# Phenomena-centered Text Analysis (PTA): a new approach to foster the qualitative paradigm in text analysis

**Thomas Krikser[1] · Benedikt Jahnke[2]** [ORCID]

## Abstract

Content or text analysis is one of the most common evaluation methods employed in qualitative research. Despite its wide application, however, a clear structure of how such evaluation should be conducted is often lacking due to the complexity of qualitative data. As a consequence, highly differentiated category systems with small-step subdivisions of categories and sub-categories are often used, leading to a loss of context both among categories and for the content as a whole. The aim of this paper is to describe the Phenomena-centered Text Analysis (PTA) as a novel form of qualitative text analysis, which takes these shortcomings into account by focusing on text-inherent phenomena. These phenomena are identified in two preceding quantitative analysis steps that identifying overlapping coding for subsequently qualitative analysis. We explain the structured code- and context-based approach of this new method and demonstrate its application with an empirical example. The PTA contributes to an increasing demand of qualitative methods especially for small-scale projects that need a structured kind of qualitative data analysis.

**Keywords** Content analysis · Text analysis · Qualitative research · Social research · Qualitative coding · Qualitative methods

## 1 Introduction

Qualitative Content Analysis is one of the methods most frequently used in qualitative research (Carrera-Fernández et al. 2014). The prevalence of this method is due to its wide-ranging applicability (including in the analysis of interview transcripts, audio-visual recordings of group discussions, and social media posts) and to the flexibility, systematicity and data-reduction possibilities inherent in this method (Schreier 2014). In spite of

✉ Benedikt Jahnke
jahnke@uni-kassel.de

Thomas Krikser
t.krikser@uni-kassel.de

1 Department of Management in the International Food Industry, Faculty of Organic Agricultural Sciences, University of Kassel, 37213 Witzenhausen, Germany

2 Department of Agricultural and Food Marketing, Faculty of Organic Agricultural Sciences, University of Kassel, 37213 Witzenhausen, Germany

these advantages, a basic problem of Qualitative Content Analysis is that there is often no clarity as to how exactly an evaluation should be undertaken. This lack of clarity arises primarily because qualitative research is highly dependent on the quality and quantity of the data. Furthermore, qualitative research questions cannot be answered in such a standardized way as is possible with statistical hypothesis tests. On the other hand, it is precisely the openness of this research method that offers opportunities for more in-depth examination and evaluation of the data collected.

The most common forms of Qualitative Content Analysis currently include various methods of Qualitative Content Analysis according to Mayring (2000, 2019) and Kuckartz (2014) and the Grounded Theory as developed by Glaser and Strauss (2009). While Grounded Theory offers a complex and methodically structured overall system of data collection, analysis and evaluation, its application is often too costly for small-scale projects in terms of time and financial resources. In contrast, the structured Qualitative Content Analysis proposed by Kuckartz (2014) constitutes a simplified evaluation model. The drawback of this type of analysis, however, is that in practice very many categories or sub-categories are often generated, so that in the subsequent data evaluation only small sequences are considered and the general context of what is said is lost. As a consequence, simple summaries are frequently presented as results which, depending on the number of categories, can easily appear confusing and incoherent. Due to the high number of unrelated small summaries generated in this way, qualitative analysis all too often loses a common thread, as well as the context and relationships. Nevertheless, all of them are important for a qualitative analysis that aims to uncover immanent connections, which are essential for their justification. The path of axial coding proposed in Grounded Theory could offer a way out of these difficulties, but as already mentioned, this method can be too far-reaching and economically unjustifiable. This applies especially for smaller-scale student projects or preliminary qualitative research stages (e.g. expert interviews) within a broader research project. (For an overview of the controversy regarding axial coding and Grounded Theory, see Kendall (1999) and Kelle (2007)). Furthermore, because of the openness for categories and sub-categories, structuring Qualitative Content Analysis can easily tempt researchers into crossing the border into *Quantitative* Content Analysis (Krippendorff 2013; Neuendorf 2017). For example, some researchers suggest that more than a 100 categories can be generated (Friese 2014; Saldaña 2009). Given such large amounts of categories, a researcher can easily become lost in the coding and no longer attain an overview of the structure and the context of the material (Richards 2006). Qualitative Content Analysis thus runs the risk of quantifying too much and thereby undermining its own strengths as an open, interpretative, and holistic analysis of content (Creswell 2018).

To overcome these risks and problematic aspects, we propose a new approach consisting of quantitative and qualitative steps. This approach incorporates different elements from Qualitative Content Analysis and Grounded Theory but develops a new practice-oriented form of data evaluation that should make it possible to use the strengths of the qualitative data while keeping time and financial expenditure within acceptable limits.

## 2 Theoretical assumptions

It can be assumed that by coding a text with categories relations between the applied categories exist. If these categories are considered and analyzed separately the relationships between the categories can get lost. Nevertheless, these relationships create the general

meaning and the context of a text. Therefore, it is necessary to capture and interpret these relationships between categories. In the context of PTA these relationships are captured by a quantitative approach as a first step in order to be analyzed qualitatively afterwards. The quantitative approach follows the assumption that categories that appear frequently in an articulated context also share a proximity regarding the content. This context of meaning builds the foundation for further analysis in PTA.

