role or meaning something useful, all that we can ignore, like prepositions and articles, etc, the whole stored in a specific dictionary), concentrating then our attention on nouns, verbs, adjectives; everything by use of dictionaries and the "lemmatizator". We can also create dictionaries with the words that we have a special interest (like positive or negative arguments, or words regarding a special field, like health, or other). We can also group words in a new word or concept (with a new name to the group or even with the name of the most frequent word).

**Body of the text**
**Lexicon**
**surfing**
Verbatim
Lemmatizator
Body of the text
lemmatized
Dictionary
Structures
Statistical
**Reduced and structured lexicon**
Figure 3. The Selective Lexicon Approach

**Level 4: Lexicon Statistics And Text Quantification.** The Research of the lexicon characteristics, to establish the statistics of the words of the text according to an external non-textual variable, like..."department" (marketing, human resources, research and development, others), gender (male / female), age (under 20, 20 and over), or level of education (graduate, undergraduate), for example.

**Body of the text**
surfing
Lexicon
Reduced and structured lexicon
Lemmatizator
Body of the text
lemmatized
Dictionary

 

structures
Statistical
**chart or table**
**Lexicon**
**Context**
**variable**
Figure 4. Looking for Lexicon Characteristics

**Level 5: Codifying the Lexicon and Generating Measures**. Describe the text by a nominal variable, created by a codification based on the presence in the answers of a group of selected among the words from the lexicon. We can do this kind of procedure many times, each time with a different purpose, each time generating new variables, closed this time (and not open-ended), which we can also cross with others like gender, age, etc., in a bi-variate analysis procedure.

**Level 6: Quantification of text**. Calculate the number of times certain words occur, starting from a "lemmatized" text and already with the tool words taken out of the lexicon (by taking out of the text the words that do not interest us using dictionaires, for such a task using special user-friendly procedures or "clicks"). It can allow us to know if certain opinion is "unique" or if it is a regular or common one.

**Body of**
**The text**
surfing
Lexicon
Reduced and
structured
lexicon
Lemmatizator
Body of the text
lemmatized
Dictionary
structures
Statistical
Verbatim
**Lexical**
**Measures:**
**intensity**
**banality**
**Closed**
**variables**
**on the**
**lexicon**
**words**
Figure 5. The Lexicon Codes and Measures, and the Text Quantification

**Level 7: The Multivariate Analysis of the Textual Data.** Data analysis made on the variables extracted from the lexicon. We can apply the methods of multivariate analysis using the new variables (lexicon tables): factor analysis or automatic classification can be done. We can even consider the variables in relation to other contextual variables (like gender, education, age , etc), and thus integrate classical data analysis with the textual data analysis.

**Body of the text**
surfing
Lexicon
Reduced and
structured
lexicon
Lemmatizator
Body of the text
lemmatized
Dictionary
structures
Statistical
**Context**
**variable**
Verbatim
**Lexical**
**measures:**
**intensity**
**banality**
**Closed**
**variables**
**on the**

 

**lexicon…**

Figure 6. Data Analysis of Variables Extracted from the Lexicon.

The access and the "surfing" regarding the text or document data or the qualitative research methods is then facilitated. The tools are now powerful. However, the analysis of textual data does not change the meaning of the data. Even if analysis reduces the "noise" in the data, the reduction of the long, and sometimes annoying readings, the faster production of conclusions and reports drives everybody to an *interpretive* reading.

*Interpretation* is often dangerous since it is fast and it appears to be falsely objective. We should so be careful in our investigation and conclusions. That is a Moscarola's observation (1995).

**Data Mining with Speech Qualitative data: The Election Debate "Dole versus Clinton"**

This is a very interesting field that will continue to be developed for far beyond what it is today. Using the Internet, we can very quickly download the speeches from a political, social, or whatever discussion, and make many different lexical analysis. Gavard-Perret and Moscarola (1995) analyze the Bush-Clinton USA presidential candidates debate. Later, Moscarola also explored some data regarding the Dole and Clinton elections. We have at least two possible applications of content analysis and lexical analysis in this cases: the content itself or the main ideas, and some statistics based on the texts, or the speech-acts. We can know very easily how many words each one spoke, what was the average length in words of each speech or contribution in the discussion, whether they repeated each one of their main arguments, whether they said something "unique", whether they were "original" in their speech, etc. We can even look only for the verbs, meaning the actions . same thing regarding adjectives or nouns. We will then be able to identify the objects or figures of each speaker, their main keywords, their main messages to the people, their main ideas concerning some important subjects like education, social security, health, etc. and then compare them this way!

