Edith Cowan University

Research Online

1-1-2020

Let the data speak: Using rigour to extract vitality from qualitative data

Alistair J. Campbell
Edith Cowan University, alistair.campbell@ecu.edu.au

Let the Data Speak: Using Rigour to Extract Vitality from Qualitative Data

Alistair J Campbell
Edith Cowan University, Perth, Australia
alistair.campbell@ecu.edu.au

Abstract: Qualitative data can be gathered from an array of rich sources of research information. One of the popular ways to collect this data is by interviewing a range of experts on the topic, followed by transcription, resulting in a database of written documents, often supplemented by other documented data that informs the topic. Thematic or Content Analysis can then be used to explore the data and identify themes of meaning that enlighten the research topic, with the themes being gathered into nodes. The researcher now has an array of nodes, which needs to be organised into a coherent model, and more importantly, one that represents the views of the research informants. To do this with some degree of rigour, the researcher needs some way of ranking the nodes in terms of their relative importance. The node ranking can be based on experience, or on the literature, but neither of these approaches looks to the data itself. If the database contains new or unexpected knowledge, neither experience nor the literature will guide us to it, and vital new insights may easily be missed. The framework outlined in this paper aims to provide a sound first-cut analysis of the data, based on the evidence in the research interviews themselves. Clearly the literature and research experience have an important role to play in shaping the results of any research. However this paper argues that one should proceed only after the data itself has been offered "the first chance to speak".The node classification matrix detailed here, identifies distinct node categories, each ranging in significance and with particular characteristics that reveal key aspects of the informants' views. In this way the researcher can use the nodes to reveal the voice of the experts, and build a scientifically rigorous set of results from a qualitative database.

1. Introduction

1.1 Qualitative Approach:

Qualitative research is quite different to a quantitative or statistical research approach which typically rests on a finite range of answers to questionnaires, or hypotheses that are put to a sample of numerical data. These statistical results may indicate a correlation, and occasionally causality, among variables that are thought to be fundamental to the investigation. The quantitative approach is however limited to the confirmation of known concepts, and assumes that we know enough about the phenomenon being investigated to ask probing questions, and pose credible hypotheses. Statistical results may therefore add to extant knowledge by polling a range of known phenomena in new geographical or social environments.

However if we are dealing with a topic that is exploring new ground, then a set quantitative questions will be unlikely to uncover anything unusual. Davidsson (2005) puts it this way: "events that are infrequent, unanticipated or extraordinary… phenomena of this kind may be difficult to capture with conventional, quantitative approaches".

This is where the open-ended nature of qualitative research, and semi-structured interviews in particular, begin to show value (Alasuutari 2010). For any interview there will of course be a topic of focus, and this is what gives it structure. However a skilled researcher will allow the respondent considerable leeway to explore the topic in their own terms (Denzin and Lincoln 2013). In this way, particularly through multiple interviews, and other sources of qualitative data, a new body of knowledge can be created by probing the experiences and insights from a range of experts in the field.

Qualitative approaches are particularly useful when investigating research topics that study creativity, where the research method needs to accommodate the unexpected, or surprises, and aspects that might be novel or even unique (Davidsson 2005, Yin 2018). A quote often attributed to Einstein asserts that: 'No problem can be solved by the same kind of thinking that created it' (Bohm and Peat 2000). In this sense, and particularly where creativity or innovation is being explored, the research approach needs to be open to accept answers that are outside a pre-conceived framework of questions. Only then can new knowledge be added, novel insights be garnered, or innovative models be proposed.

Pursuing the qualitative route however involves complexity, as qualitative data comes in many forms; for instance Yin (2018, pg.114) lists six possible sources of qualitative data: documentation, archival records, interviews, direct observations, participant-observations, and physical artefacts. An important point to make is that a mix of these sources can be used to inform a single investigation, and having multiple sources naturally strengthens the rigour and depth of the database (Yin 2014, Fabregues and Molina-Azorin 2017).

Also, when seeking to uncover new knowledge using qualitative approaches, rather than confirming existing constructs using statistical methods, the so-called Meno's Paradox comes into play. Fine (2014, pg.8) paraphrases this paradox as follows:

> *But how will you inquire into this, Socrates, when you don't at all know what it is? For what sort of thing, from among those you don't know, will you put forward as the thing you're inquiring into? And even if you really encounter it, how will you know that this is the thing you didn't know?*

Clearly in many cases we know enough to at least begin a research enquiry. However the central issue remains – if our knowledge is incomplete, how will we know what questions to ask, and will the answers expand our knowledge? Mir (2018) argues that this boils down to little better than an educated guess (which is undoubtedly preferable to an uneducated one), but the questions are still largely guesswork. It is therefore in the focussed (but unfettered) expert knowledge that the potential value lies, and it is up to the researcher to unearth and present this value.

