Qual Quant (2011) 45:859–873
DOI 10.1007/s11135-010-9399-4

RESEARCH NOTE

Mixing quantitative and qualitative content analysis: triangulation at work

Anton Oleinik

Published online: 4 December 2010

Abstract The paper discusses an original path model for triangulating the results of three types of content analysis: (i) the analysis of the co-occurrence of words; (ii) the substitution model of quantitative content analysis and (iii) qualitative coding. The model refers to the "within method" as well as the "between methods" of triangulation. It shows how this model helps assess the reliability and validity of content analysis. In the former case, the text is used as a unit of analysis; in the latter—the variable (theme, category or qualitative code). The model is tested on two sets of transcripts of semi-structured ($N = 64$) and unstructured ($N = 43$) interviews carried out independently by two research teams. Outcomes of the test show consistent patterns in both cases. Some directions for further explorations are discussed, including latent qualitative analysis.

Keywords Triangulation · Qualitative content analysis · Quantitative content analysis · Path model · Reliability · Validity · Semiotics

Introduction: the text as a key source of information in the humanities and social sciences

Although the subject-matter of the natural sciences—the objective reality—can be studied directly, the humanities and social sciences deal with the reality *as we perceive it*. For instance,

This article is a significantly abridged and adapted version of Methodological appendix of "Market as a Weapon: The Socio-Economic Machinery of Dominance in Russia", copyright Transaction Publishers 2010. Reprinted by permission.

A. Oleinik (✉)
Department of sociology, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada
e-mail: aoleynik@mun.ca; aoleynik@hotmail.com

A. Oleinik
Institute of Economics of Russian Academy of Sciences, Moscow, Russia

Weber (1968, p. 4) sees the specific purpose of sociology in interpreting social action. The external observer discovers a subjective meaning attributed by the actor only indirectly, by studying traces that his or her actions leave on material objects, visual images, texts, pieces of art and so forth.

The text, whether it is an official document, a novel, the transcript of an interview or a personal diary, is a key source of information about the subjective perceptions and justifications elaborated by the actor so that the other interacting parties correctly interpret his or her intentions. The humanities scholar “does not deal with an empirical subject, but with a *possible*, *assumed* subject as he represents himself in his other being—in the *text*” (Bibler 1991, p. 72; emphasis in the original; see also Lotman 1990, pp. 2–4).

Viewed from this perspective, the relevance of content analysis goes far beyond narrowly defined boundaries of semiotics, hermeneutics, linguistics and a few other disciplines in the humanities and social sciences. Content analysis then transforms into a key methodological tool of humanitarian and social inquiries. The fact that there exist highly heterogeneous techniques of content analysis commonly producing divergent outcomes makes such status of content analysis questionable. The main dividing line lies between qualitative content analysis with its stronghold in semiotics and quantitative content analysis, which is widely used by linguists. The present text focuses on discussing possible solutions for dealing with the instability of the outcomes of content analysis that undermine its methodological value.

White and Marsh (2006) usefully summarize key distinctions between the qualitative and quantitative versions of content analysis. They point out their different ontological and epistemological roots: quantitative analysis is positivist in its orientation whereas qualitative analysis is “interpretativist”; the first is “objectivist” whereas the second—“constructivist”. Because of this, qualitative content analysis is criticized for its highly subjective character and difficulties with controlling the impact of the coder’s personality. Criticisms of quantitative content analysis, on the other hand, suggest that the exclusive reliance on frequencies makes the humanities and social sciences a province of the natural sciences.

The existence of two alternative approaches within quantitative content analysis further complicates the situation (Hogenaraad et al. 2003). On one hand, the correlational tradition heavily emphasizes co-occurrences of words; it classifies words in function of their co-occurrence with other words (“categories”). On the other hand, the substitution tradition implies that classifications of words derive not from simple co-occurrences but from ad-hoc dictionaries (“themes”) built by the researcher to test particular hypotheses and theories.

Calls for finding a middle ground and overcoming existing oppositions are not uncommon. For instance, Eco (1990, x) proposes to combine semiotics with methods of quantitative content analysis. Ermakov et al. (2004) use methods of both qualitative and quantitative (the correlational approach) content analysis when inquiring into the image of power in Russian culture and the impact of a particular model of power observed in this country on linguistic structures. However, qualitative and quantitative methods are rarely used concurrently and applied to the same text, which complicates the task of comparing their results. An original methodology for confronting outcomes of qualitative and quantitative content analyses is outlined in what follows and tested on two different sets of texts.

Section 1 discusses various issues in the reliability and validity of various forms of triangulation applied to content analysis. It helps formulate two research questions as to how to measure the reliability and validity of alternative analyses of the same text. Path-models for various forms of content analysis of the same text are outlined in Section 2. Section 3 describes the two sets of texts—transcripts of in-depth unstructured and semi-structured interviews carried out in the framework of two separate research projects—used for testing the path-model empirically. Section 4 presents outcomes of a series of empirical tests performed with the help of software programs for content analysis *QDA Miner* version 2.0.8 and *WordStat* version 5.1.12 developed by *Provalis Research*.

1 Reliability and validity of various forms of triangulation

Practices of combining different methods of research are commonly referred to as triangulation. The idea of triangulation and its forms (Jick 1979)—"within method" (the use of multiple techniques within a given method) and "between methods" (the use of two or more distinct methods to analyze the same data)—seems easy to grasp. However, in practice, the task of integrating multiple methods, cross-checking the reliability and validity of their outcomes represents a serious challenge (Gray and Densten 1998, p. 420).

The key problem lies in finding a common denominator for qualitative and quantitative data without which meaningful comparisons of the outcomes of "between methods" triangulation are hardly possible. How can the data be converted from one format to the other, how can it be translated without losing much information? Jick (1979, p. 609) argues in favor of converting the numerical data into a format suitable for qualitative analysis. However, this approach makes replication extremely difficult. Further, there is no way to assess the reliability of triangulation and its outcomes. Similarly to the case of field-notes and accounts written by an anthropologist about his or her journey to a distant country, the perception of the outcomes of triangulation with a qualitative denominator depends on the personal reputation of their author (Geertz 1988).