Are they really discussing something? Are they only playing a game where they have known very well what they would like to say long before the debate, which means were they ignoring their opponent during the discussion? We can find answers to these questions by making lexical analysis of their speeches! For example, if we divide the Dole-Clinton speech data into five equal parts, and cross it with the main arguments used by both candidates, we have the illustration in Figure 13.

Axis 1 (33.8%)
Axis 2 (29.0%)
Senator_Dole
Mr_President
tax_cut
American_people
health_care
United_States
Brady_Bill
Thank_you
voted_against
assault_weapons
balanced_budget
tax_increase
worked_hard
21st_century
Four_years_ago
trial_lawyers
economic_package
school_choice
Social_Security
young_people
better_off_than_we_were_four_years_ago
crime_bill
Drug_use
Middle_East
people_watching
welfare_reform
balance_the_budget
federal_government
foreign_policy
public_schools
Saddam_Hussein
United_Nations
World_War
ONE
TWO

 

THREE
FOUR
FIVE

## 7. Conclusion: The Usefulness of Content Analysis

Due to its complexity, it should be investigated in which cases content analysis or lexical analysis should be applied. According to Berelson (in Grawitz, 1976, p. 623), it should be applied to all cases that require great precision and objectivity. The comparisons and the evolutions form the main content analysis field. When sufficiently defined and detailed, content analysis allows us to pass of the simple description and to reach the objective of every scientific research: the discovery of explanations and causal relationships.

The value of content analysis depends on the quality of the conceptual elaboration done a *priori* by the researcher, of the *exactitude* with which it will translate itself in variables, of the analysis outline or categories, and also of the agreement among the reality to analyze and these categories. So that such analysis deserves to be accomplished, the subjects guided by the categories should show an interesting hypothesis and they should correspond to the selected material. Anyway, it is time for "multiple methods" (Sawyer et al, 1997), it is more and more clear that we need be in touch with real situations using real data and looking for "Rurr" (really useful rigorous research), following Brown et al (1997) in order to build better relationships. It is also true that we look for relevance in our studies, as Mandviwalla and Gray discussed in a very recent paper published in the IRMJ (Information Resources Management Journal, v. 11, nr. 1, Winter 1998). In this paper, other authors share important points of view like: Beyond rigor and relevance (Robey, D. and Markus, L.), The application of IT research to organizations (Kavan, C.B.), and The challenge of relating IS research to practice (Senn, J.). Overall, it is time to go ahead with many more qualitative studies, and to educate our managers, starting with our children, that the world is not only quantitative, but also qualitative. At least, what good quantitative study could not have been preceded by a qualitative one? Creswell (1998), and also Kirk and Miller (1986) offer us some concepts and discussions concerning qualitative research and mainly about reliability and validity in this kind of study. By the way, the the MIS Quarterly (v. 21, no. 3, September 1997) published three papers based on qualitative research in the same issue. Even though it was authored by Richard O. Mason, James L. McKenney, and Duncan G. Copeland, the fact that we had three of them in the same issue is a positive sign from the editors. Mason (1997) offers us some good insights to start discussing "What is qualitative research?" and mainly "What should qualitative research be?" Like she said "Asking difficult questions". It should be: "systematically and rigorously conducted, strategically conducted, yet flexible and contextual, should involve critical self-scrutiny by the researcher, or active reflexivity, should produce social explanations to intellectual puzzles, should produce social explanations which are generalizable in some way, or which have a wider resonance, should not be seen as a unified body of philosophy and practice, whose methods can simply be combined unproblematically, should not be seen as necessarily in opposition to, and uncomplementary to, quantitative research, should be conducted as an ethical practice, and with regard to its political context".

Mason (1997) discuss five main questions regarding qualitative research which everybody should ask:

- What is the nature of the phenomena, or entities, or social 'reality', which I wish to investigate?
- What might represent knowledge or evidence of the entities or social 'reality' which I wish toinvestigate?
- What topic, or broad substantive area, is the research concerned with?
- What is the intellectual puzzle? What do I wish to explain? What are my research questions?
- What is the purpose of my research? What am I doing it for?

The use of techniques of qualitative data analysis and research is a latent theme in information science and scientific management communities (Willcocks et al, 1997; Kendall et al, 1997). We hope that this document constitutes a contribution to rescue some of its concepts and techniques, and to illustrate some of its potential applications.

## References

Bachelet, C., and Moscarola, J. La communication interne: diagnostic par l'analyse du discours des acteurs. Annecy/France: IUP/GEREG, Université de Savoie, *Cahier du GEREG*, 1995, 10p.

Bardin, L. L'analyse de contenu. Paris: PUF, 8. Ed.,1996, 291 p.

