The State-of-the-Art of Content Analysis

Peter Gavora*

This methodological study presents an overview of the current variants of content analysis which were identified by examining their constitutive components. These components are the following: types of content, types of sampling, depth of analysis, direction of analysis, types of control of coding, and modes of presentation of findings. The current paper shows how these components are framed in the works of the relevant international authors. Five configurations of these components and their contents were identified, which represent five variants of content analysis. It has been shown that configurations are not arbitrary; rather their components establish certain relationships in order to form meaningful methodology.

Introduction

For a considerably long time content analysis has been a standard method of educational research. It has been aimed at the evaluation and interpretation of text content and applied in numerous research studies, focusing on a wide range of text types and genres within the educational domain.

Since the advent of content analysis in the 1950s the world of texts has substantially changed. The onset of electronic press made publishing quicker and cheaper, which resulted in a myriad of printed texts (textbooks, journals, professional books etc.). The production of texts has increased exponentially. New text types have emerged (e.g. electronic learning programmes), other texts have undergone significant modifications (e.g. textbooks), whereas the content of texts has also changed. Paper is no longer considered the only carrier of information. In addition to verbal content, today's content analysis concentrates also on visual content like pictures, drawings, images, as well as on complex entities like websites and TV programmes.

Originally, content analysis was a quantitative research method. Berelson, whose definition is considered to be the grounding one in the field, defined content analysis as "a research technique for the objective, systematic and quantitative description of the manifest content of communication" (Berelson, 1952, p. 18). The aim of content analysis was to obtain an accurate description of a sample of communication. The emphasis was laid on the accuracy and reliability of findings – as with any other quantitative research method, such as questionnaire, structured observation, sociometry, and experiment. For this reason, in quantitative content analysis particular attention is paid to the control of sampling, objectivity of data collection and analysis, reliability of data, training of coders and statistical methods.

The development of content analysis has greatly been influenced by qualitative methodology which surfaced in 1970s. This methodology is based on ontological and epistemological beliefs that are considerably different from their quantitative counterparts. Thus, new types of investigation aims have emerged and they, in turn, have resulted in new types of research outcomes.

Content analysis was unable to resist the influence of qualitative methodology, so it adapted to it. However, the principles of qualitative methodology were so different from those of the content analysis that the latter

* Research Centre, Faculty of Humanities, Tomas Bata University, Zlín, Czech Republic. gavora@fhs.utb.cz

started to seek a new identity. In fact, content analysis is still struggling to find its unique identity (Dvorakova, 2010).

This methodological fermentation is obvious if we inspect the initial pages of some relevant books on contemporary content analysis. Authors pay much attention to carefully explain the definition and characteristics of content analysis. For instance, in their monograph Guest, MacQueen and Name (2012) define content analysis and describe its characteristics in 15 pages, Neuendorf (2002) does the same in 16 pages. In an anthology of texts on content analysis edited by Roberts (2009) a whole chapter is reserved for the definition and description of methodological characteristics. This interest in clarifying content analysis is in sharp contrast with the literature of other research methods, such as questionnaire, interview, or observation, where text sections providing definitions and descriptions are considerably shorter.[1] These latter research methods are stabilized, whereas content analysis is still under development.

Constitutive components of content analysis

In this paper we will present an in-depth examination of content analysis. In the literature content analysis is usually framed within two opposing methodologies, or paradigms: quantitative and qualitative (Elo & Kyngäs, 2007; Devi, 2009). Many authors provide descriptions for both by contrasting their properties (e.g. numerical data versus verbal data). Most frequently, mixed approach is added as a third option, which harmonizes the two different methodologies.

In our opinion this approach is rather simplistic, therefore, we propose a more sophisticated one instead by breaking down the methodology of content analysis to its *constitutive components* and then showing the range of possibilities these offer for different research scenarios. We admit that contrasting the two methodologies represents an appropriate didactical strategy adequate for teaching students the foundations of research methodology; however, it is not suited for the everyday research practice. As far as 'the mixed approach' is concerned, its connotation is rather delusive. Researchers do not mix methodological elements; rather they construct the research project by choosing the most suitable solutions for each constitutive component in a particular research project. A researcher, thus, acts as a constructor rather than as a cook. The product of this construction process is an original research project which, eventually, is accomplished in the researcher´s field work.

