The Quantitative Content Processing Methodology: Coding of Narratives and Their Statistical Analysis

Sigitas Vaitkevicius, Lina Kazokiene

Kaunas University of Technology
K. Donelaicio st. 73, LT-44029, Kaunas, Lithuania
e-mail: sigitas.vaitkevicius@ktu.lt

Kaunas University of Technology Panevezys Institute
Nemuno st. 33, LT-37164, Panevezys, Lithuania
e-mail: lina.kazokiene@ktu.lt

crossref http://dx.doi.org/10.5755/j01.ec.24.1.2350

The paper deals with the quantitative analysis of the qualitative content, which enables converting the variables used in qualitative research into quantitative variables and comparing them to variables used in quantitative research. This methodology is based on the combination of Romanticist, Phenomenological, Dialectical, Critical, and Post-structural hermeneutic systems of interpretation described by Demeterio III (2001). The specially designed content systematization technique, which describes the logical sequence of content construct identification and is based on Wilkinson’s (1991) Key and Hook technique and Merkys’s (2005) operationalization, was developed by using the dichotomous coding technique. The paper describes the methodology and the areas of its application. It also gives a practical example of its application in a research study.

An integrated testing and development of the methodology was performed by investigating evaluation efficiency of business public relations. The paper introduces the context and scope of the applied procedure in the research. It also presents the data processing steps, the performed analysis and the key results. The presented research in this paper is discussed in three phases: input, implementation and impact assessment phases, where fragments of each are provided. The article focuses on the narrative content.

Keywords: research methodology, quantitative and qualitative variables, comparability of qualitative and quantitative information, qualitative and quantitative data.

Introduction

Currently a social research methodology is so advanced that can cover the entire research process starting from the identification of phenomenon itself and concluding with the investigation of the spread of typological features in the whole population (Bryman, 2007).

Qualitative and quantitative cognition technology allows to study the separate identities of social subject and to gain new sociological knowledge in formalized and transformed way by the latest modern mathematics and computing technique achievements (Leitch, 2007).

Depending on different methodological approaches, application traditions and possibilities that are determined by problematic situation and the nature of the phenomenon under study the way of cognition can be both qualitative and quantitative (Creswell & Garrett, 2008; Vallejo-Martos, 2008, Carraher et al., 2009, Ciabuschi & Martin, 2012). As evidenced by research practice and theory these methods effectively complement each other and in order to get to know the research object properly their interaction gives more comprehensive, objective and reliable results (Denscombe, 2008; Freshwater, 2007; Greene, 2007; Creswell et al., 2003) though sometimes these groups are shown as opposing and incompatible but in scientific literature there are attempts to look for these different methodological approaches coordination opportunities (Reichard & Rallis, 1994; Tashakkori & Teddlie, 1998; Moon & Moon, 2004; Johnson & Onwuegbuzie, 2004).

Currently MIX – method direction gains popularity and treats traditional research approaches as an integrated methodology of quantitative and qualitative research (Denscombe, 2008; Freshwater, 2007; Greene, 2007; Creswell et al., 2003). (Plano et al., 2007; Johnson et al., 2007; Creswell et al., 2011) describe the mixed methods research as a research design with philosophical assumptions as well as methods of inquiry, where:

- As a methodology, it involves philosophical assumptions that guide the direction of the collection and analysis of data and the mixture of qualitative and quantitative approaches in many phases in the research process;

- As a method, it focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of quantitative and qualitative approaches in combination provides better understanding of research problems than either approach alone.

Analyzing application examples of the qualitative and quantitative research approaches (Chipunza & Gwarinda, 2010; Gollic & Davis, 2012), it was noticed that their inherent properties condition the application consistency:
a) Identification of phenomena. Fact-finding and data analysis according to the grounding theory.

b) Qualitative research. Description of the identified phenomenon and identification of its typological features.

c) Quantitative research. Identifying occurrence of the phenomenon and distribution of its typological features in the population.

Such an application seems completely rational because in the process of social cognition most often it is impossible to perform all the process contemporaneously. Due to big data volume and big quantity of carried actions it is likely that the research under all conditions will have to be carried out according to certain logical consistency therefore the researcher could potentially face some problems of data comparability.