Benzécri, J. P. L'Analyse des Données, v. II, L'Analyse des Correspondances, Paris: Dunod, 1973.

Brown, C. V., Markus, L., Rockart, J.F., Sambamurthy, V., and Shrednick, H. Building relationships with practice with RURR (Really Useful Rigorous Research). Atlanta (USA): ICIS 97, panel, December1997.

Burt, C. The factorial analysis of qualitative data. British Journal of Psychology (Statistical), 3, November 1950, p.166-185.

Carrol, J. D., Green, P. E., and Schaffer, C. Interpoint Distances Comparisons in Correspondence Analysis. Journal of Marketing Research, v. 23, August 1986, p. 271-280.

Carrol, J. D., Green, P. E., and Schaffer, C. Comparing Interpoint Distances Comparisons in Correspondence Analysis: a Clarification. Journal of Marketing Research, v. 24, Nov. 1987, p. 445-450.

Cheung, Y.L. Categorical Criteria Values: Correspondence Analysis, Omega, n° 7, July 1994, p. 371-380.

Creswell, J. W. Qualitative inquiry and research design: choosing among five traditions. Sage, 1998, 403 p.

 

Crouch, A., and Basch, J. The structure of managerial thinking: a lexical and content analysis. Journal of applied management studies, v. 6, no. 1, June 1997, p. 13-34.
Cunha Jr., M. Análise Multidimensional de Dados Categóricos: A Aplicação das Análises de Correspondência Simples e Múltipla em Marketing e Sua Integração com Técnicas de Análise de Dados Quantitativos. Porto Alegre, Brazil: Cadernos de Estudos do PPGA, UFRGS, 16/97, 1997, 26 p.
Cunha Jr., M., Freitas, H., and Slongo, L. A. A pesquisa de marketing como fator de interação universidadeempresa: estudo de caso aplicado utilizando uma ferramenta (Sphinx ®) de estruturação e de análise quantitativa e qualitativa de dados. João Pessoa/PB Brazil: Anais do 19o. ENANPAD, v. 1, no. 5,Marketing, September 1995, p.109-129.
Domenjoz, P., Gavard-Perret, M. L., and Moscarola, J. Price and communication: How do they interrelate?
An analysis of car advertisements published in English. France: Actes du congrès de la AFM (Association Française du Marketing), 1995.
Fox, R. J. Perceptual Mapping Using the Basic Structure Matrix Decomposition, Journal of the Academy of Marketing Science, 16, 1988, p. 47-59.
Freitas, H., Cunha Jr., M., and Moscarola, J. Aplicação de sistema de software para auxílio na análise de conteúdo. São Paulo: RAUSP, v. 32, no. 3, July 1997, p. 97-109.
Gavard-Perret, M. L., and Moscarola, J. De l'énoncé à l'énonciation: pour une relecture de l'analyse lexicon en marketing. Annecy/France: IUP/GEREG e CERIAM, Université de Savoie, Cahier du GEREG, 1995, 34 p.
Gavard-Perret, M. L., and Moscarola, J. From the utterance to the enunciation, for a re-reading of lexical analysis in marketing. France: 1st. International Research Seminar, Marketing Communication and Consumer Behaviour, 1995, 21 p.
Grawitz, M. Méthodologie des Sciences Sociales. Paris: Dalloz, 1976 (chap. II, p. 586-627), and 1993 (9ed., chap. 2, section 2, p. 532-565.
Greenacre, M. J. Theory and Application of Correspondence Analysis. London: Academic Press, 1984.
Greenacre, M. J., and Hastie, T. J. The Geometric Interpretation of Correspondence Analysis, Journal of The American Statistical Association, 82, June 1987, p. 437-47.
Hair et al, Multivariate Data Analysis. Prentice Hall, 1994.
Hoffman, D. L., and Franke, G.R. Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research. Journal of Marketing Research, v. 23, August 1986, p. 213-27.
Hoffman, D. L., De Leeuw, J., and Arjunji, R. V. Multiple Correspondence Analysis. Advanced Methods of Marketing Research. Bagozzi, R. P., 1994, p. 260-94.
Hofstede, G. Cultures and organizations: software of the mind. London: McGraw-Hill, 1991. 279p.
IRMJ Information Resources Management Journal, (11, 1) Winter 1998 (Beyond rigor and relevance.)
Robey, D. and Markus, L.; the application of IT research to organizations, Kavan, C.B.; The challenge of relating IS research to practice, Senn, J.).
Kendall, K., Trauth, E., Brown, M., and Boudreau, M. C. A report from 8.2 to ICIS: Three points of view on IS and qualitative research. Atlanta (USA): ICIS 97, panel, December 1997.
Kirk, J., and Miller, M. L. Reliability and validity in qualitative research. Sage, 1986, 87 p.
Krippendorff, K. Content analysis: an introduction to its methodology. The Sage CommText Series, 1980, 191 p.
Lagarde, J. Initiation à L'analyse des données, Paris: Dunod, 1995.
Lebart, L.; and Salem, A. Statistique Textuelle. Paris: Dunod, 1994, 342 p.
Lebart, L., Morineau, A., and Warwick, K. M. Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques. New York: John Wiley & Sons, Inc., 1984.
Macadar, M.; Costa, A.; Freitas, H.; Becker, J. L., and Moscarola, J. A concepção e validação de um conjunto de instrumentos quanti-qualitativos para estudar o processo decisório segundo as culturas nacionais e os níveis de experiência decisória. Angra dos Reis-Brazil: 21st ENANPAD, Organizations, September 1997, 15 p. (A quanti-qualitative method to study decision-making process).
Mandviwalla, M., and Gray, P. Is IS research on GSS relevant? Information Resources Management Journal, v. 11, nr. 1, Winter 1998, 29-37.
Mason, J. Qualitative researching. Sage, 1997, 180 p.
Miles, M. B. & Huberman, A. M. Qualitative data analysis. Sage Publ., 1994, 338 p.
MIS Quarterly v. 21, number 3, September 1997, (Three papers on qualitative research, from Richard O. Mason, James L. McKenney and Duncan G. Copeland). Moscarola, J. Enquêtes et analyse de données. Paris: Vuibert, Gestion, 1991.
Moscarola, J. Balladur, Chirac, Jospin, les mots d'une campagne: quelques exemples d'analyse lexicone avec le Sphinx. Annecy / France: IUP, cahier du GEREG, Université de Savoie, 1995, 8 p.
Moscarola, J. Analyse de contenu et analyse de données: solutions logiciels pour une intégration quali/quanti.France: Actes JADT, 1993.
Moscarola, J. Les actes de langage: protocoles d'enquêtes et analyse des données textuelles. Paris:Colloque Consensus Ex-Machina, La Sorbonne, April 1994.
Moscarola, J. La communication politique vue par l'analyse lexicale. Annecy / France: IUP, Université de Savoie, Cahier du GEREG, n° 32, 1994.