We will focus now on the description of the six constitutive components of content analysis, which are the following:

1.   types of content
2.   types of sampling
3.   depth of analysis
4.   direction of analysis
5.   types of control of coding
6.   modes of presentation of findings.

While describing these components we will rely on the relevant literature, first of all on the extensive monographs by Neuendorf (2002) and Krippendorff (2004), then on the chapter on content analysis in Berg's book (1995), a monograph by Guest, MacQueen and Namey (2012) and the overview of content analysis provided by

---

1.   We refer only to the definition and basic characteristics of content analysis, not the description of its components and processes. Obviously, these latter mentioned topics cover more pages.

White and Marsh (2006). Also, we will refer to the archetypical work of Berelson (1952), one of the founders of modern content analysis, and to Holsti (1969). Obviously, if necessary, we will also refer to other literature.

Our description of content analysis will culminate in a typology of content analysis variants which were drawn from the above literature. These variants will be presented in an overview table which shows how a particular author frames each constitutive component and how these components relate to each other.

Type of content

Obviously, the first question when dealing with the topic of content analysis is what content can be analysed (or is worth analysing). The term content analysis is frequently used without an object. This triggers the following questions: The content analysis of what? What is the object of content analysis?

Content can be defined as the matter which has been, or is being, communicated. It is a semantic entity, or a set of information, which is more or less structured. It describes reality, phenomena or processes. Anything which is communicated has content, it delivers a message. (Even an envelope which was sent empty by mistake delivers information – it communicates a message about the absent-mindedness of the sender.)

Because content always has a communicative function, it exists as a message.[2] It was created in order to be conveyed and delivered to somebody. Sometimes the content serves only to its author, like in the case of notes or diaries. Even these have a message, only that the addressees are the authors themselves. Sometimes, the communication, which was originally created as a personal document, becomes the object of content analysis by other people, such as researchers. For example, notes and diaries of authors of fiction or drama are often content analysed by critics, historians, linguists, etc.

As for the genres of messages, there is practically no limit or restriction. Originally, content analysis was mainly applied for the examination of media news. The expansion of content analysis is connected with World War II., when propaganda materials were often targets of analysis, so were published opinion reports. In 1941 Berelson defended his dissertation on the topic, and in 1948 he co-authored with Lazarsfeld a book entitled The Analysis of Communication Content. In 1952 Berelson published his book on content analysis, which became a respected monograph. From media analysis content analysis spread to other fields, such as psychology, where it was used for the investigation of personal characteristics or the analysis of the subjective meanings of words (Osgood, Suci, & Tannenbaum, 1957).[3]

Nowadays, content analysis does no longer focus solely on media messages; rather it covers all varieties of texts that are suitable for a particular research endeavour. In addition, content analysis also deals with non-verbal messages, for example images. Electronic communication has brought about the expansion of content analysis to almost any kind of electronic message, including the analysis of websites.

In the educational domain, a wide range of contents can be investigated which provide information on educational topics. Here is a brief summary of the existing source categories:

- Learning materials: books, textbooks, workbooks, collection of exercises, practice books.
- Other learning resources: maps, images, pictures.
- Students' products: essays, written assignments, homework, drawings, paintings.
- Bachelor´s and master's theses, doctoral dissertations.

---

2. Krippendorff (2004) prefers calling it communication. This expression, however, overlaps with the term 'communication' used for the process of communication.

3. Early history of content analysis is described in full in Neuendorf (2002), Franzosi (2007), Richards (2009), and Guest et al. (2012).

- Teachers´ products: notes, diaries, lesson plans.
- Documents: school legislation, school/class documentation, curriculum materials, educational programmes, evaluation materials, school codes.
- School presentation materials: bulletins, websites.
- School-parents communication: paper and e-mail messages.
- Media: articles in magazines, journals on educational topics.
- Radio programmes, TV news, TV shows, films, plays.

Some of these messages are *ready made*, i.e. they were not originally produced for educational investigation, but are fit for research purposes. Other sources are ad *hoc materials*, which were produced specifically for a certain research project. These products are collected and processed using a number of approaches.