The search for a quantitative denominator has pitfalls as well. It may involve a significant loss of qualitative information when converting it into a numerical form. Most software programs for content analysis allow for *either* its qualitative version *or* its quantitative version, which prevents the retrieval of output in a compatible numerical format. In the best case, the question about reliability, i.e. "whether the results of a study are repeatable" (Bryman 2004, p. 28), is answered *separately* for qualitative and quantitative content analyses of the same text. Reliability of quantitative content analysis involves recounting words and their clusters, themes and categories, using various software programs.[1]

The reliability of qualitative coding is conventionally measured with the help of intercoder agreement coefficients. They range from simple percent agreement $PA_O = \frac{Total A's}{n}$, where *A* refers to the number of units for which the coders agree and *n*—to the total number of coded units, to more sophisticated coefficients that allow for the discounting of high levels of agreement due to chance. Their list includes Cohen's *kappa*, Scott's *pi* and Krippendorff's *alpha* (Neuendorf 2002, pp. 153–158; Warner 2008, pp. 833–834). Of these three, Cohen's *kappa* $\kappa = \frac{PA_O - PA_E}{1 - PA_E}$, where PA_E refers to the chance level of agreement, is most widely used. An inter-coder agreement coefficient exceeding 0.7 suggests an acceptable level of reliability (Gray and Densten 1998, p. 423).

The assessment of the reliability of qualitative coding requires time and substantial human resources. Budget constraints limit the scope of involving at least one additional coder, let alone several, especially in the case of doctoral research. Concurrent qualitative content analyses also appear very time consuming as the coders need to get together and discuss common coding policies and their application.

[1] Sometimes these multiple runs are not without problems. For instance, in the process of working on this text the author used two computers. To his great surprise, despite using the same content analysis software program, they initially produced slightly divergent results when calculating frequencies of categories. After the intervention of the software producer's representative, who made some adjustments in the algorithm for retrieving words, the problem was solved.

The task is further complicated by the necessity to assess the reliability of triangulation, or the joint reliability of quantitative and quantitative content analyses of the same text. Strategies for assessing it are neither known nor have they been empirically tested. This prompts the first research question as to *how can the reliability of triangulation be measured, both "within method" and "between method", in content analysis*. An ideal algorithm should also satisfy the criterion of efficiency by saving time and human resources: "How can the reliability of triangulation be measured in a reasonable amount of time?"

The situation for assessing the validity of content analysis presents even more challenges. "Validity is concerned with the integrity of the conclusions that are generated from a piece of research" (Bryman 2004, p. 73). A repeatedly observed character of patterns of themes and categories does not guarantee that they accurately tap the meaning of the text. In fact, the text does not necessarily have a unique meaning. Lotman (1990, Chap. 3) differentiates two types of the text: logical and rhetorical. The purpose of the former lies in conveying the author's message—single and unambiguous. A valid content analysis should then "decipher" this message as accurately as possible. The rhetorical text, on the contrary, is intended to prompt discussions and various interpretations. Bakhtin (1979) concept of "polyphonic novel" containing a plurality of "voices" with their own "truths" along with the author's concept clearly corresponds to this type of the text. The validity of the outcomes of content analysis, then, has to be judged with respect to a particular position and "reading" chosen by the coder.

The two traditions in quantitative content analysis allow for switching positions in keeping with the type of the text. The correlational tradition appears appropriate for analyzing the logical text. This approach allows for a "representational" reading. The substitution tradition fits the spirit of the rhetorical texts with multiple readings that it prompts. "When a researcher understands texts representationally, they are used to identify their sources' intended meanings. When a researcher understands texts instrumentally, they are interpreted in terms of the researcher's theory" (Roberts 2000, p. 262). This citation highlights ambiguities with identifying the exact type of the analyzed text: in some case it can be classified either in one category or in the other. Nevertheless, the structure of the text can be considered a proxy for its type: "The less structured the text, the more structuring and categorizing must be the analysis" (Hogenaraad et al. 2003, p. 226). A textbook chapter conveys a unique message about knowledge accumulated in a particular field, which calls for a representational reading. Qualitative analysis—as a result of its subjective component—has some "elective affinity" with instrumental reading.

The existence of multiple points of reference—the author's intention, the readers' interests—turns into an advantage when triangulating outcomes of qualitative and quantitative content analysis. "The [convergent] validity of a measure ought to be gauged by comparing it to measures of the same concept developed through other methods" (Bryman 2004, p. 73). This explains the second research question addressed here, namely *how to measure validity of triangulation, both "within method" and "between method", in content analysis simultaneously controlling the reference point*.

2 Path-models for content analysis

A series of path-models help address the issues of reliability and validity in content analysis. They derive from the idea of mapping the outcomes of quantitative and qualitative content analyses of the same text and quantitatively measuring the strength of the relationship between different points on this map. Content analysis of a logical text can be represented as a path starting at a point corresponding to the text's intended meaning. In quantitative terms, this point can be described by a vector of themes, or clusters of co-occurring words: $T = \langle t_1, t_2, t_3 \ldots t_n \rangle$. The correlational approach in quantitative content analysis serves to specify its parameters. The end point of the path is represented by a constellation of qualitative codes $Q = \langle q_1, q_2, q_3 \ldots q_n \rangle$. These codes are attributed by the reader (coder) in his or her attempts to decipher the original meaning of the text conveyed by the author. The path goes through a middle point, a vector of categories $C = \langle c_1, c_2, c_3 \ldots c_n \rangle$. C's parameters refer to frequencies of particular entries in the ad hoc dictionary based on substitution. When creating an entry, the reader again thinks about the author's original meaning. The fact that the themes, the categories and the qualitative codes all have the same purpose of conveying the author's intentions makes quantitative relationships between vectors T, C and Q meaningful.