Qualitative data is therefore rich (Yin 2018, Yang 2013, Neergaard and Ulhoi 2007), and extracting just the relevant themes from the database can be a challenge, requiring focus, dedication, skill, and repeatability from the researcher. Nowell et al (2017) point out that Thematic Analysis is often used as a basis to guide this exploration, given that it is a relatively easy concept to understand, especially for early career researchers or those that are new to qualitative approaches.

Nowell et all (2017) goes on to set out a framework to improve trustworthiness in Thematic Analysis, while noting that it is one of few contributions in this field. This paper will therefore seek to add to this general aspect of rigour as well. However the qualitative researcher still faces another important hurdle.

Once a collection of relevant themes has been amassed, it is vital to understand the hierarchy among them – are some themes more relevant than others? There are a number of ways this hierarchy could be achieved – using the researcher's experience could be a way, but the question of impartiality would be the first issue to arise. Another option would be to base the hierarchy on extant models or constructs in the literature – in this case the question is whether the research is then just destined to discover what is already known? Crucially, neither of these options leverages the unique power of qualitative research approaches to uncover the new, the unexpected, or the surprising.

This paper will therefore seek to develop a coherent and repeatable framework for determining a hierarchy in the identified themes in the data, by using attributes that are found in the database itself, and therefore reflect the value of the expert knowledge that has been gathered into that database.

1.2 Aim and Objectives

To summarise the above introductory discussion, the Aim and Objectives are stated separately here:

Aim:
This research will seek to develop a rigorous framework for researchers using qualitative methods, that supports the extraction of coherent, novel and validated results from the gathered data.

Objectives:
- Leverage off the ubiquitous and extant nature of thematic and content analysis, using software search tools and concepts from phenomenology to augment thoroughness, meaning and rigour when identifying themes.
- Develop a framework that analyses the relative value of themes identified in the database, to reflect the value inherent in the expert knowledge in the data.

- Illustrate the developed framework using real-world examples.

1.3 Working with Qualitative Data

This paper is therefore focused on qualitative research, and as many experienced researchers can attest, in some fields of exploration this still runs against the current (Yang 2013; Mir 2018; DeGregorio, Tagliafico and Verde 2018). Mir (2018) has some strong views on this, asserting that "splitting empirical research in the social sciences into "qualitative" and "quantitative" camps is an act of discursive and epistemic violence". Likewise Yang (2013) describes her experiences as "Surviving as a qualitative researcher in a quantitative world". While quotes like these may not typify the majority of qualitative researchers, it is certainly enough evidence to be mindful that the status quo may include critics of qualitative approaches.

It is therefore important that the qualitative methods of analysis are rigorous, and can support results and conclusions that are both valuable and robust (Neergaard and Ulhoi 2007, Fabregues and Molina-Azorin 2017).

This paper explores a framework that can be applied to a wide range of qualitative data, with a particular focus on documented data from multiple sources, and these could be emails, interview transcripts, promotional material, or other written documents.

1.4 Thematic and Content Analysis – a search for Meanings

Some research makes a distinction between Thematic and Content Analysis (Crowe, Inder and Porter 2015; DeGregorio, Tagliafico and Verde 2018) although the differences can be quite subtle at times, with a large degree of overlap and similarity (Vaismoradi et al 2016). The framework outlined here is a qualitative analysis of the documents, and will consider the terms Thematic Analysis and Content Analysis interchangeable, thus siding with those authors that refer to Thematic Content Analysis as a single concept (DeGregorio, Tagliafico and Verde 2018; Braun and Clark 2006).

Returning to the focus of this paper, in essence we are searching for meanings that the interviewee (or document author) is intending to communicate, and in this context it is useful to borrow the concept of meaning-units from the field of phenomenology (Giorgi 1985, 2009). This helps emphasise that it is the meaning that is being sought (not the words), and so "meanings have words" rather than the other way around. Since an array of possible words might be used to convey the same meaning, it is up to the researcher to determine which words in the database might point to a particular meaning being communicated. This underlines the strong argument for including "a human in the loop" when employing software analysis tools (Michaud 2017, Basit 2003). The use of words as "search-clues" to identify key meanings, is elaborated on later under the heading "meanings not words".

With this human intelligence in focus, we then move on to identify the meanings in the qualitative database. Here there are two possibilities.