With the introduction of qualitative methodology new textual materials started to be content-analysed. Originally, they were collected by qualitative data collection methods such as deep interview, semi-structured interview, narrative interview, focus groups interview, and unstructured observation. The voice recordings of interviews are transcribed and the transcript is used as the object of content analysis. Analogically, field notes from observations are formatted so that they constitute a text suitable for content analysis.

If we look at the preferences of the sample of authors included in the present study, only Berelson, (1952) and Holsti (1969) stick to *media content*. Neuendorf (2002), Krippendorff (2004), Berg (1995), Guest et al. (2012), and White and Marsh (2006) allow for any content. The different views of these two groups of researchers reflect the past developments of content analysis. Berelson and Holsti belong to the formation period when the range of content was rather narrow. The rest of the authors represent the modern era, during which, among other matters, qualitative methodology strongly influenced the focus of content analysis.

In the following sections of the paper we will narrow our perspective and focus predominantly on the content analysis of *text messages*. Non-verbal messages (e.g. images) or mixed messages (websites) have specific contents which require a special approach, and therefore, will not be considered in the present study.

Types of sampling

The process of establishing content analysis units is called *unitizing*. Three units of analysis can be discerned. Krippendorff (2004) distinguishes sampling units, recording/coding units, and context units. Similarly, Neuendorf (2002) defines sampling units, units of data collection, and units of analysis. Sampling units, as the wording reveals, are units selected for the research. They are selected from a population of texts. Recoding/coding units are units of data collection that researchers record and code from a particular sample of texts. Context units are "textual matter that set limits on the information to be considered in the description of recording units" (Krippendorff, 2004, p. 101). The relationship between the three units is often unclear, and sometimes they overlap, e.g. in the case of units of analysis and units of data collection.

There are two possibilities of selection of sampling units from texts. These are: (a) *stochastic* (probability) sampling and (b) *purposive* (intentional) sampling. Stochastic sampling aims to relate the data to the representativeness of the population. In purposive sampling "[…] decisions aim at the material which promises the greatest insight for developing theory, in relation to the state of theory elaboration so far (Flick, 2002, p. 64)". The third option is convenient sampling, which is, however, not a systematic way of sampling and is therefore suitable for pilot investigations only, but not for rigorous research projects. Convenient sample is a sample of 'take what you have', it is not, however, a random sample because the units were not taken randomly but arbitrarily.

9

From the point of view of sampling preferences, the authors discussed in the present paper form three groups. Those who incline to stochastic sampling are typically quantitative researchers (Berelson, 1952; Holsti, 1969; Neuendorf, 2002; Krippendorff, 2004). These authors favour stochastic sampling because only this 'corresponds to the standards of good science'. They opt for purposive sampling in rare – and well justified – cases. Those who choose purposive sampling favour qualitative research (Guest et al., 2012; White and Marsh, 2006). The third group is, in fact, represented by one author only, Berg (1995). According to Guest et al. (2012, p. 7) purposive sampling is employed typically in exploratory investigations, which are not guided by initial hypotheses. On the other hand, confirmatory research is organized so that it tests the initial hypotheses. Berg (1995), who is a symbolic interactionist, describes the steps used in the grounded theory technique, which is the flagship of qualitative research. However, this does not prevent him from promoting the use of both purposive and stochastic sampling.

Furthermore, the type of sampling affects the size of the sample. Purposive samples are in general smaller, but are carefully selected and thoroughly described as far as their social, cultural, and educational characteristics are concerned. Purposive sampling is usually used for the analysis of the latent structures of text (see section under heading: Depth of analysis), which requires larger amount of time to conduct than the analysis of the manifest structure. The latter usually relies on stochastic sampling.

The size of the sample in qualitative investigation is given by a simple rule: the researcher has to extend progressively the sample until reaching theoretical saturation. The researcher stops acquiring new data when no new codes or analytic categories appear and when the theory emerges (Strauss & Corbin, 1998). This procedure prevents the researcher from collecting redundant data once analytic categories have been saturated.

The size of the sample in quantitative investigation is determined by the statistical sampling theory. The size depends on the level of significance the researcher wants to achieve. The higher the significance, the larger the sample should be.