We can distinguish comparison issues of ontological and methodological qualitative and quantitative research data. Ontological investigation problem is determined by complex, multidimensional, contextual, stochastic and constructional psychosocial analysis of social reality. This means that if a researcher performs a full-length research it can take several years to accomplish it starting with the identification of the phenomenon and concluding with the description of its typological features present in the population. Another situation, which calls for alternative solutions, is comparison of data gathered during qualitative and quantitative parts of the study. Such situations are examined in the works of (Morse, 1991; Greene et al., 1997; Creswell et al., 2003; Eliot, 2005; Ivanka et al., 2006; Bryman, 2006; Creswell et al., 2007; Teddie et al., 2009).

During the research studies qualitative and quantitative data more often are collected relatively separately and the results are usually compared at a qualitative level. Of course, one can use various verbal content quantification programs, such as T-lab, NVivo, Kokybis etc. However they are not powerful enough to quantify qualitative content from scientific point of view, because they create qualitative variables based on syntax of the sentence rather than on real authorial intention. As a result, even when verbal constructs have been recorded and coded in order to compare them with quantitative data, in most cases the interpretation of the psychosocial reality depends on the ability of the researcher to penetrate the phenomenon, its structure and its qualitative features.

Methodological research problem is also determined by psycho-social complexity which leads to the idea that cognition is impossible only with qualitative and quantitative means. In most cases, to understand the nature of the phenomenon completely it is not enough to employ only case study analysis, focus groups, non-structured interviews or questionnaire surveys. The limitations become especially apparent when only quantitative methods are used without combining them with qualitative research methods. Furthermore there is one more important issue that can arise due to a large number of questions, namely the learning effect on the part of respondents. When conducting various studies it was noticed that questions which respondents did not mention as important in the qualitative study, in quantitative research sometimes outweighed the ones they had mentioned because a quantitative questionnaire, due to its more comprehensive nature, tends to reflect respondent’s social opinion rather than personal one. In the case of qualitative research respondent tends to talk about issues that concern them directly, failing to identify general social stimuli and factors of social environment, which influence them indirectly with their state without a special reminder.

In order to solve this methodological problem, the attempts were to integrate open-ended questions into close-ended questionnaires. This helped to shorten the questionnaire and get some new data about the phenomenon. However, lack of methodology for comparing the results at quantitative level resulted in limited scope of the research. Such integration helped to expand the structure of attributes but don’t analyze their distribution in the population. Moreover, it was noticed that when an open-ended question expanded the scope of a close-ended question, when it was located after a set of questions it expanded upon, phrases used in close-ended questions dominated the answers of the respondents despite the fact that they had already answered such questions. This showed that while filling in the questionnaire respondents not only adapt to its content, but also start to imitate it developing an idealistic archetype of their identity. Summing up, the longer the questionnaire the greater the probability that questionnaire responses will show an idealistic model of behavior instead of a realistic one.

The research problem is a design of methodology, which would allow the using of the Hermeneutic systems for statistical coding of the authorial intentions extracted by the qualitative methods such as interviews and narratives or open-ended questions.

The aim of the paper is to familiarize the reader with the quantitative content processing methodology, the course of its compile and to give an example of its application.

This methodology must meet the following quality criteria: objectivity, representativeness, reliability, validity, efficiency and utility.

Novelty. So far, information was collected relatively separately using qualitative or quantitative cognition way and the results of these researches were compared most often at qualitative level. Prepared the quantitative content processing methodology enables to compare the information gathered in a qualitative and quantitative cognition way and to get reliable complex reality research results.

Object of the research – coding of narratives for statistical analysis.

Methods of research – theoretical modeling, quantitative and qualitative analysis of empirical data.

Description of the proposed Methodology

The main idea of the quantitative content processing methodology is that the principles used for coding verbal content can be used for statistical coding of respondents’ opinions, attitudes and feelings obtained using the hermeneutics systems described by Demeterio III, (2001). In order to ensure repeatability of the results a special content systemizing sequence was employed for coding. It was based on Wilkinson’s (1991) Key and Hook technique, which describe logical sequence of the content construct identification, and on Merkys’s (2005) operationalization technique supplemented by dichotomous coding.
Combining all the pieces into a single structure produced the following content processing and coding sequence (Figure 1):

1. Reading the text and understanding its overall meaning (based on romanticist, phenomenological, dialectical, critical, or post-structural hermeneutics systems of interpretation).
2. Identification of text’s logical structure (based on Wilkinson (1991) Key and Hook technique)
3. Dividing the text into logical passages.
4. Comparative analysis of the logical passages and their conflation into coherent logical sequence.
5. Operationalization of the logical structure:
   a. Interpretation of theoretical (extensive) concept T₁ used by the respondents;
   b. Classification of T₁ into constrictive concepts t₁, t₂, ..., tₙ according to the respondents’ intention, which have a lower degree of abstraction than T₁;
   c. Creation of t₁, t₂, ..., tₙ empirical indicators E₁, E₂, ..., Eₙ (Expressed by notions or by the logical construct)
   d. Development of measurement scale and confirmation of E₁, E₂, ..., Eₙ metrical criteria e₁, e₂, ..., eₙ.
6. Dichotomous coding of the notions, which were mentioned as 1 and those which were not mentioned as 0.