- First, the researcher could be looking for a particular attribute or construct that is identified in extant theory, and the aim is to ascertain whether that theoretical construct is found in the context under scrutiny.
- The second possibility is that the research project is exploratory, and here the aim is to search the data to see what possible meanings and themes emerge.

An example of the first instance is exploring the 5-dimensional construct of Entrepreneurial Orientation (EO) among managers. EO has dense definitions for each of the 5 dimensions (Lumpkin and Dess 1996, 2001; Dess and Lumpkin 2005) and these can be used to search the interview data in a rigorous and well-defined manner.

The second instance is exemplified by interviews with a range of successful entrepreneurs, to explore and describe the paradigms which guide their thinking, their decision-making, and new venture creation.

In both instances we are searching for meanings, and to do that we need words; words that typify the meaning that we suspect might be in the database. The second, exploratory case is a bit more difficult, but finding potential search-words can be aided by reading and re-reading the transcripts (Corbin and Strauss 2015, Strauss and Corbin 1998).

1.5   Coding the Themes, Meanings

Coding can be described as the process of 'mining' the database (Corbin and Strauss 2015) to extract value (meanings), and then placing them in distinct piles (nodes), that all share a common meaning. Nodes can also be regarded as 'containers' for the separate themes that are identified from the data.

Sometimes the researcher is the interviewer as well, and in such cases the data coding can be informed by the data gathering, but typically the sheer volume and richness of the data will mean that this analysis is largely at an intuitive level. For example, the research for which this methodology was originally developed had 35 individual transcripts and documents as its database, comprising around 200,000 words.

Manual coding of nodes is still practiced by some experienced researchers. Basit (2003) for instance, describes each node of meaning in the database being assigned a different colour pen, and marked up accordingly. This however limits the number of identifiable themes to the range of available colours (Bengtsson 2016). Clearly this is not ideal, and only suited to smaller qualitative databases.

In contrast to intuitive or manual analyses, this paper agrees with those like Basit (2003) and Bengtsson (2016) that advocate the use of software analysis tools such as NVivo® and many others. First, this helps guide a systematic process that allows the full extent of the data to be analysed comprehensively (Alasuutari 2010), and thus bolsters the aspect of rigour which is especially important in qualitative research. Second, this also ensures a repeatable and verifiable procedure, which again reinforces rigour, while using a well-known and trusted method such as thematic analysis. And thirdly, using software to search the database, sometimes totalling hundreds of pages, reduces fatigue in the researcher, and therefore increases the likelihood that the research will accesses more of the informants' vitality that is recorded in the qualitative database (Michaud 2017).

2.   Let the data speak

This is an especially helpful phrase in qualitative research data when seeking to reveal the nuances of new and novel insights. Having gone to the trouble of gathering a range of expert views, it makes sense to explore that value as best we can. Essentially this aim can be achieved by focussing on three main aspects:

- the research informants (respondents)
- meanings not words
- classifying nodes

The word "informants" is one borrowed from Yin (2018) and is a useful synonym for the respondents in a research project, in that it emphasises that these people are there to inform the research, by adding perspectives and insights that are unknown to the researcher at the beginning of the project.

Qualitative databases typically include interviews recorded as transcript documents, and when searching through the database for coherent threads of meaning (nodes), it is also useful to emphasise that the search is for "meanings not words". This will be explained in the next section.

When classifying nodes it is essential to allow the most significant nodes to "emerge from the data" rather than placing an emphasis on nodes that is based on factors external to the data (Vaismoradi et al 2016). This requires a framework which analyses the database itself, and identifies how well each of the nodes are represented in that database. Once this hierarchy of nodes is understood, one can proceed to relate the nodes to one another, and then also to extant theory and practice, to form a coherent model.

2.1   Transcripts to Nodes

2.1.1   *Meanings not Words*

As mentioned before, when searching for nodes the concept of "meanings not words" is useful especially when a variety of informants or sources are used. The concept of "meaning units" (Giorgi 1985, 2009) was introduced previously. By focussing on the meaning during the analysis, rather than the words, this helps to minimise the drift of meaning across a diversity of contexts in the research. In this way the qualitative researcher can bolster the all-important aspect of academic rigour, when using the ubiquitous method of

thematic analysis. Meaning units therefore become important in extracting coherent meaning from a range of inputs, by helping to identify the different words that interviewees may use to describe one particular meaning (meaning unit).

Take for instance the word "integrity" which typically describes an ethical approach. However in a research project one informer used the same word "integrity" to convey a completely different meaning – to denote "integrity" (cohesion) among multiple aspects of new venture creation. To complicate matters, ethics was also one of the aspects of the research question being investigated. Simply accepting the word "integrity" as a valid contribution to a node (theme) would have resulted in confusion, and nonsense, despite the power of sophisticated computer-aided tools. To maintain academic rigour, these sorts of verbal acrobatics require careful attention from the researcher (Michaud 2017), and "meaning-units" is therefore a useful concept which makes it possible to triangulate qualitative evidence rigorously, despite a diversity of sources in the database.