However, in contrast to other research methods, sampling in content analysis is not as simple as it may seem to be. Compare the sampling of respondents in a survey investigation with the sampling of images, e-mail messages, video recordings, or websites. The trouble begins with the fact that in many situations it is difficult to determine the population of these messages. Another obstacle in content analysis is that „sampling units and recording units tend to differ. Texts have their own connectivity, and recording units may not be as independent as the theory requires. Textual units tend to be unequally informative […]"(Krippendorff, 2004, p. 121–122). In such cases the researcher decides about the sample „so that the research question is given a fair chance to be answered correctly"(Krippendorff, 2004, p. 121–122).

Depth of analysis

The researcher has a choice to decide whether to analyse the manifest or latent content of a text. Manifest content is the visible content, the one which is on the surface of a text. Latent content concerns underlying meanings, the meanings which are anchored in the deep structure of a text. For Berelson (1952) and Holsti (1969) only the analysis of manifest content is acceptable. This is clearly shown by Berenson's definition of content analysis (Berelson, 1952, p. 18). As Krippendorff (2004) states, this approach indicates that content is contained in messages and waits for its discovery by an analyst. For this Krippendorff uses the metaphor of a container. "[…] the content is a tangible entity contained in messages and shipped from one place to another, that researcher presume to be analysed through the use of objective (i.e., observer-independent) techniques" (Krippendorff, 2004, p. 20). The level of independent interpretation of text content by analysis is rather low and thereby analysis is a

technical matter: meanings that were inserted in a text by its author and meanings derived from the text by a researcher are identical. This, of course, favours the reliability of text analysis (see section under heading: Type of control of coding). The manifest analysis concentrates on frequencies of words, phases, objects or signs.

In contrast to earlier views, the majority of present-day professionals hold a flexible position – to proceed as needed: either stay on manifest level or also include latent content in the analysis (e.g., Neuendorf, 2002). Qualitative researchers analyse only latent content (Guest et al, 2012; White & Marsh, 2006). In their 300-page-long book, Guest et al. make no mention whatsoever of manifest content.

The situation is, nevertheless, more complex than it might seem. Researchers have to start by analysing manifest content and then proceed to latent content, that is they shift from 'visible' to 'non-visible' content. Therefore manifest content must always be taken into consideration, since it represents the gate to latent content. More precisely, a researcher moves circularly between manifest and latent content.

Two or more researcher need not understand latent messages identically, thus, the results of their analyses may differ. Every analysis of latent content is a unique interpretative action, which is a characteristic trait of qualitative investigations. In a manifest interpretation the content is a container filled with some content. The author and the researcher have the same container and an identical content. In a latent interpretation, each content is unique, and it requires subjective interpretation from the researcher.

Some authors accept the priority of latent analysis but call for additional verification of its interpretation. For instance, Berg (1995), whom we introduced as a symbolic interactionist, stresses the importance of the examination of coding by independent coders, which is a compulsory procedure in quantitative content analysis. Berg also offers a more liberal standpoint: a possibility to support the interpretation by data: "[…] researchers should offer detailed excerpts from relevant statements (messages) that serve to document the researchers´ interpretations" (Berg, 1995, p. 176). He also proposes a rather arbitrary solution: "A safe rule of thumb to follow is the inclusion of at least three independent examples for each interpretation" (Berg, 1995, p. 176). This, of course, allows a reader of a research study to control the coding. The result of interpretation can be triangulated with other research methods, or similar content. This is a well-known strategy of qualitative investigation.

Direction of analysis

A key issue in the research process is the route from texts to findings. The direction of this route is considered to be a primary distinguishing mark of research philosophy, design, or methodology. Therefore, the researcher's position is often described early in the research project, together with the philosophy. Basically, there are two routes. The researcher moves from text (data) to findings (theory) or from theory to text. The first case is called *inductive process*; the other is the *deductive process*. The decision on which direction to take lies entirely on the researcher, though it may be strongly influenced by the other components of the research project, such as the research questions or hypotheses, sampling, focus on manifest or latent content, and type of control of data analysis.

Those researchers who hold a radical position accept only inductive progression and consider the deductive process to be inappropriate for content analysis. "This wholly deductive approach violates the guidelines of scientific endeavour. All decisions on variables, their measurement, and coding must be made before observations begin. In the case of human coding, the code book and coding form must be constructed in advance. In the case of computer coding, the dictionary or other coding protocol must be established a priori" (Neuendorf, 2002, p. 11).