The overlap of two or more categories that appear in different contexts of meaning is described as a phenomenon in PTA. Within a phenomenon all categories that belong to the phenomenon have to have a strong proximity of context and content among each other. By merging these categories to a phenomenon, the manifest statements in a text can more easily display their latent relationships. This leads to a new structure of the content and thus, to a reduction of the units that have to be analyzed. Since categories can pertain to more than one phenomenon the multidimensionality of a category can be captured and presented in the light of the adequate context. By analyzing these phenomena, the broader context of association between the categories and the latent context of meaning is included in the qualitative analysis. This seems to lead to a deeper understanding of the internal structure of the data.

## 3 Preparation

Phenomena-centered Text Analysis (PTA) is a method of text analysis that bridges Quantitative Text Analysis (Ahuvia 2001; Neuendorf 2017; Riffe et al. 2019) and Qualitative Text Analysis (Kuckartz 2014; Mayring 2000; Schreier 2012) and includes minor aspects of Grounded Theory (Charmaz 2006; Corbin and Strauss 2015; Glaser and Strauss 2009). PTA is a systematic way of describing the central meanings of qualitative material and can be used to identify the most important patterns behind texts. The focus of this method is on text documents obtained through semi-structured interviews (Flick et al. 2007) such as expert interviews, although narrative interviews and group discussions might also be analyzed via PTA. In accordance with the method of Qualitative Text Analysis presented by Kuckartz (2014), PTA does not prescribe a specific type of data collection and therefore the sampling can be orientated to the particular research question. Depending on the research question, theoretical sampling (Charmaz 2006) but also typical case sampling, extreme case sampling and other qualitative sampling approaches can be applied (Bryman 2012).

PTA is a code-based method. As shown in Fig. 1, the analysis starts with the development of categories. Such analytical categories are at the center of the research. According to Kuckartz (2014), categories can be obtained deductively or inductively or by a mixture of deductive and inductive approaches. In a deductive approach, categories are developed on the basis of theoretical aspects that emerge in the course of the literature review and
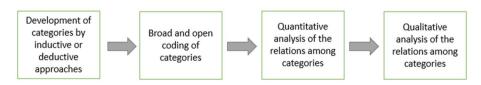


**Fig. 1** General approach of Phenomena-centered Text Analysis (PTA)

desk research. Typically, these categories are the categories found to be most salient to the research question. In an inductive approach, the categories used for the text analysis are derived out of the data, entailing a very close and careful reading of the entire text while noting down any categories that might be relevant for the subsequent analysis (Mayring 2000). In PTA, both deductive and inductive ways are possible and can be combined. In practice it often happens that a number of crucial categories are identified during the desk research and then, because qualitative analysis is an explorative approach, additional categories emerge during the research. Both forms of category development can thus be undertaken in a sequential manner.

The coding process for PTA is similar both to the process of open coding in Grounded Theory (Charmaz 2006; Glaser and Strauss 2009) and to the process of coding in thematic Qualitative Text Analysis (Kuckartz 2014). The difference in PTA, however, is that the number of categories used in this method is limited in two ways. First, PTA requires broad and open coding that is not based on words or sentences but on the context of meaning. For example, it may happen that several paragraphs are coded continuously with one or more categories. Since PTA aims to foster a qualitative paradigm of openness to new theories and a holistic perspective on the meanings of text, it requires the researcher to keep a constant eye on the *context* of data. Categories in PTA should therefore be broad, standing for a wide range of topics that belong to this category and should not be further divided into sub-categories. For example, if we use the category 'purchasing site' in analyzing shopping behavior, we should not divide this category further into 'supermarkets', 'shopping malls' and 'small and private retail shops' but confine ourselves solely to the category of 'purchasing site'. This means that PTA only works with main categories. Second, since the aim is to undertake a close analysis of the data, we need to identify phenomena that will help us understand the meaning of the text. In order to retain the big picture, we thus recommend using only a limited number of main categories. From our work on some research projects we have found that a number of 8–10 main categories seems a good compromise between the necessary number of categories and the goal of keeping the approach qualitatively context-based. Given these two limitations on the number of categories, the definition used to describe the categories should be a very general one that allows for open qualitative analysis in the subsequent steps.

## 4 Analysis

PTA involves three steps: a quantitative part that consists of two steps, and a qualitative part that interprets the quantitative findings. The quantitative part is split into a quantitative analysis of the category relations within the interviews or groups of interviews (Step 1) and a quantitative analysis of the category relations within the coding (Step 2).

### 4.1 Quantitative analysis

*Step 1: Category relations within interviews*

The first step consists of a simple but relevant counting of the number of categories per interview. This should be shown as matrix, with the interviews displayed in columns and the categories and the total number of codings displayed in rows (see Table 1). The tabulation of findings in this way is not only useful for analysis but also helps the reader gain a direct and structured overview of the amount of codings per category per interview.