Using this approach of "meanings-not-words" consistently will therefore allow the researcher to identify a number of significant and coherent meanings within the database, and to code these as a number of individual nodes.

While a simple word-search is often used, backed up by the researcher's scrutiny for the correct meaning, software analysis tools typically include other more complex search tools as well. One example is a proximity-search for two words that form a single meaning, such as "strategic marketing". The search will then return those passages where the two words are in close proximity, for instance in the same paragraph, or a certain number of words apart. So the meaning could be expressed in the database as "marketing but with a strategic approach" which would be a valid inclusion in that node of meaning, but would not be picked up by a simple word-search for "strategic marketing".

However one of the most important features of the search-tools in analysis software, is their ability to sort through a large volume of diverse documents, with different file types, in one search action. Once the search is complete, it also records a score for that search, indicating the number of passages identified, how many documents contained this theme, etc. As observed before, these tools also do not suffer the human frailty of researcher fatigue, and will faithfully search through hundreds of pages of documents, accurately identifying every occurrence of the search-item. This again helps to bolster the rigour of qualitative research by reducing researcher fatigue and human error, and by increasing the scope and repeatability of the search process.

By using a combination of software searches and attention to the meanings of each theme, the researcher therefore can maintain rigour while assembling a collection of nodes from the database, using the familiar and dependable basis of thematic or content analysis.

Once the researcher has a collection of identified nodes, the next question becomes:
- Among the many nodes of meaning identified, which are the most significant?

This is where the Matrix Sorter comes into play.

2.2   The 2x2 Matrix

The 2x2 matrix was first used by the Boston Consulting Group (BCG) to categorise a portfolio of products within a market context (Henderson 1970). This matrix divided the products into 4 groups: cash cows, stars, pets, and question-marks, which then allowed firms to follow a different strategy for each product type, and thus maximise their profits.

This 2x2 matrix tool has since been used across a range of other contexts (Lowy and Hood 2011), as an aid to strategic, organisational and personal decision-making. Some examples are the 4-quadrant "I'm ok, you're ok" analysis, the prisoner's dilemma which presents 4 outcomes of co-operating or not, the sorting of teams into 4 distinct types, and using the virtue matrix to guide corporate responsibility.

The essence of this conceptually simple 2x2 matrix is that two key elements are combined, to form four distinct quadrants, each with a particular character. The power of the matrix derives from the choice of the two elements – it is essential that these two are central to the issue being analysed (Lowy and Hood 2011). The

matrix then examines a combination of scenarios where the two key elements are alternately assumed to be high or low. For instance the BCG focused on market-growth and market-share, so high market-growth and high market-share was clearly the quadrant to be in, and those products were identified as "stars".

By using these two key elements, the 2x2 matrix thus considers 4 distinct scenarios, and giving each one a name helps to guide how each sceanrio is considered. The complexity of a 4-way decision-making process is thus sequenced, as each of the four scenarios is considered individually, without being confused or clouded by the other 3 alternatives.

The relevance of the matrix to a range of applications can be attributed on the one hand to its largely uncomplicated structure, and on the other to its strong persistence in the business world. Madsen (2017) puts it this way: "The BCG Matrix is simple and powerful, in much the same way as other strategic tools and frameworks such as SWOT analysis and Porter's Five Forces framework". This certainly locates the matrix in familiar territory, and within a business context. Madsen (2017) also analysed critics of the matrix and noted: "the review of criticism levelled at the BCG Matrix reveals a certain level of scholarly disdain". The article goes on to point out that this academic criticism occurs "in spite of [the matrix] being relevant and widely used by practitioners… [and] still commonly referred to in marketing and strategy textbooks, as well as in MBA education". (Madsen 2017).

Given the above, and especially the underpinning analysis provided by Lowy and Hood (2011), the 2x2 matrix is considered a sound and appropriate basis for this paper that is located in business research methods.

2.3 Applying the 2x2 matrix to node significance

The next step is to apply this 2x2 matrix to analyse the relative significance of the nodes that have been identified in our qualitative data. It is important to emphasise that this approach explores what our experts considered to be more significant, or less so. Organising the nodes into a hierarchy dictated by a particular theory is perhaps convenient, and may be well-received in that it fits within a prevailing paradigm. In some instances this may be appropriate, depending on the research design.