The codes mentioned here are based on the theory of the particular research problem, on the researcher's field experience and discussion with colleagues. The analysis is a rather mechanical process of assigning the chunks of a text to a priori codes or categories. For Krippendorff (2004, p. 82) research is based on hypotheses, which implies a deductive process. The need to replicate the research in order to prove reliability and validity is also very strong in this case. Therefore, the research process should be described as minutely as possible to enable replication. Findings are repeatedly sustained by evidence, so they become solid. If they are disproved, the source of deviation is to be found in the primary theory, sampling, low reliability or validity or calculation errors.

The inductive approach represents the route from text to codes and from codes to theory. Codes are generated in the course of analysis and are the unique product of the researcher. Therefore, they are intransferable to other projects and are considered the result of the subjective interpretation of a text by the researcher. The induction aims to generate theory from data: "The investigator builds a theory that accounts for patterns that occur within a restricted research domain. The researcher then proceeds to investigate whether this theory still holds in another comparable domain, or whether it is to be modified to account for patterns in both domains" (Popping, 2000, p. 4).

The credibility of the researcher's interpretation is evidenced by judging the research procedures which are described in detail in the research report. This standpoint is represented by the views of Berg (1995, p. 184–185), [4] White and Marsh (2006) and Guest et al. (2012, pp. 7, 40).

Even though the steps in an inductive process are rather typical, the details of this process are unrepeatable. Therefore, the findings of two inductive coding on the same text usually vary. Comparing the results of two independent coders (or inter-rater reliability) in this framework is useless. We will return to this issue in the section entitled Type of control of coding.

However, the situation with induction is more complicated than it might seem. After the researcher has created the first codes, he/she uses them in further text sections. This is not exactly an inductive process. The established codes are applied in the consecutive sections of analysis. Thus, we have to be more subtle when talking about induction in research. The researcher proceeds in cycles: he/she returns to text segments that have been coded, and checks their meaning. If codes seem to be inappropriate they are modified or omitted. Strauss and Corbin (1998) support the cyclical movement from data to preliminary conclusions (hypotheses) and back, as well as they defend the position of finding evidence that verify, falsify or adjust the results.

Type of control of coding

When two researchers code the same text they usually do not reach to the same results. Not only that their codes may be different, but even the segments of text that are coded may vary (they code different segments). Why does this happen? First, a text is a complex body of meanings that everyone may understand in a specific way. This influences already the very first step of coding, namely the identification of text segments for coding. Some people code larger chunks of text, while others code smaller segments, which will result in uneven number of codes in a particular text (or density of codes). Another difference may originate from the different understanding of codes. The code that has been established and agreed upon is understood in different ways by the coders. So the four sources of disparity (size of segments, density of codes, interpretation of a segment, and interpretation of a code) may result in different research conclusions.

---

4.   Berg suggests the use of either the inductive or the deductive approach; however, he broadly describes only the technique of grounded theory, which starts with induction (he quotes Denzin, and Glaser and Strauss).

Quantitative content analysis solves this problem by carefully training coders to achieve a good agreement in coding. It should be noted that reliability in content analysis refers to coder consistency and not to data consistency (Popping, 2000, p. 132). To calculate intercoder reliability statistical methods are used. Coding and reliability are explained in detail by Popping (2000, p. 132-140). In his book Krippendorff (2004) has an entire chapter dealing with the selection and training of coders.

Advocates of quantitative content analysis (Berelson, Holsti, Krippendorff, and Neuendorf) claim that good reliability (0.80 or higher) is the basic precondition of a good content analysis. Reliability relates to validity. Though high reliability cannot ensure high validity, low reliability will lead to poor validity.

Advocates of qualitative content analysis do not use intercoder reliability, and many do not even mention this concept. Text analysis is performed on a deeper level and it results in original findings and conclusions. This leads, in their opinion, to a more valued outcome than a standardized, objective procedure. Qualitative analysts rely on the conformability and credibility of the process of investigation, of which the researcher provides thorough description (Lincoln & Guba, 1985). Richards (2009, p. 108) stresses the necessity to check the consistency of coding of *one* coder, rather than checking the inter-rater agreement. After the researcher has generated the system of codes, it must be used truthfully and consistently throughout the research project. This may be an obstacle if the project lasts over a longer period of time or if the aim is to analyse unusually large amounts of text.