![Figure 1. The dichotomous coding model for concept T](image)

The use of the hermeneutic system for content processing enabled: first, to decode methodically the meaning of the text; second, to perform analogical actions with visual, non-verbal, emotional, audio and symbolic content (Demeterio, 2001); third, working experience with hermeneutics allowed to notice that the type of hermeneutic system not only determines content recognition, but also defines its logical structure - therefore this technique can be applied in wider research context. Finally, the hermeneutic system can be related with the typological model of the research aims. For example, in social research of individuals’ or groups’ attitudes sometimes one needs to understand the real intention of their attitudes. This aims the application of romantic hermeneutics, which is especially valued in intercultural research. Such aims can be attributed to the first group of typological models. The second group of research aims involves understanding of respondents’ or their groups’ attitudes in research context. It is related to dialectical hermeneutics and especially useful in confirmative research. The third group involves understanding the interaction of opinions that forms the images in researcher’s sub-consciousness and is related to phenomenological hermeneutic which is useful in exploratory research. The fourth group involves the aims of diagnosing hidden pathology of opinion and at the efforts to remove ideological distortions. These aims are related to critical hermeneutics and they are especially important when one needs to eliminate conformist attitudes and political ideology of the respondents. Finally, the fifth group involves the aims of understanding the principles behind formation of one’s opinion, which are closely related to sub-structural hermeneutics and especially useful in exploratory research used for analyzing the reasons behind population’s behavior and for assessment of the face validity of research tools.

Recognition of the logical structure of the text is used as a tool to ensure content validity. It enables identifying the statistical potential of the text and researcher’s notes used for recording the opinions, attitudes and feelings. The methodology is designed in such a way that after coding the text it becomes possible to measure the statistical relationship between constructs. For this purpose, Wilkinson’s (1991) method for identifying the logical structure of content is used. It enables integrating the insights gained at the stage of hermeneutic analysis into one logical structure which, depending on the hermeneutic system used, will reflect the logical setup of the text, respondents’ opinions, attitudes and feelings, or researcher’s insights.

Identification of logical passages of the text, their analysis and integration into one logical sequence is a preparatory step before operationalization. Its purpose is to identify the topics and sub-topics discussed in the text which qualitatively describe the subject of the study. The methodology is based on the view that mathematical testing is primarily a tool for statistical confirmation of the theoretical insights rather than an exploratory tool. For this reason, the structure of logical relationships has to be uncovered first and only then coded. Finally, the models of these relationships are examined.

Analysis of the logical passages of the text can be horizontal and vertical. In this methodology, vertical analysis means identifying the same logical passages in different parts of one respondent’s text and conducting their comparative analysis. Meanwhile horizontal analysis means identification and comparison of the same logical passages across the texts of several respondents.

Operationalization of logical structure is performed as a background for concept’s dichotomous coding (Figure 1). During the operationalization, original concepts are extracted up to a conditional empirical indicator, whose presence in the text is assessed based on whether it is mentioned, i.e. if it is mentioned it is coded as 1, if not – as 0. Other strategies can be applied for concept T coding. For example: when analyzing the intensity of relative use of certain concepts, the frequency of mentioning them can be coded instead of the fact of them being mentioned.

\[E₁, E₂, ..., Eₙ\]
The Context of the Proposed Methodology Application

Testing of the methodology and development of methodological background was done due to Evaluation of Effectiveness of Public Relations in Business Organizations study in the doctoral dissertation written by the author and supervised by the other author of this paper. Saving the space of this paper only an illustrative example of this methodology is presented. Overall study structure is presented in Figure 2.