However the power of a qualitative research approach is that it can reveal unexpected aspects of the topic under investigation – facets that were not necessarily predicted or included in extant theory (Davidsson 2005). In this way the existing body of knowledge in that field is fleshed out, added to, and gradually better aligned with real-world observations.

To examine the relative significance of nodes we need two key parameters (Lowy and Hood 2011) that identify the ranking that our experts placed them in. There are any number of parameters that qualitative analysis software can display regarding each of the nodes or themes that have been identified in the database. These include the number of characters, words, paragraphs, passages, documents, references or files that are stored at each node, and each software package uses different terms for these parameters. To narrow down the discussion, three main parameters will be examined, using the labels found in NVivo®.

For each node (theme), the software typically records the following parameters:
- References
- Files
- Coverage

First, the number of references refers to the individual passages that were found to match the theme that has been coded at that particular node. The number of references is irrespective of the source of the passage, so it could be from interviews, promotional material, annual reports, etc. Each time any passage in the database refers to the theme of that node, the number of references for that node increases.

In terms of node significance, the more references that match the theme of that node, the more it was mentioned in the database of documents, and therefore the more significant the node becomes. This is clearly a key parameter in determining the relative significance of the nodes in the data.

Second is the parameter called files. In most databases, a separate file is created for each data source (informant), and if not, it is relatively easy to organise the collected data into an individual file for each data source. Data files could be interview transcripts, company documents, field notes, and so on. The number of files in a particular node will indicate how many different data sources mentioned the particular theme that is contained at that node. If only a few files are found at a particular node, then only a few of the data sources have mentioned the theme at that particular node. In contrast, if many of the data sources have mentioned that theme, then many files will be counted at that node. Again, in terms of significance, having many files at a particular node will indicate that many of the data sources (e.g. interviewees) made mention of that particular theme. This would identify that particular node as more significant than nodes with a low number of files, and that the theme (node) has a diversity of support among many of the research sources. Files is therefore also a key parameter in deciding the hierarchy among our nodes.

Third, the parameter of coverage is expressed as a percentage, and refers to the proportion of a particular document that is devoted to the theme at that node. In a short document, the coverage for a particular node (theme) would typically be quite high, since most of the document consists of just that node. In contrast, a lengthy interview might mention a particular theme many times, but it will be buried in amongst the introductory parts of the interview, the interviewer's questions, talking about the weather, and other bits of chit-chat. The theme might be very eloquently expressed, many times over, and therefore coded at that node each time it occurs, but the coverage will be quite small, since it makes up only a small percentage of the overall interview.

In the case of the longer interview, the theme is significant, but the coverage percentage is low. In contrast, in a short document the coverage is high, as the passage containing the theme is a larger percentage of the document, but without indicating that the theme (node) itself is any more significant than in the lengthy interview. It is therefore evident that coverage is not a parameter that has a straightforward link to the relative significance among a collection of nodes.

2.3.1 Two key factors for Node Significance

Following the discussion above, the two major factors to emerge which determine node significance are therefore:

- the number of References at that node
- the number of Files at that node

Before proceeding, it is useful to reiterate that in this article the NVivo® descriptors are used as the default. The following summarises the two key parameters and provides examples of different words that are used to denote the same concepts.

- **References** – an instance where the particular meaning is referred to. Alternate descriptions for References are:
1. Passages
2. Citations

- **Files** – a record in the database from a specific person or source – with each independent source being recoded in a separate file.
  Alternate descriptions for Files are:
1. Sources
2. Documents
3. Interviewees

These two key parameters can now be used to evaluate the relative significance among the nodes (themes) in the data. Typically the results for each node will be listed in a table, for instance:

Table 1

Node	References	Files	notes
First node	20	15	
Second node	80	18	most significant
Third node	51	12	
Fourth node	15	5	least significant
…etc…			

The number of references can also be expressed as the number of passages coded to that node, the number of "mentions" that a theme received, or the number of times that the theme at that node was cited in the research data.

The number of files can also be stated as the number of different respondents, or sources of information, in which the theme for that node is identified. For instance if most of the interviewees mention a particular theme, then the theme (node) can be regarded as widespread among the experts interviewed for the research.

2.4 The Matrix Sorter – a Hierarchy of Nodes

These two key indicators of relative node significance, are now placed into the 2x2 matrix format in Fig.1. The number of references is located at the top axis of the matrix, and the possible outcomes are that the node will have either few or many references.

Likewise for the number of files identified at each node, there might be many files at each node, or only a few, and this is indicated on the left axis of the matrix.