**Table 1** Category matrix

| Category system | Interview 1 | Interview 2 | Interview 3 | Interview 4 | Interview 5 | Interview 6 | Interview N |
|---|---|---|---|---|---|---|---|
| Category A | 17 | 31 | 21 | 15 | 9 | 10 | … |
| Category B | 25 | 42 | 19 | 12 | 7 | 12 | … |
| Category C | 8 | 12 | 9 | 21 | 31 | 18 | … |
| Category D | 6 | 8 | 3 | 25 | 41 | 23 | … |
| Category E | 2 | – | – | – | 1 | – | … |
| Category F | – | 1 | – | 1 | – | – | … |
| Category N | … | … | … | … | … | … | … |
| Total | 58 | 94 | 52 | 74 | 89 | 63 | … |

There are two main reasons for conducting such quantitative analysis of the category relations within interviews. First, it allows the researcher to identify any categories that seem negligible for further analysis. Especially if the categories have been generated through deductive category development, we can identify differences between the data analyzed and the current state of the literature. Often a research project has a field of interest that is more specific than the general theory on the research topic, or the research is located in one country that differs in some aspects from other countries that have already been analyzed in the literature. For example, Table 1 shows results that might occur after coding interviews with a deductive category system. In this example the categories E and F were only mentioned by a limited number of interviewees. Although these categories are relevant for general discussion about the research topic, they do not fall within the particular area of interest of the research. This difference should be explained either theoretically or by further research in the field. Even if the category had been obtained inductively, this first step gives a hint as to which categories are negligible. Depending on the research question and the research area, limited coding can be seen as a first result of the quantitative step and these categories should be analyzed in a separate step independently of the further analysis of the phenomena.

Second, this quantitative analysis step is important to control for any relevant differences in the weighting of categories among the interviews. These differences can arise due to the setting of the sampling process or as a consequence of interviewing different groups in society. For example, a researcher might interview three different groups, such as a group of scientists, a group of interviewees from the political administration and a third group from NGOs. If the interviews differ in the relation of categories among the interviewees and subgroups can be defined, the further analysis should use these subgroups. As another example, researchers might undertake a consumer study in which interviews are conducted with people who have vegetarian or vegan diets and thus differ from people with an omnivore diet. In this case, two subgroups could be necessary to avoid losing the specific views of vegans and vegetarians in the results of the analysis. If such heterogeneity is detected, it can be assumed that there are different clusters and that they should be analyzed separately. Therefore, the interviews should be divided into relevant subgroups and the second step of the analysis be applied to each subgroup.

In Table 1, such differences in the amounts of codings within the interviews are made clearly visible. While in interviews 1, 2, and 3, the highest number of codings can be found for the categories A and B but less for C and D, it is the other way around in interviews 4,

5, and 6. Based on this finding, two subgroups should be used for the subsequent analysis, with Subgroup 1 consisting of interviews 1, 2, and 3, and Subgroup 2 consisting of interviews 4, 5, and 6.

*Step 2: Category relations within codings*

Step 2 is undertaken for the purpose of identifying relevant phenomena for the qualitative analysis to be conducted in step 3. A 'phenomenon' here is defined as a strong relation between two or more categories. To identify such phenomena, we use a distance-based approach. Depending on the data, cross-sectoral phenomena can also be identified. Both methods can be applied using Computer-Assisted-Qualitative-Data-Analysis-Software (CAQDAS) such as RQDA or MAXQDA. In a project, both methods should be used for identifying phenomena within the complete data or within subsets.

*Step 2.1: Distance-based identification of phenomena*

For the identification of phenomena through the distance-based method, we use a matrix of overlapping codings. Many computer programs for the analysis of qualitative data include this option, such as MAXQDA (Code-Relations-Browser) and RQDA (Cross-Codes). For this analysis it is necessary to only include those coding relations that have overlaps, and are not just in the same paragraph. Overlaps in this sense include exact overlaps and the inclusion of a coding in another. Most programs provide a matrix like the one presented in Table 2. In such a matrix we count the amount of overlaps between two codings. For the identification of phenomena, we concentrate on the most relevant coding relations. It has been shown (Krikser 2013) that relevant relations can be found by taking the highest amount of overlaps and concentrating on all overlaps that are higher than half of the maximum overlapping. Since in some cases we have still lost a relevant relation in this way, however, we consider a small deviation of 10% of the maximum number of overlaps as a good and reasonable compromise to allow for some variance that can occur in qualitative interviews and coding. The following formulae can thus be used to calculate relevant coding relations:

Formulae 1 and 2:

$$PI_{50} \geq \frac{\max O}{2} \quad PI_{40} \geq \frac{\max O}{2} - 10\% \max O$$

In formula 1, the phenomena identifier (PI) is equal to or higher than half the maximum amount of overlaps of two variables (max O) ($PI_{50}$). It is also possible to use a PI that is

**Table 2** Distance-based analysis of phenomena

| Category system | Category A | Category B | Category C | Category D | Category E | Category F | Category G |
|---|---|---|---|---|---|---|---|
| Category A | – | 7 | 1 | 2 | | 11 | 2 |
| Category B | 7 | – | 13 | 10 | | 7 | 6 |
| Category C | 1 | 13 | – | 8 | 1 | 4 | 6 |
| Category D | 2 | 10 | 8 | – | | 5 | 6 |
| Category E | | | 1 | | – | | 1 |
| Category F | 11 | 7 | 4 | 5 | | – | 3 |
| Category G | 2 | 6 | 6 | 6 | 1 | 3 | – |

equal or higher than half of the maximum overlaps observed (as in formula 2), including a deviation of 10% of the maximum overlaps (max O) ($PI_{40}$). Unfortunately, it is not possible to give a definite statement for the optimal PI, since it depends on the quality and the quantity of the data. Based on tests with different kind of data and based on experience we recommend to start with $PI_{50}$ and to carefully look at the data. If important categories get lost, it should be considered to compare the phenomena of the $PI_{50}$ with phenomena from $PI_{40}$.