Content analysis of the rhetorical texts involves changing the start point. In this case, the path starts from a constellation of qualitative codes $Q_j = \langle q_{1j}, q_{2j}, q_{3j} \ldots q_{nj} \rangle$ attributed in keeping with a particular reader's interests and theories. Since, usually, there is no single reader but rather several ($j = 1, \ldots$, m) readers, it is worth speaking about the paths in the plural. Even the same reader can adapt divergent approaches to the same text and read it from various perspectives. All paths lead to the same end point, the text, even if they highlight different parts of it and, hence, different parts of vector T, which produces a spark-like path-model. Each of these paths goes through a particular middle point whose exact position can be located with the help of vector $C_j = \langle c_{1j}, c_{2j}, c_{3j} \ldots c_{nj} \rangle$. However, categories in the dictionary based on substitution this time derive from a reader's interests and theories instead of the author's.[2]

Links between vectors at the same level of the path model, e.g., between all vectors of categories C or between all vectors of qualitative codes Q, can also be explored. "It is necessary to ensure that comparisons are made at the same level (e.g., that quantitative manifest variables are compared with qualitative manifest data, or quantitative latent variables are compared with qualitative latent variables)" (Gray and Densten 1998, p. 421). However, the interpretation of associations will depend on the research question. If reliability is at stake, then one set of associations counts; if validity—the other.

2.1 Assessing reliability of content analysis

In this case, reliability refers to uniformity and consistency in coding texts using a specific frame of analysis: either the author's perspective or the reader's. If a single reader codes a series of texts, the degree of reliability of his or her content analysis can be assessed without involving additional coders and training them to apply the same set of criteria. The principal emphasis shall be placed on measuring associations between vectors T, C and Q lying on the same path while using text as the unit of analysis.

Each of the points on the path can be represented in a matrix form. Rows in these matrices correspond to cases (texts), columns—to variables (themes in matrix T, categories in matrix C and qualitatively coded segments in matrix Q). Cell values then contain the information about the presence or absence of a particular theme, category or code in a given text or their frequencies. Distances between rows or columns—depending on the choice of the unit of analysis—can be measured by means of factor analysis. The assessment of reliability calls

[2] In terms introduced by Skinner (2002, p. 93), categories C refer to "meaning$_3$" and codes Q—to "meaning$_2$". Skinner's "meaning$_1$" differs from the interpretation of co-occurrences T proposed here: according to him, they are independent of both the author's and the reader's intentions and represent the "the arbitrariness of the sign" (Derrida 1967, p. 74).

for Q-mode of factor analysis, whereas the choice of the variable as the unit of analysis necessitates R-mode (Basilevsky 1994, pp. 278–282).

At this stage, only one element of factor analysis, the use of the cosine of the angle between two vectors to measure the distance between them, is used (Salton and McGill 1983, p. 203). The cosine coefficient of similarity $SIM_{Cos}\left(text_i, text_j\right) = \frac{\sum_{k=1}^{t}(VAR_{ik} \times VAR_{jk})}{\sqrt{\left(\sum_{k=1}^{t} VAR_{ik}^2 \times \sum_{k=1}^{t} VAR_{jk}^2\right)}}$ accounts for the length of the text: a short one can still be compared with a long one (Grossman and Frieder 2004, p. 19). The frequency with which a variable VAR_k occurs can be weighted by the inverse frequency of the texts containing it $TF{*}IDF = VAR_k \times \log\left(\frac{N}{n_k}\right)$, where N is the total number of texts, n—the number of texts containing VAR_k, in order to increase the precision of the instrument (Salton and McGill 1983, p. 62; Grossman and Frieder 2004, p. 2).

A reliable content analysis implies a high level of congruence between matrices T, C and Q. In other words, a particular text occupies a similar position in all of them with respect to the other texts as a result of consistency in identifying themes and categories and in attributing qualitative codes provided that the categories and qualitative codes derive from the same perspective (the author's or a reader's). In this case, triangulation serves to assess a reader's consistency in qualitative coding across texts that he or she coded, with themes and categories serving as "yardsticks" against which qualitative codes are judged. Coding patterns are compared across texts, not across readers. If the values of the cosine coefficient for the same text in matrices T, C and Q are associated, then content analysis produces reliable outcomes.

The list of possible measures of association, or congruence, between the matrices includes the Pearson correlation coefficient r between raw values of the cosine coefficients and Spearman's rank correlation coefficient rho. If the latter is chosen, the texts have to be rank-ordered according to the degree of their similarity. This option involves a significant loss of information, which is particularly visible when the sample of analyzed texts is small. The Pearson correlation coefficient does not have this shortcoming.

The calculation of partial correlation coefficients, for instance, between matrices T and Q, also makes sense. This serves to control the impact of the dictionary based on substitution that lies at the origin of matrix C and answer the question as to whether it can be considered a confounding variable, an intervening variable or a source of spuriousness.

Two strategies exist in regard to selecting the $centroid$, or the vector selected as the origin for the purpose of comparisons. In principle, "the choice of point selected as origin does not affect the result" (Basilevsky 1994, p. 283). However, measurement errors and even minor inconsistencies in coding may affect the overall result when changing the centroid. To minimize the impact of non-systematic inconsistencies and errors, one of vectors with the highest similarity index in a matrix shall be chosen as the centroid. For rhetorical texts, matrix Q seems appropriate for identifying the centroid that will be used for calculating the cosine coefficients in matrices T and C as well; for the logical texts—matrices C or T. The other related problem consists in deciding whether one or several centroids are to be selected for multiple readings of a set of rhetorical texts. "Pure" rhetorical texts call for the latter option, "mixed" cases (rhetorical texts with some elements of logical texts) make it possible to select just one centroid.