	Few References	*Many References*
Many Files	**Widely Held View**	**Well Supported**
Few Files	**Weak Support**	**The Committed Few**

Figure 1: Node Significance

Combining these possible outcomes using the matrix therefore gives us 4 distinct scenarios. Each one is given a description that refers to its significance.

The two extreme scenarios are more readily understood, beginning with the scenario in the bottom left quadrant.

In this case the node has few references and also few files, and it is clear that this theme has only weak support from the data that we collected. Not only was this meaning very seldom mentioned, it was also quoted by very few of our sources, and the description Weak Support is therefore apt.

Conversely, the other extreme in the top-right quadrant, identifies those nodes that are referenced many times, and that also have many files which indicates that many different sources in our database identified this theme. Clearly this quadrant contains the most significant nodes in our database – hence the description Well Supported. So these first two node classifications are literally at opposite ends of the scale, and are vital indicators in determining node significance. It is important to emphasise at this point that this difference in significance arises, not from theory, not from the literature, not from existing models, but as signalled by the expertise represented in the database.

The two remaining quadrants are mixed scenarios. Starting at the bottom-right, these nodes have many references, but relatively few files, and this requires a bit of thought. In this outcome only a few sources in the data mentioned the theme, but those that did mention it, were quite enthusiastic about it, and mentioned it many times over. This scenario is characterised as "The Committed Few".

As an example, in a particular research project the issue of ethics was found to be important to certain respondents. However less than half of the interviewees mentioned this aspect, but those that did mention it, emphasised it by repeating it many times over. This type of node classification therefore identifies meanings that are somewhat out of the mainstream, but nonetheless very important to a key subset of the informants. These nodes are quite different to the other classifications, and can often indicate potential sources of new research directions or sub-divisions.

Lastly, the top-left quadrant gathers those nodes that have a large number of different files, but with relatively few references. Again this requires a bit of thought, as the theme was clearly identified by many different sources in the database. However these sources did not refer to the theme (node) very often. An example of this type of node is where "everybody knows". For instance in new venture creation, innovation is a given, and there is no need to repeat it over and over, although nearly all the respondents mention it at least once or twice. This type of node is therefore characterised as a "Widely Held Position". Again, this classification is quite distinct from the others, and offers the researcher yet another option when considering the relative significance of the nodes that have been identified.

This now gives us four distinct categories for our nodes:
- Well Supported
- The Committed Few
- Widely Held View
- Weak Support

Each of these scenarios identifies nodes that are have quite different levels of significance. It is also not a simple winners-losers ranking either – it is a recognition that all of these four categories have a different role in the way they convey meaning to our research project.

For instance one might be tempted to discard the last option of Weak Support as trivial, since it identifies only low endorsement from the informants. However particularly given the rich and complex nature of qualitative data, it can be equally valuable to know what is unimportant, as this can eliminate distraction by fringe-aspects. In this way our qualitative research can remain focused and rigorous, despite the sometimes distracting richness of the data.

2.5   Examples: using the Meanings and Matrix Sorter in practice

2.5.1   *Searching for Meanings*

First, during the word-search phase there is a need for words that typify the meaning we are looking for. One example is the concept of Risk-taking – does the respondent show instances where risk is acknowledged and managed? The obvious choice is the word "risk" itself, and by using this short four-letter version in the search, the software will also identify related extensions such as "risky" and "riskiness". As outlined before, it is then up to the researcher to determine which of the passages containing that word, also fit with the meaning being sought. This can be thought of as the first level of word-search.

The search can be extended more comprehensively, by exploring other ways that our respondents may have described a risky situation they faced. For instance they might describe a situation that had "a real prospect of failure" – in other words the manager identified a risk-taking event. So by using the search-word "fail" this will identify all the passages containing that root-word and its related extensions (fails, failure, etc.). Repeating this with other possible words related to the meaning will lead to a more exhaustive and richer range of nodes, thus increasing the quality and robustness of the research (Fabregues and Molina-Azorin 2017, Michaud 2017).

A caveat in extending the word-search in this way, is that when moving away from the root-word (risk) the researcher has to increase the level of vigilance in examining the passages that the software identifies, to ensure that they still fit the core meaning of that node (Basit 2003).

### 2.5.2   *Searching for Known Constructs*

Then to an example where the Matrix Sorter was used to good effect. A research project mentioned before, explored the Entrepreneurial Orientation (EO) among a range of arts managers using interviews. EO has 5 dimensions, and the expectation is that in most contexts all 5 will be found (Lumpkin and Dess 1996, 2001;

Dess and Lumpkin 2005). Our question was whether this sample of managers would show an entrepreneurial slant (EO), and if so, to what extent?