Whether a $PI_{40}$ or $PI_{50}$ should be used depends on the research question and on the data being analyzed. While $PI_{40}$ allows for the inclusion of more categories in the final phenomena, it can also happen that some interrelations appear that are not relevant for the research question. By contrast, $PI_{50}$ uses a stricter criterion that excludes more categories from the final results. Accordingly, $PI_{40}$ is recommended for smaller projects with a lower number of interrelations while $PI_{50}$ might obtain better results for bigger projects that are easier to interpret.

These phenomena identifiers are used in the next step to identify relations between more than two categories. For this identification we first look at the highest amount of overlaps between two categories. Following this identification, we need to ascertain whether there is a category that also has phenomena identifiers with both categories from the highest overlapping. If such a third category is identified, we then look for a fourth category, and so on. If we find two or more categories that are related with each other by phenomena identifiers, we then have a phenomenon within our data that we can analyze in a qualitative manner in step 3. We repeat this procedure with all the phenomena identifiers to find different phenomena in the data.

Table 2 shows an example of the overlaps among 7 categories. The highest amount of overlaps can be found between categories C and B, with 13 overlaps. Based on this, the PI could be calculated as follows:

$$PI_{40} \geq \frac{13}{2} - 13 * 0.1 \quad PI_{50} \geq \frac{13}{2}$$

In this case the $PI_{40}$ is higher than 5.2, hence all categories with 6 or more overlaps will be included in the subsequent analysis. The $PI_{50}$ would be > 6.5 and would only include categories with 7 or more overlaps. A closer look at Table 2 reveals that category G has three overlaps with a number of 6 and none with 7 or higher. With $PI_{50}$ we would therefore lose this entire category for further analysis. In this case, a more accommodating approach using $PI_{40}$ seems more appropriate and thus would be used in this example.

Starting with the relation of B–C, category D has PI with B (10) and C (8). The combination B–C–D again has PI with G (6–6–6). The category F cannot be part of this phenomena, since there are PI with B, but not with C, D and G. Therefore, a phenomenon can be identified in the combination of B–C–D–G. Other phenomena could be found in the combination A–B–F. Based on these combinations, a Qualitative Content Analysis should be conducted that analyses the meanings and associations of each combination.

*Step 2.2: Analysis of cross-sectoral categories*

It can happen that several categories belong to more than one phenomenon and thus can be identified as cross-sectoral categories. These categories should be extracted in the results and used for further discussion of the relevance of these categories to the research objectives. In the example in Table 2, category B can be identified as such a cross-sectoral category.

## 4.2 Qualitative analysis

*Step 3: Qualitative analysis of phenomena*

Once the phenomena have been identified in the first two quantitative steps, a qualitative analysis is conducted. In this analysis we concentrate on the coded text in each of the categories that are comprised by a phenomenon. By using the text retrieval function of a CAQDAS, the coded text for each category can be extracted. In some CAQDAS it is also possible to obtain an output with the overlapping sequences of several categories. Since the aim is to capture the whole phenomenon, however, each coding must be taken into consideration. The goal here is to develop a micro-theory about the contexts in which the categories that are part of a phenomenon are associated with each other. These micro-theories should not be too broad, since the description should be concentrated only on the categories that are included in the identified phenomena.

Having elaborated a micro-theory for each phenomenon, we then take a closer look at potential cross-sectoral categories. To explain the meaning of these categories, we thus need to look at the associations of the phenomena. Additional desk research might be necessary to make statements about the meaning and importance of the cross-sectoral categories. These cross-sectoral categories are further used as a bridge from the micro-theories about the phenomena to the development of a broader macro-theory that addresses the research question.

## 5 Assuring the quality of the approach

As recommended by most textbooks on Qualitative Content and Text Analysis (e.g. Kuckartz 2014; Schreier 2012), intercoder and intracoder reliability should be tested to verify the quality of the coding process in PTA. However, it must be taken into account that broad and open coding can lead to a greater variance in the accuracy of the match between two codings. For this reason, it is recommended that a small margin of tolerance be allowed for the exactness of overlaps. In a software-based analysis, an accuracy of 80% seems sufficient to assume double coding with two different categories. This does not affect the alpha score that is decisive for the quality assessment of intercoder and intracoder reliability. Here a value greater than 0.8 should be obtained (O'Connor and Joffe 2020). Furthermore, it is important to ensure that exact agreement on the coding units is reached within the research team in advance, since the use of broad coding in meaningful contexts has so far been the exception rather than the rule.

In addition to the calculation of intercoder and intracoder reliability, the creation of text memos also plays a decisive role in quality assurance in PTA. This is because in quantifying the codings according to their overlaps in the first and second steps of quantitative analysis, individual categories can be excluded from further analysis even though they are highly relevant to the research focus. Such exclusion arising from the quantitative analysis steps can be counteracted by appropriate text memos, as known from Grounded Theory (Glaser and Straus 2009) in order to allow for the analysis of such text passages at a later time independently of the PTA.