2.2 Assessing validity of content analysis

If words and text fragments are consistently placed in the same theme, category or qualitative code, another question arises: do these measures adequately tap the author's or the reader's concepts? To address this question, the degree of the validity of content analysis needs to be gauged. This can be done after making a few adjustments and corrections in the path model outlined above. First, the unit of analysis is no longer the case, but the variable (theme, category or qualitative code), which paves the way for the R-mode of factor analysis. It helps see whether similar constellations of qualitative codes and categories characterize similar texts. Second, validity means congruence in the distribution of themes, categories and qualitative codes. Equal distances separate, on one side, two given qualitative codes in matrix Q and, on the other side, two corresponding categories in matrix C if both qualitative codes and categories in the dictionary based on substitution derive from the same underlying concepts. This involves calculating the cosine coefficients of similarity between themes, categories and qualitative codes and then cross-correlating these coefficients.

The valid within (quantitative) method of content analysis implies that values of the cosine coefficients of similarity between variables in matrices T and C are correlated. This is possible, when the dictionary based on substitution does not corrupt the original language of the text. A strong correlation between the values of the cosine coefficients of similarity between variables in matrices Q, C and T indicates the validity of the between (qualitative and quantitative) method of content analysis. Finally, the validity of triangulation increases one's confidence in the validity of the outcomes of qualitative and quantitative analyses. Valid triangulation conditions the convergent validity of the categories and qualitative codes used in content analysis. The latter serves as the criterion indicator for the former and vice versa.

2.3 Criteria for assessing the path model

Several criteria have to be kept in mind when evaluating the proposed path model. First, there is a systematic measurement error when relying on the values of the Pearson correlation coefficients to assess the degree of congruence between the cosine coefficients in the three matrices (except between matrices T and C). In fact, the values the Pearson correlation coefficients depend not only on the degree of the reliability and validity of content analysis, but also on the inherent differences between qualitative and quantitative content analyses. For instance, qualitative coding can have different levels of "depth", ranging from taking the text at face value (manifest qualitative coding) to reading "between lines" (latent qualitative coding). Quantitative coding has far fewer degrees of freedom in this respect. Consequently, one can predict stronger correlations between the cosine coefficients of similarity for matrices T and C (both refer to quantitative content analysis) than for matrices C and Q (because matrix Q contains the outcomes of qualitative content analysis).

Conventionally calculated inter-coder agreement coefficients provide a yardstick for assessing the values of the Pearson correlation coefficients for matrices Q and C, as well as Q and T. Inter-coder agreement coefficients arguably account for possible variation in the levels of depth in qualitative content analysis. Thus, the Pearson correlation coefficients can be "standardized" to the level of agreement deemed acceptable for the purpose of assessing the reliability of content analysis.[3] In fact, in this case, the Pearson correlation coefficient is nothing other than a particular inter-coder agreement coefficient calculated for the same coder at different moments in time. For instance, if the value of the inter-coder agreement coefficient deemed acceptable is 0.8, then a Pearson correlation coefficient for matrices Q and C equal to 0.6 in fact suggests a stronger association—and the degree of reliability—in the order of 0.75 $\left(r = \frac{0.6}{0.8}\right)$.

[3] This fruitful idea was initially suggested by Normand Péladeau of *Provalis Research* in a personal communication dated April 11 and 13, 2008.

Second, the dictionary based on substitution and, thus, matrix C, can be built either ex ante or ex post qualitative content analysis. The ex ante built dictionaries have the following flaw: they "lift the words out of context" (Hogenaraad et al. 2003, p. 225). This flaw brings matrix C closer to matrix T by increasing the strength of the association between them and farther from matrix Q. The ex post built dictionaries have the opposite effect on the path-model: entries in the ex ante built dictionaries are "customized" to a particular phrases suitable for describing concepts in general, whereas the ex post dictionary shows how these concepts are described by a given author.

Third, long texts tend to deflate the values of the cosine coefficients of similarity. Long vectors in which they are transformed include many variables. As a result, the probability of a perfect fit between long vectors decreases (Salton and McGill 1983, p. 203).

3 Description of the data: content analysis of in-depth interviews

The path model was empirically tested on two sets of in-depth qualitative interviews with state representatives and experts on the issues of state service in the Russian Federation. In one case, interviews were more structured than in the other, which allows applying the path model to the content analysis of various types of texts. The interviews were conducted within the framework of two separate research projects, by two independent research teams, each containing several interviewers.[4] Although both studies focused on the power elite—people "in positions to make decisions having major consequences" (Mills 1957, pp. 3–4), they addressed different research questions: (i) what are the key characteristics of a particular model of power that structures interactions within Russia's power elite and between its members and ordinary people; (ii) whether members of Russia's power elite play a stabilizing or destabilizing role in socio-economic development. In what follows, the set of semi-structured interviews conducted in the framework of project (i) is referred to as the "bureaucrats" set ($N = 64$) and the set of unstructured interviews carried out in the framework of project (ii), the "crisis" set ($N = 43$).

The ideas and conclusions generated by these more than 100 interviews can hardly be generalized since, in specialized and elite interviews, "the population cannot be satisfactorily randomized or stratified in advance" (Dexter 2006, p. 43). This prevents one from safely generalizing findings beyond the particular samples. However, the two sets of transcripts represent a good fit for testing the path model.

First, the transcripts represent a diversified collection of texts. As the mean standard deviations suggest, their length varies significantly (more—in the "bureaucrats" set, less—in the "crisis" set). If the proposed strategy of triangulation works under these conditions, one could assume that it will perform better when applied to less heterogeneous sets, such as summaries of scientific articles (Hogenaraad et al. 2003).