In this project each EO Dimension was assigned as a Node. The number of Passages (=References above) at each node was counted, and it also tallied how many of the Interviewees (= Files above) mentioned that particular node. This resulted in a table with the two variables for each node (e.g. Table 1), and the table can then be used to create an x-y scatter chart which has the basic format of the 2x2 matrix sorter.

The x-y scatter chart is helpful in that it shows the results graphically, and this can be produced using standard spreadsheet applications. In cases where the node parameters (references, files) have a wide variance (10:1 or more), it can be helpful to use logarithmic axes in the x-y scatter-plot, for a more even spread of nodes.

In Chart-1 below the results are graphed, and the matrix sorter then identifies the relative significance among the five dimensions of EO within this sample of interviewees. As in Figure 1 before, node significance is low in the bottom-left quadrant, and increases towards high significance in the top-right quadrant of the matrix. So in this case Innovativeness (Innov) is clearly the most significant meaning to emerge from these informants, while all the other dimensions show a strong presence, except for one.
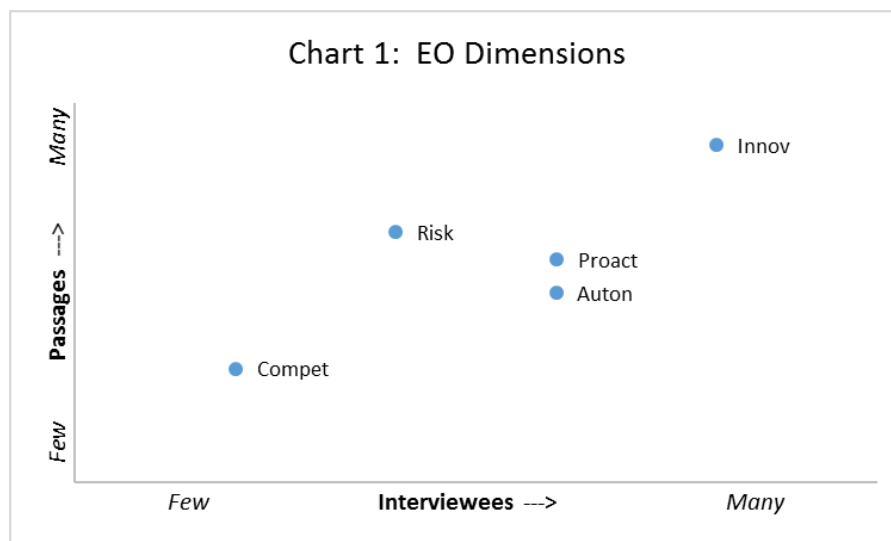
The node Compet is firmly in the quadrant of Weak Support (Fig.1) and therefore very much less significant than the rest. This node represents the EO dimension of Competitive Aggressiveness, and therefore showed very clearly that among our respondents, this dimension was "lacking". So while the original intention was just to confirm if the arts managers were entrepreneurial (or not), the answer was much more interesting than imagined. By using the Matrix Sorter we had discovered a small but intriguing anomaly. Without this framework and its reliance on two key significance-factors, it is quite likely that this anomaly would have been missed.

The data therefore confirmed that the respondents were certainly entrepreneurial, but in this case they were not aggressively competitive (with other arts companies), which is not what the EO construct suggests. This key insight later initiated an extension of the research to identify an important counter-dimension of collaboration (rather than competition) among the respondents, all drawn from what might be described as the Arts Community. So although these arts managers showed all the other EO traits, and were running profitable businesses, a key difference was that they achieved this without the need for "competitive aggressiveness".

This is a clear case of the Matrix Sorter rigorously guiding the research, and importantly also in uncovering "the unexpected" – something novel, new, surprising – and thus allowing a vital strength of qualitative research to emerge (Davidsson 2005). This insight also spawned a new line of research that explores potential contextual modifiers for EO. The evidence was always there in the qualitative database, but without the methodical approach described here, it is quite possible that this knowledge would have remained undiscovered.

2.5.3   Searching for the Unknown

In another case example, this time in an exploratory research context, the Matrix Sorter was used to construct a model from the nodes of meaning uncovered in the database. In this case the coding had identified 20 individual themes, gathered into nodes, and the question was then how to organise them into a coherent model. Importantly, that model had to represent the emphasis that the respondents had placed on each meaning. In this case the key factor named "references" in this article was termed "passages", and the factor that this article calls "files" was termed "documents". The raw results of the thematic content analysis are shown in the table below. Only the two right-hand columns are relevant here.