Svaricek and Sedova (2007) presented an interesting peer-coding experiment. Each of them open-coded the same narrative text, and then they qualitatively examined the outcome. There was a surprising agreement in the number of codes (which may be a chance), but they only had three identical codes. The authors describe in detail the reasons of disparity. The basic difference between the codes was in the perspectives they took, which resulted in the discovery of new aspects and topics, rather than in the confirmation of a coding agreement. This is not a shortcoming but an advantage of qualitative inquiry.

For Altheide (1987) as well as White and Marsh (2009) the examination of reliability is not considered an essential issue. For Berg (1995) credibility is important, but he also stresses reliability, similarly to Guest et al. (2012, p. 85). According to Guest et al. identical coding is essential in a project conducted by a team of researchers to achieve consistent results. In large projects it is usual to hire professional coders. They, however, do not participate in data collection on the field, nor do they establish contact with the research subjects. Experiencing field authenticity is, however, an important factor in the interpretation of text. It was the primary factor that affected the individual perspectives in the example illustrated by Svaricek and Sedova (2007).

Research studies that deal with qualitative content analysis reveal a tendency of treating the data governed by issues of credibility. This is especially true of ad hoc texts, i.e. texts that were created for the purposes of research project. Such texts are transcripts of interviews, narrations, essays and the like. On the other hand, there is some indication that with ready-made texts – in spite of their qualitative orientation - coding should be checked by means of inter-coder agreement. This is, for instance, the case of school documents, media programmes etc.

Mode of presentation of findings

This component of content analysis shows the form or genre of the presentation of research results. There are three options: numerical, verbal or both. Of course, the presentation option is not an isolated element in content analysis. It is related to other components: research questions, sampling, control of coding etc.

Numerical presentation is employed when the research data is processed with statistical methods. The au-thor condenses large amounts of data into tables and schemes. The objects of quantification are words, phrases, topics, and the usual outcomes are sums, frequencies, less frequently averages and standard devi-ations. Relationships are manifested by coefficients. More advanced methods are: regression analysis, factor analysis, multidimensional scaling and cluster analysis. Since in quantitative research projects hypotheses are tested, the presentation focuses on their confirmation or refutation.

Qualitative presentation of results rests on verbal expressions. Findings are condensed but the final verbal condensation depends on the research concept and style preferences of the author. Guidance on writing up qualitative research findings can be found in Strauss and Corbin (1998), Flick (2002), and Guest et al. (2012). Usu-ally, the description begins with the presentation of key categories, central concepts, or the most important top-ics. These categories relate to the theoretical framework which is explained at the beginning of the research re-port, and of course, to research questions to which the project seeks to find answers. The report continues with the introduction of other categories, and the overall system is usually graphically displayed (schemes or tables). The description is supported with authentic quotes of data, which illustrate (but not substitute) theoretical con-cepts. Findings are related to the results of other research projects.

As far as the preferences for modes of presentation are concerned, the authors can be divided into three groups. Berelson, Holsti, Neuendorf and Krippendorff prefer the numerical style of presentation. Write and Marsh favour the verbal style. Berg and Guest et al. adhere to both modes.

Typology of content analysis variants

In the previous sections we have discussed the constitutive components of content analysis, their major charac-teristics and the relationships between these. Six constitutive components have been identified, which are the following:

1. types of content
2. types of sampling
3. depth of analysis
4. direction of analysis
5. types of control of coding
6. modes of presentation of findings.

Building on the literature discussed previously, in this section we will provide an overview of the constitutive components of content analysis, the main issues related to each component as well as describe possible config-urations. Our findings are shown in Table 1, which comprises authors, constitutive components and key issues associated with them.

Table 1
Configuration of variants of content analysis

Authors	Berelson (1952); Holsti (1969)	Neuendorf (2002); Krippendorff (2004)	Berg (1995)	Guest et al. (2012)	White & Marsh (2006)
Type of content	media	no restriction			

Type of sampling	stochastic		stochastic and purposive	purposive
Depth of analysis	manifest	manifest and latent		latent
Direction of analysis	deductive		inductive and deductive	inductive
Types of control of coding	inter-subject agreement		inter-subject agreement and credibility	credibility
Modes of presentation of findings	numerical		verbal and numerical	verbal

Table 1 shows that there are extreme variants of content analysis which showcase conflicting issues (Berelson and Holsti versus White & Marsh) and which in fact represent what in the traditional terminology is referred to as quantitative and qualitative approaches (paradigms). Those variants of content analysis which are situated between these extreme poles, and which are represented by four authors, show varied configurations. Only in one case (type of content) they are identical, in other cases they largely differ.