**Table 3** Code-matrix-browser in MAXQDA11

| Codesystem | (002) I... | (003) I... | (004) I... | (005) I... | (006) I... | (007) I... |
|---|---|---|---|---|---|---|
| Professionalization | 7 | 6 | 5 | 5 | 5 | 1 |
| Projects | 7 | 7 | 7 | 6 | 8 | 8 |
| Public relations | 7 | 6 | 4 | 5 | 4 | 4 |
| Donations/Gifts | 10 | 5 | 4 | 8 | 8 | 6 |
| Motivations | 1 | 2 | 1 | | 3 | 1 |
| Board Members | 7 | 6 | 6 | 7 | 5 | 4 |
| Relation to Community | 10 | 8 | 7 | 4 | 7 | 5 |
| Competitors | 2 | | | | | 1 |
| Networks | 2 | 2 | 3 | 1 | 2 | 1 |

## 6 Examples of how to conduct phenomena-centered text analysis

In this section we present an example from a research project of how to conduct PTA, providing an example for each of the three steps of PTA. Since we focus here on an example with no subsets, we also provide a second example in the appendix of a sample from which we could identify subsets.

The example is used both for the quantitative and the qualitative parts of the analysis in order to show how to work with PTA. Although the focus is on the quantitative steps, we also provide a brief explanation of the qualitative step. Readers seeking further guidance on the qualitative analysis should refer to the literature cited in this paper.

### 6.1 Project without subsets

*Step 1: Category relations within interviews*

The example we use is taken from a small-scale project about community foundations in Germany, where six interviews were conducted with representatives of small community foundations (Krikser 2013). A deductive approach was used for the development of the categories, with 8 main categories deduced from a review of the literature. Inductively, the further category of 'Motivations' was added. Table 3 shows the Code-Matrix-Browser generated by the qualitative data analysis tool MAXQDA (Kuckartz and Rädiker 2019; VERBI Software 2014).

The results of this quantitative analysis show that a similar distribution of codings was obtained from each interview. This indicated that we did not need to divide the sample into different subgroups. In the subsequent analysis, therefore, all of the interviews were analyzed together.

A second finding from Table 3 was that two of the main categories we had deduced from the international literature about community foundations were not applicable in the case of German community foundations. While the literature revealed that the categories of 'Competitors' and 'Networks' play a major role in discussions about community foundations in other countries, especially in North America, the interviews conducted in Germany for this project revealed these categories to be negligible. This finding may be attributed to the fact that community foundations had only been introduced in Germany comparatively recently. A result like this should be mentioned briefly in the quantitative analysis of the

**Table 4** Distance-based phenomena identification using MAXQDA11



category relations among the interviews and should later be described and explained in the qualitative analysis in the subsequent step of PTA.

*Step 2: Category relations within codings*

In step 2 the identification of phenomena is explained. Based on the example from the community foundations project, we can generate Table 4 through the Code-Relations-Browser in MAXQDA (VERBI Software 2014).

From the output displayed in Table 4 we can calculate the phenomena identifiers. The highest amount of overlapping codings are found in the relation between the categories 'Projects' and 'Public relations' with a total of 13 overlaps (see subscript 1 in Table 4). Based on this number, the phenomena identifiers (PI) are defined as 6 or higher:

$$PI_{40} \geq \frac{13}{2} - 13 * 0.1$$

For the first phenomenon we use the interrelations between 'Projects' and 'Public Relations' as a starting point. We then look for the next category that has the highest number of overlaps with either 'Projects' or 'Public Relations', i.e. the category 'Donations/Gifts', which has 10 interrelations with 'Projects' and 8 with 'Public Relations' (see subscript 2 in Table 4). We can therefore add 'Donations/Gifts' to the 'Projects' and 'Public Relations' categories.

The next categories we should examine are those of 'Professionalization' and 'Board Members', since both of these categories have 7 interrelations with 'Projects'. Given that these two categories have no PI with 'Public Relations' and that the category of 'Professionalization' also has no PI with 'Donations/Gifts', neither of these categories will be added to the phenomenon. The final category we need to check is 'Relation to Community', which has 6 interrelations with all of the 3 categories already added to the phenomenon (see subscript 3 in Table 4). On basis of these results, we have thus identified the first phenomenon as 'Project–Public Relations–Donations/Gifts–Relation to Community'.

A second phenomenon can be identified in the interrelations of 'Professionalization' with 'Board Members' and 'Projects'. A third can be identified with 'Board Members', 'Donations/Gifts' and 'Relation to Community', and a fourth with 'Board Members', 'Projects', and 'Relation to Community'. A fifth phenomenon consisting of only two categories can be found in the overlap between 'Relation to Community' and 'Networks'. These five phenomena will therefore be included in the next step of qualitative analysis.