Second, transcripts of in-depth interviews can be read in an instrumental as well as in a representational manner. Structured interviews have two "authors": the interviewee and the interviewer, who produces the interview guide. After conducting and transcribing the interview, the interviewer transforms into the reader attributing codes and creating entries in the dictionary based on substitution. In other words, the format of the qualitative interview

[4] Project "Particularities of power in the post-Soviet context: theoretical considerations and empirical studies of bureaucracy" was supported by the Social Sciences and Humanities Research Council of Canada (award No 820-2005-0004). Some of its outcomes are discussed in a collective monograph (Oleinik 2008). Project "Post-Soviet elite" was conducted by members of the Levada Center (Moscow) with the financial support of the "Liberal Mission" Foundation and resulted in a separate book (Gudkov et al. 2007).

provides a rare opportunity for switching from the author's position to the reader's and, consequently, from a representative reading to an instrumental one. Multiple readings of the same text further allow for assessing how close it lies to the logical and rhetorical formats. This can potentially be done by comparing the strength of associations between matrices T, C and Q lying on various axes of the spark-type path-model. If the values of the Pearson correlation coefficients in one set substantially exceed the corresponding values in the others, it indicates a rhetorical nature of the text.

For that reason, the content of both sets of transcripts was analyzed from multiple perspectives, i.e., using multiple "frames". The "crisis" framework (a code book containing descriptions of thirteen qualitative codes grouped under three headings: "instability", "stabilizing factors" and "destabilizing factors" plus a dictionary based on substitution with a similar structure) is derived as closely as possible from the interview guide for the "crisis" project: this refers to the "author's perspective" in this case. However, the "crisis" framework was also used in the content analysis of the "bureaucrats" set where it represented a "reader's perspective"—one of several. Vice versa, the "bureaucrats" frame (a code book with forty-one qualitative codes grouped under five headings: "business", "constitution", "constraints", "critical situation" and "hour-glass society" plus a dictionary based on substitution with a similar structure) suggests a representational reading of the "bureaucrats" set and an instrumental reading of the "crisis" set. A third frame, "power" (a code book consisting of fifteen qualitative codes grouped under three headings: "power in a pure form", "techniques for imposing will" and "domination by virtue of a constellation of interests" plus a corresponding dictionary based on substitution), lies closer to the research questions explored in the "bureaucrats" project, but provides a new perspective for reading the transcripts.

The dictionaries based on substitution derive from a combination of the ex ante and the ex post techniques for building them. At the first stage, a list of words and phrases that presumably fit each code from the code book was prepared. Then, after the qualitative content analysis was carried out, they were contextualized and "fine-tuned" using particular expressions frequently found in the transcripts.

Two modules of a software program, *QDA Miner* and *WordStat*, can be used to perform all forms of content analysis in virtually any language without spending much time and resources on converting the data sets when switching from qualitative to quantitative content analysis and to retrieve outcomes in a compatible format. These two modules can be used to produce the T, C and Q matrices.[5] Only the author was involved in coding fragments of the interviews and building the dictionaries based on substitution.

4 Discussion: on the plurality of readings

Several iterations in attributing qualitative coding appeared necessary to achieve a satisfactory—at least moderately so—strength of associations between matrices Q, C and T (the Pearson coefficient of correlation exceeding 0.25 or 0.3125 on the "standardized" scale). Depending on the particular set of interviews and the framework applied in a specific reading,

[5] Unfortunately, the version of *WordStat* used calculates only the Jaccard coefficient of similarity $SIM_{Jaccard}\left(text_i, text_j\right) = \frac{\sum_{k=1}^{t}\left(VAR_{ik}\times VAR_{jk}\right)}{\sum_{k=1}^{t} VAR_{ik}+\sum_{k=1}^{t} VAR_{jk}-\sum_{k=1}^{t}\left(VAR_{ik}\times VAR_{jk}\right)}$ (Salton and McGill 1983, p. 203) for matrix Q. It does not affect the meaningfulness of using the Pearson coefficients of correlation to gauge congruence between the matrixes, yet it certainly decreases the precision—reliability and validity—of the qualitative content analysis. The Jaccard coefficient takes into account the occurrence of a code, not its frequency in a particular text.

the desired outcome was achieved after 2–4 iterations. In other words, each of 107 transcripts was read and coded 9–10 times on average, which took more than six months of work in all (not counting the time spent on transcribing, the initial reading, developing and improving the interview guides). Furthermore, fragments of various transcripts coded in a similar manner were also retrieved and re-read "across transcripts", in order to ensure consistency in qualitative coding. This amount of time far exceeds that necessary to process a purely quantitative data set, but certainly falls short of the time budget required for content analysis by a group of coders. It also seems affordable for a PhD student working on his or her thesis.

As a result of the arguably rhetorical nature of the analyzed texts, matrix Q was used to identify the centroid. Two versions of the path model for assessing the reliability of the content analysis were tested both on the "bureaucrats" set and the "crisis" set: with a single centroid identified through the path of the "original" reading (the "bureaucrats" framework for the "bureaucrats" set, the "crisis" framework for the "crisis" set) and with three centroids identified for each path.

The values of the Pearson correlations between the cosine coefficients of similarity in matrices lying on the same path correspond to the pattern predicted in Sect. 2.1 (Fig. 1). Associations between matrices T_i and C_i (the within method triangulation) tend to be stronger than between matrices C_i and Q_i, on the one hand, and matrices T_i and Q_i (the between methods triangulation), on the other. Association only between two matrices, C_1 and Q_1 (one pair out of eight), suggests that a further improvement in qualitative coding using the "crisis" code book may be desirable.