Let us have a look at *types of content* first. Berelson and Holsti focused predominantly on media news, which can be explained by their professional interests in the early history of content analysis. Other authors do not have any restrictions concerning the content, though in their own research they usually worked with only certain types of content (e.g. transcripts of interviews in case of Guest et al.). This constitutive component shows most clearly the developmental shift which occurred in the course of the development of content analysis methodology. Nowadays, the content analysed is not limited to verbal messages but also includes images, material entities, websites, and TV programmes.

As for the *sample,* three possibilities were identified. Besides the exclusive use of stochastic samples and the exclusive use of purposive samples there is only one author, Berg, who is inclined to use both sampling alternatives. This seems to be a rather odd choice – but only until we notice Berg´s solution regarding the depth of analysis. Again, he opts for two alternatives and suggests linking stochastic sampling to deductive processes, and purposive sampling to inductive processes. These both alternatives are methodologically meaningful.

In the *depth of analysis* component again we have found three alternatives: manifest, latent and both. Entirely manifest content and entirely latent content are opposites. Similarly to the types of content of messages and samples, they represent two extreme alternatives of content analysis. Berg is faithful to his broad conception of content analysis also in the case of this component, since he acknowledges both alternatives. By recognizing the analysis of latent content as scientific Krippendorff somewhat softens his rather positivistic approach to content analysis.

Direction of analysis is an essential constitutive component of content analysis. Deductive analysis employs the use of a priori set codes whereas inductive analysis involves code extracting from data. The former represents a top-down while the latter a bottom-up process. Berg is comfortable with both inductive and deductive approaches. Berelson, Holsti, Neuendorf and Krippendorff are strict deductivists. Guest et al. and White and Marsh represent the inductive branch.

Type of control of coding is related either to the preference for the use of inter-subject agreement (for checking the reliability of coding), or the acceptance of the subjectivity of the researcher (credibility, truthfulness, conformability). Guest et al. have a different perspective, since they offer two solutions. This can be explained by

the nature of the research they conducted, in which the coding process involved a team of researchers. There-fore, they had to ensure the uniformity of coding in order to obtain consistent results.[5] However, they acknowl-edged conformability too.

The analysis of the last component, *modes of presentation,* reflects a very similar picture to that of the previ-ous component. The presentation in form of numerical data corresponds to the research which aims at the gen-eralisation of findings, whereas verbal presentation represents the wish to show the perspectives of subjects. Again, Berg and Guest et al. propose both options, which is similar to their treatment of the control of coding. Others represent one solution only, either numerical or verbal.

The findings presented in Table 1 allow us to draw the following conclusions. If we accept the metaphor of *construction* for a research project involving content analysis, that is the construction from constitutive compo-nents rather than by mixing the elements of the content analysis method, then in case of each component the researcher has three alternatives options (except for the first constitutive component, sampling, where there are only two options). However, the researcher´s choice is limited in part. The contents of constitutive components are arbitrary only in some configurations, whereas the number of possibilities is limited, and illustrated in the ta-ble. It is not possible, for instance, to have a purposive sample, deductive analysis and verbal presentation of findings. At least, this configuration is not methodologically meaningful. Or, it is doubtful to have a stochastic sample, inductive analysis and numerical findings. In other words, the contents of constitutive components exist in certain configurations, or patterns, which the researcher has to conform to. The researcher cannot mix the content in whatever way he or she wishes. This supports our metaphor of constructing rather than mixing the el-ements of content analysis.

Discussion

In this paper we presented an overview of the current variants of content analysis which were identified by ex-amining their constitutive components. We used six such components, which can be utilized in two or three ways. Five configurations of these components and their contents were detected, which form five variants of content analysis. Two variants can be termed radical quantitative and radical qualitative content analysis, and three variants represent positions between these extreme alternatives, and they vary depending on the five out of six constitutive components. The conclusion is that a researcher has a range of options to combine elements of components of content analysis in order to create a reasonable research project.