For the identification of cross-sectoral categories we can use a table (such as Table 5) that shows the frequency with which these categories appear in the phenomena. Although the qualitative analysis should consider any categories that occur in more than one phenomenon, not all categories will be relevant for the research question. The

**Table 5** Frequencies of categories within phenomena

| Category | Frequency |
| --- | --- |
| Relation to community | 4 |
| Projects | 3 |
| Board members | 3 |
| Donations/gifts | 2 |
| Public relations | 1 |
| Professionalization | 1 |
| Networks | 1 |

relevant categories should be described in terms of their interrelation with different phenomena.

*Step 3: Qualitative analysis of phenomena*

In step 1 we identified the two categories 'Competitors' and 'Networks' as negligible for the purposes of this research project. Both of these categories were obtained by deductive category development, since the literature review showed that many researchers in the international discussion on community foundations agree that a larger non-profit sector leads to a larger number of competitors (Saxton and Benson 2005). This leads to a stronger dependence on the non-profit market and, as a further consequence, to an increase in donor control (Eikenberry 2009; Ostrander 2007). To tackle these challenges, these non-profits are compelled to establish long-term partnerships and networks in order to achieve stability in their revenue structures (Harrow and Jung 2011; Knott and McCarthy 2007; Paarlberg and Meinhold 2012).

To explain why these findings did not apply to the German community foundations that took part in the project it is necessary to know that while community foundations have a long tradition in other parts of the world it was only in 1996 that the first community foundation in Germany was founded and that it took several years before this new non-profit model was established (Hoelscher and Hinterhuber 2005). As Saxton and Benson (2005) have shown, although competitors and networks often become relevant for community foundations in the years after their establishment, in their first years of existence they have different priorities and other challenges to tackle. This explanation can be related to the fifth phenomena, which shows that 'Networks' are only based on a local level and are not well established.

The following five phenomena were identified by our quantitative analysis in step 2:

Phenomenon 1: 'Projects–Public Relations–Donations/Gifts–Relation to Community'.

Phenomenon 2: 'Professionalization–Board Members–Projects'.

Phenomenon 3: 'Board Members–Donations/Gifts–Relation to Community'.

Phenomenon 4: 'Board Members–Projects–Relation to Community'.

Phenomenon 5: 'Relation to Community–Networks'.

In the qualitative third step of PTA each of these five phenomena need to be explained. Most CAQDAS include a retrieval function that is able to sort coded text according to the categories used for coding. Through this function, quotes can be retrieved from the data that can be used to explain the interrelation between the categories in the identified phenomena.

The implication behind the first phenomena we identified (i.e. 'Projects–Public Relations–Donations/Gifts–Relation to Community') is that each of the interviewees mentioned that finding new donors was the most important task for their community foundation. Aside from their general goal of achieving good project results, they saw their most important challenge in convincing current and potential donors in the community to contribute to future projects. Accordingly, their main focus was on issues likely to elicit positive responses in local media. Although the interviewees acknowledged that there were other more virulent issues and challenges, they therefore chose to support projects with high visibility (Krikser 2013).

In the next three phenomena it was the exceptional role of the board members that are responsible to create a professional monitoring and evaluation process for the project (Phenomenon 2), to get in contact to the community by their private networks to recruit new donors (Phenomenon 3), and to give their name for the projects because there are known as serious members of the community (Phenomenon 4). In this specific case we have a cross-sectoral category that has a high influence on a variety of phenomena. Therefore, the role of board members can be used later in the discussion to bridge the observations with the Theory of Change and the role of a champion for small non-profit organizations (Thompson et al. 2006).

## 7 Discussion and conclusion

PTA has been presented as a new method intended to supplement the existing repertoire of qualitative analysis methods by enabling qualitative data analysis to be adapted to the reality of many research projects, including small-scale projects. The decision for or against the use of PTA as an analysis method needs to be made on an individual basis, taking into account the special knowledge interest and all other project-related factors such as time, money, personnel, etc. For example, a solely explorative study undertaken for the primary purpose of attaining familiarity with an unknown research field would seem less suitable for PTA than a theory-based content analysis of semi-structured interviews. Additionally, the kind of text can have an influence on the applicability of PTA. Since the PTA is based on associations between codings it is necessary that the text offers the possibility for these associations. A narrative interview therefore is more suitable for using PTA than a legal text or a result log.

Furthermore, in relation to structured process of data analysis within the PTA, it should be noted that not all content-relevant categories necessarily overlap with other categories. For example, it would be possible, at least theoretically, to imagine that an individual category might 'float around' unconnected to other categories in the Code-Relations-Browser. Effective memos should therefore be written during the initial research stages to ensure that any such 'floating' categories that might be important for understanding the content and its integration in the wider context are not 'lost' in the PTA. The same applies to all those categories that fall below the PI threshold but which should nonetheless be included in the interpretation on the basis of preliminary theoretical considerations.

In general, it should be emphasized with regard to PTA that the mere frequency of a category or the number of overlaps with other categories does not by itself allow conclusions to be drawn about the quality of content. The two quantitative analysis steps involved in PTA carry a certain risk of slipping into quantitative content analysis. Here

it is important for researchers to be reflexive and alert to this tendency and to perceive the quantitative steps as a way of pre-structuring the content analysis in accordance with the qualitative paradigm.