It also makes sense to calculate coefficients of the partial correlation between the cosine coefficients of similarity in matrices T_i and Q_i and compare them with the Pearson coefficients of correlation. This serves to control for C_i, i.e., for the impact of the substitution tradition of quantitative content analysis. In all three cases, the correlation decreases, but does not drop significantly: $r_{T1Q1.C1} = 0.213(\text{cf.}r_1 = 0.295)$, $r_{TQ2.C2} = 0.317(\text{cf.}r_2 = 0.483)$ and $r_{T3Q3.C3} = 0.283(\text{cf.}r_3 = 0.405)$. This means that the dictionary based on substitution partially mediates the association between matrices T and Q (Warner 2008, pp. 407–409). This further supports the logic underpinning the path model: matrix C lies "halfway" between T and Q yet has a dynamic of its own. The substitution tradition of quantitative content analysis, if it combines the ex ante and ex post techniques for building dictionaries, represents a "hybrid" form with elements of the representational quantitative analysis and qualitative coding yet acquiring a unique identity.

The values of Spearman's rank correlation coefficients for the same path model are predictably lower than the corresponding values of the Pearson correlation coefficients because of the loss of information. In fact, a moderately strong association between them characterizes the "bureaucrats" set (r=0.666), which may indicate that the loss of information has a consistent character.

It is worthwhile to discuss the relationships between matrices of the same type, e.g., between C_1, C_2 and C_3, or between Q_1, Q_2 and Q_3, with a single centroid in triangulation (the values of the Pearson correlation coefficients are indicated in square brackets in Fig. 1). Out of six pairs, only two (C_2 and C_3, C_1 and C_3) do not show at least a moderate strength of association. Association along the diametral paths suggests either that the "author's message" "overpowers" the alternative readings, or a pattern that may emerge by chance in relationships between any long vectors (Grossman and Frieder 2004, p. 19). The regularity formulated above may well apply to the case under consideration: two particular transcripts, with especially significant word counts, when read from different angles, tend to appear close to each other because of their richness in relevant information. In the same vein,

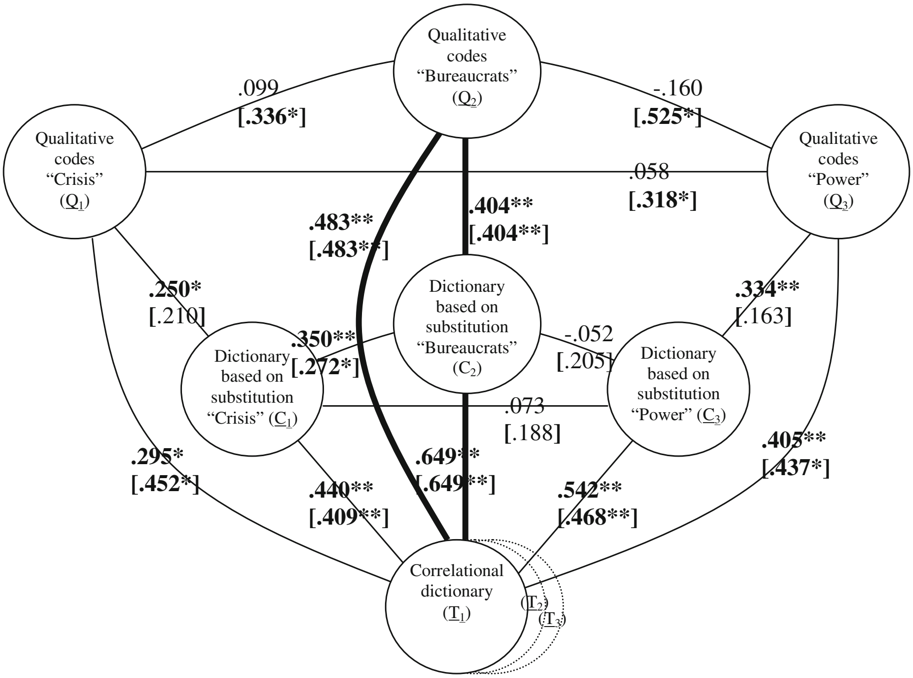

Fig. 1 Path model for assessing reliability, the "bureaucrats" set. Moderate-strong correlations are indicated in *bold*. The level of statistical significance (* for correlations significant at the 0.05 level, ** for correlations significant at the 0.001 level) is indicated for the sake of reference: the non-random character of the sample of transcripts makes statistical inferences meaningless. In this case, the values of the Pearson correlation coefficients have a purely descriptive meaning. The strength of association between matrices for the case of a single centroid—from the "bureaucrats" framework (from matrix Q_2)—is indicated in *square brackets*. The path that refers to the "original" ("bureaucrats") framework is indicated by *thick lines*

short transcripts appear close to each other independently of the content analysis framework because of their lack of relevant information.

A pair-wise comparison of the Pearson coefficients of correlation for, on one hand, the path model with a single centroid, and, on the other hand, one with three centroids, may help in assessing a possible impact of the relative power of the author's message, even if the text does not seem to be the most appropriate unit of analysis compared with the variable. When using the text as the unit of analysis, the impact of the author's message can be measured only indirectly: the closer the content analysis framework is to the author's intentions, the more consistent its outcomes. The observed pattern, weaker associations between matrices T_i and C_i, C_i and Q_i, and stronger association between matrices T_i and Q_i lying on the same "non-original" path,[6] provide some further evidence confirming the relative strength of the author's message. After all, the "bureaucrats"—"original"—path shows the strongest associations in the path model with a single centroid. Yet even this additional evidence does not serve to rule out the alternative explanations, as the discussion below demonstrates.

The path model for assessing content analysis reliability was also tested on the "crisis" set of transcripts with different—to some extent—parameters. Some results were completely

[6] Relatively strong associations between matrixes Q_i, Q_j and Q_k and, as a result, stronger associations in the path model with a single centroid between T_i and Q_i, may be due to the use of the Jaccard coefficient of similarity in matrixes Q (see Note 5).

replicated, while others were not. Practically all measurements referring to narrowly defined reliability confirm their precision. Associations along all the paths, including the one corresponding to the "original" framework for this set (the "crisis" framework), appear either moderate or strong, which suggests a higher level of content analysis reliability in this case compared with the previous one. The values of the coefficients of the partial correlations also show very similar patterns consistent with the interpretation proposed above.