<table>
<thead>
<tr>
<th>INTERVIEW</th>
<th>SURVEY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input stage</td>
<td>Public relation activities' implementation stage</td>
</tr>
<tr>
<td>Assessment of objectives</td>
<td>Public relation influence stage</td>
</tr>
<tr>
<td></td>
<td>Press release assessment criteria</td>
</tr>
<tr>
<td></td>
<td>Assessment in implementation criteria</td>
</tr>
<tr>
<td></td>
<td>Assessment criteria for expected results</td>
</tr>
<tr>
<td></td>
<td>Final results assessment criteria</td>
</tr>
<tr>
<td></td>
<td>Outcome assessment criteria</td>
</tr>
<tr>
<td></td>
<td>Productivity assessment criteria</td>
</tr>
</tbody>
</table>

Figure 2. The context and scope of the methodology application

Research study had three stages: (1) Input stage. Qualitative method was used at this stage due to its creative nature, relatively little knowledge available and the qualitative diversity of decisions that had to be taken. (2) Public relations activity implementation and (3) Public relations influence stages, due to relatively clear and sufficiently researched typology and structural simplicity, were analyzed using notions-attitudes questionnaire with close-ended answer format. This enabled to assess the applicability of this method in the research.

Dichotomous coding of interview results on the subscale for assessment of the predicted productivity of public relations (PR).

The meaning of the text was identified using the system of dialectic interpretation (Demeterio, 2001). It was chosen because of the research aims. In this case, uncovering the unified meaning of the text was less important than identifying the meaning relevant to this particular research study.

While the texts were repeatedly listened to and re-read, a kind of discussion between the researchers’ biased position, their presumptions, and the biased text with presumptions contained in it took place (Figure 2).

The reading cycle started with the projection of the researcher’s biased position and presumptions onto the text and then allowing the text to reveal its position and presumptions in this context. Such reading mode enabled to maintain the integrity of the context without losing the track of the respondents’ flow of thought. Next using Wilkinson’s (1991) Key & Hook technique the logical structure of the text was analyzed (see the right column in Table 3).

Table 1

<table>
<thead>
<tr>
<th>Logical structure and logical passages of text fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical passages based on dialectic hermeneutics</td>
</tr>
<tr>
<td>Question – What should the lower level goals cover?</td>
</tr>
<tr>
<td>Respondents audience size indicators (Fragment No. 1)</td>
</tr>
<tr>
<td>Purpose of the goal</td>
</tr>
<tr>
<td>Purposefulness of the goals</td>
</tr>
<tr>
<td>Limiting factors</td>
</tr>
<tr>
<td>Effectiveness of public relations</td>
</tr>
<tr>
<td>Respondents audience size indicators (Fragment No. 2)</td>
</tr>
</tbody>
</table>

-31-
For the analysis of logical structure the special meaning of the text was used. It was identified using the system of dialectic hermeneutic interpretation and related to the key mark (underlined and wave underlined) integrated into the question. This was the starting point in the identification of text’s logical structure. In the first question, the logical structure was related to the term “size” while in the second one it was related to the term “goal”. The analysis revealed that the logical structure of “goal” is broader than that of “size” because in this study size was one of the goals. In this way, the hierarchy of the entire content in terms of meaning was underlined too.

Table 2

Horizontal and vertical analysis of logical passages

<table>
<thead>
<tr>
<th>Fragment No. 1 (chronological sequence)</th>
<th>Determinants:</th>
<th>Fragment No. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>How big an audience I will reach having in mind that the circulation is 100 thousand. We can assume that 3-4 people have read the newspaper.</td>
<td>- Potentially reached audience or part of it o 3-4 have read o 10 or 20 per cent of it will learn</td>
<td>10 or 20 per cent of it will learn</td>
</tr>
<tr>
<td>And radio has bigger audience that means how many of people really noticed and positively remembered</td>
<td>- No. of positive, neutral, negative articles about the company. o how many... positively remembered o Potentially reached audience or part of it o how many... remembered and noticed</td>
<td></td>
</tr>
<tr>
<td>So out of these 17 million positively remembered 4 and noticed such. So... lets totalize. The mostly who speaks is... and then more precise, i.e. how many articles and in which publication...</td>
<td>- Potentially reached audience or part of it o out of these... 4... remembered and noticed o No. of positive/neutral/negative articles about a company... positively remembered 4...</td>
<td></td>
</tr>
</tbody>
</table>

The next step was identification of logical passages of the text (the left column in Table 1). The purpose of this exercise was to create logical indexes, which would enable further horizontal and vertical analysis of text fragments. Such analysis based on the “Audience size indicators” logical passages is presented in Table 2. This analysis identified eight statements falling within three logical definitions: the number of press releases, the potentially reached audience or part of it, and the number of positive / neutral / negative articles about the company. Analysis shows that although some of the responses are expressed in figures, they can only be regarded as providing qualitative rather than quantitative data. This is because when for example a respondent talks about circulation of the printed material he or she do not provide exact numerical information offering a rather general example instead. This means one cannot use the ratio scale for quantitatively defining this variable but a nominal scale can be used.