Notwithstanding these limitations and risks, there are a number of strengths associated with PTA that make it an important addition to the existing repertoire of research methods. One of these benefits is that only a small number of phenomena need to be described instead of a multitude of individual categories. This is not only a great advantage in terms of the time and financial costs of research but also makes it possible to prepare results in a more compact and dense form for publications. Another advantage of PTA directly related to the smaller number of phenomena to be described is that the contexts and the relationships among the categories are much better preserved by this method than in conventional category-based content analysis. This makes it easier to identify micro-theories and combine these micro-theories into larger overall theories. Overall, PTA strengthens the qualitative paradigm through its open and holistic approach.

Recognizing that the development of any method of analysis is an iterative process, this paper has aimed primarily to provide a theoretical framework to enable the application of this new method to empirical projects. On the basis of our discussion, we call on researchers to use their own empirical data to contribute to the fine-tuning of our proposed method by demonstrating the wide range of possible applications of PTA to texts beyond interview data and thus ultimately to arrive at an empirically based validation of this method.

## Appendix 1: Phenomena-centered text analysis (PTA)

### Example of PTA with subsets

In this example we present some preliminary results of a research project on consumer trust in organic food. In the first stage of this project, interviews were conducted with consumers to attain an initial impression of different levels of trust in different aspects of the supply chain of food. For the purpose of this example, we use only the first eight interviews that were analyzed during this project.

As shown in Table 6, the 'packaging', 'freshness' and 'seasonality' of the product were only mentioned by a few consumers, while the terms 'regional' and 'organic' were coded quite often. This finding is in line with the findings of other studies of food consumption

**Table 6** Category matrix

| Category system | Int1 | Int2 | Int3 | Int4 | Int5 | Int6 | Int7 | Int8 |
|---|---|---|---|---|---|---|---|---|
| Packaging | 1 | | | 1 | | | | 1 |
| Information | 2 | 2 | 3 | 1 | 1 | 2 | 1 | 2 |
| Freshness | 1 | | | | | 1 | | |
| Purchasing site | 4 | 9 | 7 | 2 | 4 | 4 | 4 | 4 |
| Seasonality | 1 | | 1 | 1 | | | | |
| Regional | 2 | 4 | 5 | 4 | 1 | 4 | 4 | 3 |
| Seals | | 3 | 1 | | 5 | 2 | 1 | |
| Trust | 5 | 8 | 2 | 2 | 7 | 5 | 4 | 2 |
| Organic | 3 | 4 | 5 | 1 | 10 | 3 | 1 | |

behaviour and can thus be addressed as negligible categories in the discussion and conclusions of the study.

The table also shows, however, that three of the interviewees only mentioned 'organic' once or not at all. Given that the focus of the project was on trust in organic food, we split the sample into two groups according to their answering behaviour in relation to organic food consumption. Accordingly, for the next step the interviews are split into the following two subsets: a subset of five interviews showing an affinity for organic food (Interviews 1, 2, 3, 5, and 6), and a subset of three interviews showing no affinity for organic food (4, 7, and 8). In the next step we analyze the category relations within codings, as shown in Table 7.

The $PI_{50}$ in the first subset is > 6 and the $PI_{40}$ is > 4.8. In this case there is no difference between $PI_{40}$ and $PI_{50}$. There are three phenomena in this subset:

Phenomenon 1: 'Trust–Purchasing Site–Organic'.

Phenomenon 2: 'Trust–Seals–Organic'.

Phenomenon 3: 'Trust–Regional–Purchasing Site'.

As cross-sectoral categories, we can identify 'Trust' (3), 'Organic' (2), and 'Purchasing Site' (2).

In the second subset the $PI_{50}$ is > 3 and the $PI_{40}$ is > 2.4. There is thus also no difference between the PIs. As shown in Table 8, there is only one phenomenon in this subset:

Phenomenon 1: 'Trust–Regional–Purchasing Site'.

This phenomenon can also be identified in the first subset.

From this example we can argue that trust in local purchasing sites with regional food offerings is very high both for consumers with an affinity for organic food and those

**Table 7** Distance-based analysis of phenomena in subset 1

| Category system | Packaging | Information | Freshness | Purchasing site | Seasonality | Regional | Seals | Trust | Organic |
|---|---|---|---|---|---|---|---|---|---|
| Packaging | | | | | | | | | |
| Information | | | | 4 | | 2 | 2 | 4 | 2 |
| Freshness | | | | | 1 | 2 | | | |
| Purchasing site | | 4 | | | | 6 | 2 | 12 | 12 |
| Seasonality | | | 1 | | | 2 | | | |
| Regional | 2 | 2 | | 6 | 2 | | 1 | 7 | 4 |
| Seals | 2 | | | 2 | | 1 | | 8 | 7 |
| Trust | 4 | | | 12 | | 7 | 8 | | 11 |
| Organic | 2 | | | 12 | | 4 | 7 | 11 | |

**Table 8** Distance-based analysis of phenomena in subset 2

| Category system | Packaging | Information | Freshness | Purchasing site | Seasonality | Regional | Seals | Trust | Organic |
|---|---|---|---|---|---|---|---|---|---|
| Packaging | | | | | | 1 | | | 1 |
| Information | | | | 1 | 1 | | | 1 | |
| Freshness | | | | | | | | | |
| Purchasing site | | 1 | | | | 6 | | 4 | 1 |
| Seasonality | | 1 | | | | | | | |
| Regional | 1 | | | 6 | | | | 5 | 2 |
| Seals | | | | | | | | 1 | |
| Trust | | 1 | | 4 | | 5 | 1 | | |
| Organic | 1 | | | 1 | | 2 | | | |

without such an affinity, while the former group also have a high degree of trust in 'organic food stores' and 'organic seals'.