The loss of information when relying on Spearman's rank correlation coefficients in this case seems to be even more substantial and less consistent along various paths. The Pearson correlation coefficient between pairs of measures (r and rho) gauging the strength of association between the same pairs of matrices is close to zero. This may be due to a more substantial similarity (indicated by the mean Pearson correlation coefficient standard deviation) between texts in this set. As a result, there are more cases with the same rank. A small size of the "population" (15 pairs of the correlation coefficients) makes it impossible to draw too far-reaching conclusions (Warner 2008, p. 269).

However, simple references to the size of the sample do not serve to explain a different pattern of associations along the diametral paths. In fact, the "original" path does not show the strongest associations in the path model with a single centroid. Furthermore, the "alternative" frameworks (the "bureaucrats" and the "power") serve to strengthen the associations between matrices lying on the corresponding paths contrary to the outcomes of the content analysis of the "bureaucrats" set. As there is less variability in the word count of the transcripts, the impact of the length of the vectors may be less significant in the "crisis" set. Thus, the observed differences may be due to a less structured character of interviews in this case: non-structured texts prompt multiple readings. As state officials and experts were the principal interviewees in the "crisis" set as well, the "bureaucrats" framework may well appear more appropriate despite being "non-original".

Two empirical tests of the path model for assessing content analysis validity—one done on the "bureaucrats" set, the other one on the "crisis" set—served to collect additional evidence with respect to assessing the relative strength of the author's message. In this test, the variable (qualitative code or category in the dictionary based on substitution) is used as a unit of analysis. Two caveats have to be made before the outcomes of the tests are discussed. First, restricted options in *QDA Miner* and *WordStat* with respect to using themes as the unit of analysis resulted in the exclusion of matrix T from the path model for assessing the validity of the content analysis. For a similar reason, it includes the Jaccard coefficients of similarity instead of the cosine coefficients, which makes the path model less precise. Second, because of the small size of the populations ($n = 13$ in the "crisis" framework for content analysis, $n = 41$ in the "bureaucrats" framework and $n = 15$ in the "power" framework) the interpretation of the values of the Pearson coefficients of correlations between the Jaccard coefficients of similarity between variables calls for extreme caution.

Associations between all matrices show a consistent pattern: there are moderate to strong correlations (r ranging from 0.616 to 0.892) between the Jaccard coefficients of similarity in matrices C_i and Q_i lying along all the three paths (Fig. 2). There is a substantial difference in this regard between the "bureaucrats" set and the "crisis" set. The observed pattern suggests that the outcomes of the content analysis have a valid character: qualitative codes co-occur with categories corresponding to them. To weaken the effect of the smallness of the population, the two frameworks, "bureaucrats" and "power", were merged (both refer to the issues of power and actors vested in it, which can justify such operation). The association between matrices C and Q remains moderate: $r = 0.676$ in the "bureaucrats" set, $r = 0.579$ in the "crisis" set ($n = 56$).

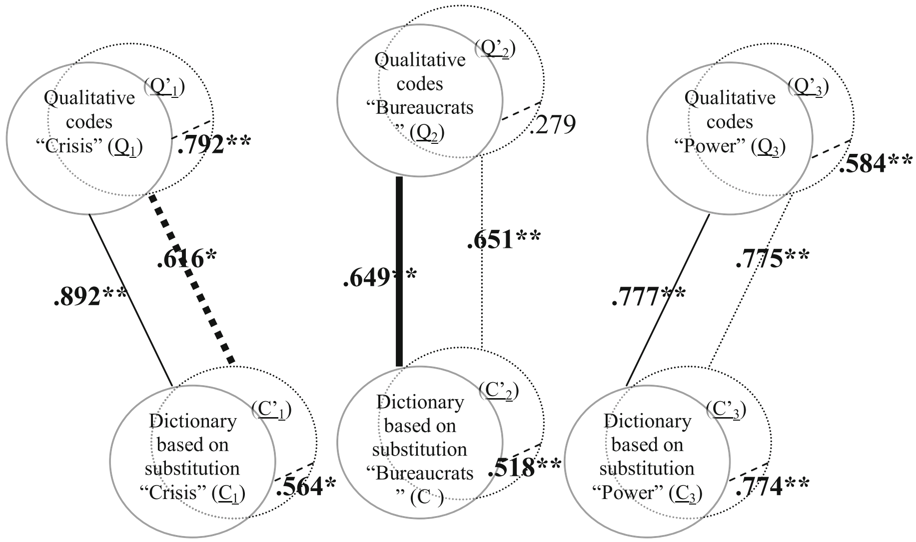

Fig. 2 Path model for assessing validity. The upper path model refers to the "bureaucrats" set, the lower—to the "crisis" set. The paths that refer to the "original" ("bureaucrats" and "crisis" correspondingly) frameworks are indicated by *thick lines*. The "power holders" code (heading "destabilizing factors") is the centroid in the "crisis" framework, the "vertical of power" code (heading "constraints")—in the "bureaucrats" framework, the "asymmetrical" code (heading "power in a pure form")—in the "power" framework

A moderate association between the constellation of qualitative codes in the "bureaucrats" set, on the one hand, and that in the "crisis" set (this refers to "horizontal" links in Fig. 2, e.g., between Q_i and Q_i'), on the other, appears relevant to the task of comparing the relative strength of various perspectives, the author's and the reader's. Certain congruence also characterizes the constellations of quantitative categories in the two sets. Out of six correlations of this type, only one ($r = 0.279$) is rather weak. This provides some support for the idea about a plurality of possible readings of the same text. Namely, transcripts of semi-structured and unstructured interviews allow a variety of their readings and interpretations. The transcripts included in the "bureaucrats" set are relatively more structured, yet the level of precision of the path model does not make it possible to determine if the "bureaucrats" framework for reading it is any better than the "crisis" one.