From a chronological point of view we can notice a shift from radical quantitative positions, as represented by Berelson and Holsti (whose works were published in 1950s and 1960s), to soft quantitative positions in the works of Neuendorf and Krippendorff (which were published between 2000 and 2004). Berg and Guest et al. represent soft qualitative models, which were formed during a 10-year period, whereas White and Marsh repre-sent a hard qualitative position.

Our aim in this study was to show the variants within content analysis. The method we used was an exami-nation of a sample of works of the relevant authors. The authors were selected so that they represent a wide range of views. By no means have we indicated here that our overview shows the frequency of occurrence of the content analysis variants in the actual research. To obtain such data on frequency, one must examine a represen-tative sample of research studies.

---

5.   A similar approach was employed, for instance, in the research of Moretti et al (2011). In this case, however, the research did not only involve a team of researchers but also an international one.

References

Altheide, D. (1987). Ethnographic content analysis. *Qualitative Sociology, 10*(1), 65–77.

Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.

Berelson, B., & Lazarsfeld, P. F. (1948). *The analysis of communication content*. Chicago, IL: University of Chicago Press.

Berg, B. L. (1995). *Qualitative research methods for the social sciences* (2nd ed.). Boston, MA: Allyn and Bacon.

Devi, N. B. (2009). Understanding the qualitative and quantitative methods in the context of content analysis. *QQML2009: Qualitative and Quantitative Methods in Libraries*, International Conference, Chania Crete Greece, 26-29 May 2009. Retrieved from file:///C:/Users/Pouzivatel/Dropbox/CAnalysis/Literat%C3%BAra/Devi-Understanding_the_Qualitative_and_Quantatitive_Methods_PAPER-QQML2009.pdf

Dvořáková, H. (2010). Obsahová analýza / formální obsahová analýza / kvantitativní obsahová analýza. *Antropowebzin,* 2, 95–99. Retrieved from http://antropologie.zcu.cz/webzin/index.php/webzin/article/view/97/97

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107–115.

Flick, U. (2002). *An introduction to qualitative research*. London, UK: Sage.

Forman, J. & Damschroder, L. (2008). Qualitative content analysis. In L. Jacoby, & L. A. Siminoff (Eds.), *Empirical Methods for Bioethics: A Primer. Advances in Bioethics* (Vol. 11, pp. 39–62). Oxford: Elsevier.

Franzosi, R. (2007). *Content analysis: Objective, systematic, and quantitative description of content.* Retrieved from http://scholar.google.com/citationsview_op=view_citation&hl=sk&user=I5SYOqoAAAAJ&citation_for_view=I5SYOqoAAAAJ:0EnyYjriUFMC

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory. Strategies for qualitative research.* Chicago, IL: Aldin.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. Los Angeles, CA: Sage.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.

Krippendorff, K. (2004). *Content analysis. An introduction to its methodology* (2nd ed.). Thousand Oaks: SAGE.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Moretti, F., van Vliet, L., Bensing, J., Deledda, G., Mazzi, M., Rimondini, M., Zimmermann, Ch., & Fletcher, I. (2011). A standardized approach to qualitative content analysis of focus group discussions from different countries. *Patient Education and Counselling, 82*, 420–428.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: Urbana University Press.

Popping, R. (2000). *Computer-assisted text analysis.* London, UK: SAGE.

Richards, L. (2009). *Handling qualitative data* (2nd ed.). Los Angeles, CA: Sage.

Roberts, C.W. (2009). *Text analysis for the social sciences.* New York, NY: Routledge.

Shapiro, G. & Markoff, J. (2009). A matter of definition. In C. W. Roberts (Ed.), *Text analysis for the social sciences* (pp. 9–31). New York, NY: Routledge.

Strauss, A. L., & Corbin, J. M. (1998). *Basics of Qualitative Research: Grounded theory procedures and techniques* (2nd ed.). Newbury Park, CA: Sage.

Švaříček, R. & Šeďová, K. (2007). Technika audit kolegů. In R. Švaříček, & K. Šeďová (Eds.), *Učební materiály pro kvalitativní výzkum v pedagogice* (pp. 62–75). Brno: Masarykova univerzita.

White, M. D., & Marsh, E.E. (2006). Content analysis: a flexible methodology. *Library Trends*, *55*(1), 22–45.