Table 3

Operationalization model for the “Productivity” concept

<table>
<thead>
<tr>
<th>Concept, T</th>
<th>Question / examples of response fragments (concepts, t₀)</th>
<th>Empirical indicators Eᵣ of concepts tᵣ</th>
<th>Eᵣ, metrological criteria, eᵣ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity action plan</td>
<td>What should the lower level goals cover? / “Potentially reached audience or part of it”</td>
<td>- No. of positive, neutral, negative articles about the company. - Topic where company is mentioned - No. of times the company was mentioned in the media. - Potentially reached audience or part of it</td>
<td>e₁ - ...will read... newspaper...; e₂ - ... plan to reach potential audience...; e₃ - ... target audiences [part of it]; e₄ - ...</td>
</tr>
<tr>
<td></td>
<td>... if the circulation is 100 thousand. We can assume that 3-4 people have read the newspaper. And radio has bigger audience that means how many of people really noticed and remembered positively. So out of 17 million positively remembered 4 and noticed such. So... lets totalize....;</td>
<td></td>
<td>Mentioning 1, Not mentioning 0</td>
</tr>
<tr>
<td></td>
<td>What should the lower level goals cover? / “Potentially reached audience or part of it”</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>... planned to reach a potential audience...;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>What is the usual purpose of public relations in your company? / “Potentially reached audience or part of it”</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>...Then we can introduce quantitative goals. Let’s say about product x ... nobody... still do not know, such product like doesn’t exist, but we introduce a parameter that in... that in one year will learn from target audience; let’s say this target audience is formed, by... well by the demographic, social and other parameters and thereof from it will learn say 10 or 20 per cent...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next using the operationalization procedure the operationalization model of the “Productivity concept” was created (Table 3). It unifies the entire sequence of content interpretation and coding into one logical structure. This model differs from the more commonly accepted ones because in this case operationalization of the concept takes

---

**Inzinerine Ekonomika-Engineering Economics, 2013, 24(1), 28-35**
place at the level of understanding rather than semantic terms. This enables unification of the meaning of various definitions and analyzing them using mathematical methods.

**Construction of the Predicted PR Productivity Index**

The indicator of the predicted PR productivity assessment was constructed by creating a mean of four primary variables: the number of positive/negative articles, topics, the number of times the company was mentioned in the media and the potential audience (Table 4).

The results of reliability analysis show representative capacity of the index. However, due to a relatively small number of mentions, it is more appropriate to use induction rather than deduction for assessing the indicator’s quality.

Reducing the potential number of testing insights to the scientifically justified level it can be argued that the constructed indicator in similar conditions will definitely reflect the totality of its attributes. On the other hand, if we interpret the results of such tests from the empirical theory of truth perspective, the results confirm that the model of transformation of the primary variables into the index is correct, and that index construction models would also be suitable for defining the overall change in attributes in the larger scale research studies.

<table>
<thead>
<tr>
<th>PR impact assessment</th>
<th>Corrected Item-Total Correlation</th>
<th>Construct validity parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of times the company was mentioned in the media.</td>
<td>0,537</td>
<td>Inter-Item Correlations</td>
</tr>
<tr>
<td>Topic where company is mentioned</td>
<td>0,537</td>
<td>r_{min}</td>
</tr>
<tr>
<td>Potentially reached audience or its part</td>
<td>0,537</td>
<td>r_{max}</td>
</tr>
<tr>
<td>No. of positive, neutral, negative articles about the company</td>
<td>1,000</td>
<td>Cronbach's Alpha</td>
</tr>
</tbody>
</table>

**Conclusions**

During the compile procedure there was found, that the Hermeneutic analysis could be useful tool for quantitative content processing. The hermeneutic analysis allows the understanding of the real authorial intention and is useful for the preparation of it for the further coding.

Testing results of quantitative content processing methodology based on verbal narrative coding principles confirmed:

- The methodology creates premises to compare statistically the information received from the narratives not only from the context of words or their collocations but from the context of the real authorial intention as well.
- Introduced methodology enables to solve information gathered through qualitative and quantitative way comparability issue and expands the boundaries of scientific understanding of the social research process.
- Gives to the researchers more possibilities to explore the coded text mathematically and to link obtained insights into solid logical structure by identifying and positioning in assessment system new not investigated attributes of the research object and instantly identifying its influence on the behavior of the object.

**References**