## Declarations

## References

Ahuvia, A.: Traditional, interpretive, and reception based content analyses: improving the ability of content analysis to address issues of pragmatic and theoretical concern. Soc. Indic. Res. **54**(2), 139–172 (2001)

Bryman, A.: Social research methods. Oxford University Press, New York (2012)

Carrera-Fernández, M.J., Guàrdia-Olmos, J., Peró-Cebollero, M.: Qualitative methods of data analysis in psychology: an analysis of the literature. Qual. Res. **14**(1), 20–36 (2014)

Charmaz, K.: Constructing grounded theory: a practical guide through qualitative analysis. SAGE, London (2006)

Corbin, J.M., Strauss, A.L.: Basics of qualitative research: techniques and procedures for developing grounded theory. SAGE, Boston (2015)

Creswell, J.W.: Qualitative inquiry and research design: choosing among five approaches. SAGE, Los Angeles (2018)

Eikenberry, A.M.: Refusing the market. Nonprofit Volunt. Sect. Q. **38**(4), 582–596 (2009)

Flick, U., Kvale, S., Angrosino, M.V., Barbour, R.S., Banks, M., Gibbs, G., Rapley, T.: The sage qualitative research kit. SAGE, London (2007)

Friese, S.: Qualitative data analysis with ATLAS.ti. SAGE, Los Angeles (2014)

Glaser, B.G., Strauss, A.L.: The discovery of grounded theory: strategies for qualitative research. Aldine, New Brunswick (2009)

Harrow, J., Jung, T.: Philanthropy is dead; long live philanthropy? Public Manag. Rev. **13**(8), 1047–1056 (2011)

Hoelscher, P., Hinterhuber, E.M.: Von Bürgern für Bürger? Bürgerstiftungen in Deutschlands Zivilgesellschaft. MAECENATA, Berlin (2005)

Kelle, U.: "Emergence" vs "forcing" of empirical data? A crucial problem of "grounded theory" reconsidered. Hist. Soc. Res. **19**, 133–136 (2007)

Kendall, J.: Axial coding and the grounded theory controversy. West. J. Nurs. Res. **21**(6), 743–757 (1999)

Knott, J.H., McCarthy, D.: Policy venture Capital. Adm. Soc. **39**(3), 319–353 (2007)

Krikser, T.: The potential of German community foundations for community development. Maecenata Institut für Philanthropie und Zivilgesellschaft, Berlin (2013)

Krippendorff, K.: Content analysis: an introduction to its methodology. SAGE, Los Angeles (2013)

Kuckartz, U.: Qualitative text analysis: a guide to methods. Practice and Using Software. SAGE, Washington DC (2014)

Kuckartz, U., Rädiker, S.: Analyzing qualitative data with MAXQDA. Springer International Publishing, Cham (2019)

Mayring, P.: Qualitative content analysis. Forum Qual. Sozialforschung/Forum Qual. Soc. Res. (2000). https://doi.org/10.17169/fqs-1.2.1089

Mayring, P.: Qualitative content analysis: demarcation, varieties, developments. Forum Qual. Sozial-forschung/Forum Qual. Soc. Res. (2019). https://doi.org/10.17169/fqs-20.3.3343

Neuendorf, K.A.: The content analysis guidebook. SAGE, Los Angeles (2017)

O'Connor, C., Joffe, H.: Intercoder reliability in qualitative research: debates and practical guidelines. Int. J. Qual. Methods. (2020). https://doi.org/10.1177/1609406919899220

Ostrander, S.A.: The growth of donor control: revisiting the social relations of philanthropy. Nonprofit Volunt. Sect. Q. **36**(2), 356–372 (2007)

Paarlberg, L.E., Meinhold, S.S.: Using institutional theory to explore local variations in united way's community impact model. Nonprofit Volunt. Sect. Q. **41**(5), 826–849 (2012)

Richards, L.: Handling qualitative data: a practical guide. SAGE, London (2006)

Riffe, D., Lacy, S., Watson, B.R.: Analyzing media messages: using quantitative content analysis in research. Routledge, New York (2019)

Saldaña, J.: The coding manual for qualitative researchers. SAGE, Los Angeles (2009)

Saxton, G.D., Benson, M.A.: Social capital and the growth of the nonprofit sector. Soc. Sci. Q. **86**(1), 16–35 (2005)

Schreier, M.: Qualitative content analysis in practice. SAGE, Washington DC (2012)

Schreier, M.: Qualitative content analysis. In: Flick, U. (ed.) The SAGE handbook of qualitative data analysis, pp. 170–183. SAGE, London (2014)

Thompson, G.N., Estabrooks, C.A., Degner, L.F.: Clarifying the concepts in knowledge transfer: a literature review. J. Adv. Nurs. **53**(6), 691–701 (2006)

VERBI Software: MAXQDA. Berlin (2014) Retrieved from https://www.maxqda.de