 

Oswick, C., Keenoy, T. and Grant, D. Managerial discourses: words speak louder than actions? In: Journal of applied management studies, v. 6, nr. 1, June 1997, p. 5-12.

Ruas, R.L., Pires, M., Freitas, H., Antunes, E., Cunha Jr., M. and team. Research report "Indicadores da qualidade e produtividade". Partnership PPGA/UFRGS, IEL/FIERGS, SEBRAE, Executive report, Dec. 1994, 68 p.

Sawyer, S., Cooprider, J., Galliers, R., Gallivan, M., and Kaplan, B. Using multiple methods in studying IS: some experiences. Atlanta (USA): ICIS 97, panel, December 1997.

Silverman, D. Interpreting qualitative data: methods for analysing talk, text and interaction. Sage Publ, 1993, 224 p.

Simon, H. A. et al. Decision making and problem solving. Interfaces, v. 17, n.5, Sept. 1987, p. 11-31.

Simon, H. A. Administrative behavior: a study of decision-making processes in administrative organizations. New York: MacMillan, 1947. 259p. and The Free Press, 4th. Edition, 1997.

SPHINX LEXICA (v. 2.0). Sphinx User's Guide. Annecy / France: Le Sphinx Développement, 1995,1997, and Porto Alegre, Brazil, 1997 (e-mail: f&c_cons@portoweb.com.br -http://www.freitasecunha.com.br)

Strutton, D., and Pelton, L. E. A Multiple Correspondence Analysis of Telephone Contact Rates. The Mid-Atlantic Journal of Business, 30, no. 1, March 1994, p.27-40.

Stumpf, M. K., and Freitas, H. In search of the definition of the "patient colors record" of the Clinics Hospital of Porto Alegre. Angra dos Reis/Brazil: Proceedings of Annual Meeting of ANPAD, 1996, Sept. 1996.

Thurow, L.C. "Mapping the Economic Surface of the Earth", "Plate Four: A Global Economy," and "The Forces Remaking the Economic Surface of the Earth" ... The Future of Capitalism. New York: William Morrow & Co., 1996.

Weber, R. P. Basic content analysis. Sage Publ., 2nd. Ed., 1990, 96 p.

Willcocks, L., Dubé, L., Paré, G., and Robey, D. Pointers in doing and writing up case research. Atlanta (USA): ICIS 97, panel, December 1997.

 