Conclusions: possible directions for further developments

A first, rather straightforward, direction for further improvements in the path model lies in comparing its outcomes when only one reader performs the content analysis and when a team of coders simultaneously apply a single analytical framework. This strategy serves to compare measurements of reliability derived from two independent sources: the "within method" of triangulation for team coding and a combination of the "within method" and "between methods" of triangulation for the path model.

Second, the path model should ideally be tested on two divergent sets of texts: on one hand, highly structured, purely "logical" texts (e.g., abstracts of scientific articles) and, on the other hand, completely unstructured, purely "rhetorical" (e.g., poems or transcripts of unstructured interviews). This approach serves to compare the two perspectives, the author's and the reader's, in a more direct manner. The theory predicts that the more structured the text is, the narrower its reading will be. However, the level of precision achieved with respect to the path model does not make it possible to clearly distinguish the unstructured interviews from the semi-structured interviews in this regard (a reliable measure of structuration also seems desirable).

Third, the precision of the path model can be improved by making the option of calculating the cosine coefficient of similarity available in the content analysis programs. Reliance on the Jaccard coefficient of similarity when triangulating results of qualitative content analysis leads to the consideration solely of co-occurrences but not frequencies or TF*IDF. Further, the other relevant improvement involves changing the way in which the cosine coefficient of similarity is calculated. The standard formula does not take into account the mean distance separating two codes or categories. In practice, this means that vectors with two codes lying close to each other and with the same codes separated by a long text will be considered highly similar, which reduces the precision in gauging congruence between matrices, especially when matrix T—the longest one—is involved. Alternatively, one can introduce "dummy variables" to measure distances between codes or categories in average units (expressed in the number of words, for example) of codes or categories. Let us assume the mean length of coded segments in the text is N words, which gives us the value of the unit of the dummy variable. Then the mean distance between two variables can also be expressed in k dummy variables $k = \frac{mean\ dist.}{N}$. As a result, the conventional matrix form transforms into a matrix with dummy variables corresponding to distances between each pair of variables, i.e., cell values for "regular" variables refer to their frequencies, cell values for "dummy variables"—to "frequencies" as mean distances between them expressed in N. Thus, to each variable a number of dummy variables measuring distances from it to all other variables shall be added.

References

Bakhtin, M.: Problemy poetiki Dostoevskogo. Sovetskaia Rossiia, Moskva (1979) [Problems of Dostoevsky's Poetics]

Basilevsky, A.: Statistical factor analysis and related methods. Wiley, New York (1994)

Bibler, V.: Mikail Mikhailovich Bakhtin, ili Poetika Kultury. Progress, Moskva (1991) [Mikhail Bakhtin, or Poetics of Culture]

Bryman, A.: Social research methods, 2nd edn. Oxford University Press, Oxford (2004)

Derrida, J.: De la grammatologie. Editions de la Minuit, Paris (1967)

Dexter, L.A.: Elite and specialized interviewing, 2nd edn. ECPR Press, Colchester (2006)

Eco, U.: Introduction. In: Lotman, Y. (ed.) Universe of the mind: a semiotic theory of culture., pp. vii–xiii. Indiana University Press, Bloomington and Indianapolis (1990)

Ermakov, S., Kim, I., Mikhailova, T, Osetrova, E. and Sukhovol'skij: Vlast' v russkoj iazykovoj i ethnicheskoj kartine mira. Znak, Moskva (2004) [Power in the Russian linguistic and ethnical world view]

Geertz, C.: Works and lives: the anthropologist as author. Stanford University Press, Stanford CA (1988)

Gray, J., Densten, I.: Integrating quantitative and qualitative analysis using latent and manifest variables. Quality & Quantity **32**, 419–431 (1998)

Grossman, D., Frieder, O.: Information retrieval: algorithms and heuristics, 2nd edn. Springer, Dordrecht (2004)

Gudkov, L., Dubin, B. and Levada, Y.: Problema "elity" v segodniashnei Rossii: razmyshleniia nad rezul'tatami sotsiologitcheskogo issledovaniia. Fond "Liberal'naia Missiia", Moskva (2007) [Problem of the elite in today's Russia: discussing results of a sociological study]

Hogenaraad, R., McKenzie, D., Péladeau, N.: Force and influence in content analysis: the production of new social knowledge. Quality & Quantity **37**, 221–238 (2003)

Jick, T.: Mixing qualitative and quantitative methods: triangulation in action. Administrative Science Quarterly **24**, 602–611 (1979)

Lotman, Y.: Universe of the mind: a semiotic theory of culture. Indiana University Press, Bloomington and Indianapolis (1990)

Mills, W.C.: The power elite. Oxford University Press, New York (1957)
Neuendorf, K.A.: The content analysis guidebook. Sage Publications, Thousand Oaks, London and New Delhi (2002)
Oleinik, A. (ed.): Reforming the state without changing the model of power? On administrative reform in post-socialist countries. Routledge, London and New York (2008)
Roberts, C.: A conceptual framework for quantitative text analysis. Quality and Quantity **34**, 259–274 (2000)
Salton, G., McGill, M.: Introduction to modern information retrieval. McGraw-Hill Book Co., New York (1983)
Skinner, Q.: Visions of politics regarding method, vol 1. Cambridge University Press, Cambridge (2002)
Warner, R.M.: Applied statistics: from bivariate through multivariate techniques. SAGE Publications, Thousand Oaks, CA (2008)
Weber, M.: Economy and society: an outline of interpretative sociology. In: Roth, G. and Wittich, C. (eds.) Bedminster Press, New York (1968 [1922])
White, M.D., Marsh, E.E.: Content analysis: a flexible methodology. Library Trends **1**(55), 22–45 (2006)