Table 6.2: Extent of Support for each Meaning Unit (Node)

Node	Characters	Paragraphs	Passages	Documents
Ambiguity, Risk Tolerance	7609	19	18	8
Cool with Chaos	3678	13	13	8
Don't Tell yet	2773	15	5	4
Flexibility	5926	33	24	11
Holism	1536	19	15	10
Look Ahead	5856	54	51	15
Market Feedback	9649	86	85	23
Money Matters	14232	127	139	26
Multiple Experts	12034	67	66	16
NL Approach	10488	73	72	18
NL Environment	8956	78	79	24
Non-Linear	44159	144	97	27
Pattern Recognition	5798	31	28	11
Ratchet in Assumptions	1866	11	12	8
Relational Balance - Fit	28040	117	63	20
Relationships in NVs	3526	29	10	8
Strategic Minimalism	7038	40	31	13
Sustainability & Ethics	9597	30	26	11
The Entrepreneur	16751	75	72	19
What's Changed	5744	58	56	18

NVivo® allows these "free" nodes to be organised into main (parent) nodes with sub-nodes (siblings), and if required, further levels of sub-nodes. While every node has a distinct meaning, carefully defined in the research, there are also some relationships among the nodes that emerged. In this context the hierarchy of the nodes is paramount, so that significant nodes are brought to the fore. To this end the raw data above is again entered into an x-y scatter-graph, as shown in Chart 2 below.

Chart 2: Nodes for NLRP
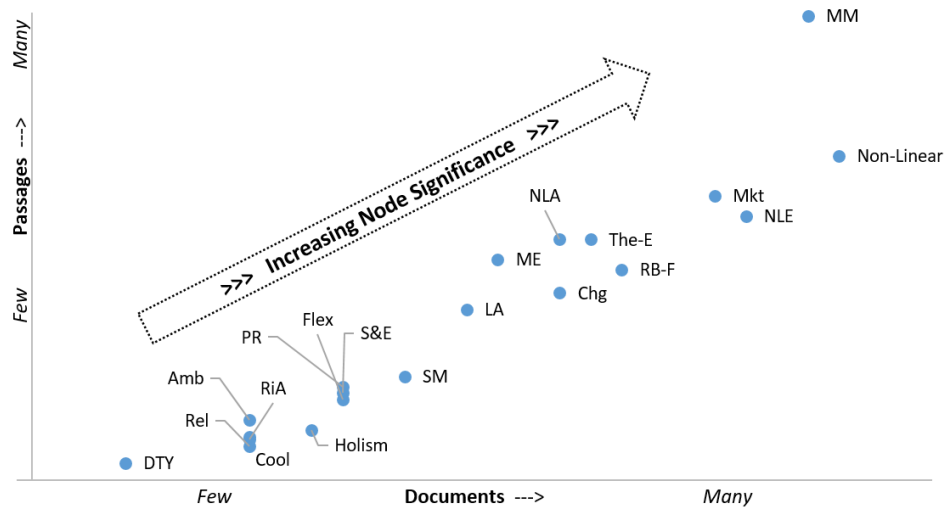
The chart uses cryptic abbreviations to identify the nodes, but the important observation here is that the random assortment of free nodes (Table 6.2) has now been arranged in a clear hierarchy. The low-ranking nodes are in the bottom-left corner, with significance increasing to its maximum in the top-right corner. It is also interesting to note that in this case there are no nodes of meaning in either of the outlier (high-low) categories of "The Committed Few" or "Widely-held View".

Having this hierarchy revealed now means the nodes can now be organised into a clear and cogent model, with each meaning taking a place that depicts the significance accorded it by the research informants. Without the framework of this matrix, it is hard to know how the researcher would assign the relative importance of nodes in accordance with the priorities of the research informants.

To illustrate this example further, the resulting model is shown below, with some nodes clearly more significant than others, and yet inter-related as well. It is not necessary to understand the model itself, only its structure, and the logic behind that structure. Importantly, the structure is guided by the Matrix Sorter, which in turn relies on the relative importance of each node that emerged from the database.

To explain the structure and the inter-relationships a bit further, the node and sub-node arrangements are helpful. Within each node-square, the numbers below the description indicate the significance of the node itself, while the numbers above add the significance contributed by its sibling nodes as well. For instance to be "Cool with Chaos" requires "Ambiguity and Risk Tolerance" and an attitude of "Don't Tell Yet" (because things are fluid and unconfirmed). The parent node has a score of 13 itself, but since it is supported by the two sub-nodes (score: 18 and 5), these two siblings inform the parent node, adding to its meaning and significance. The node "Cool with Chaos" thus ends up with a score of 36 (=13+18+5).

Similarly the parent node of "Strategic Minimalism" (score 31) is informed by the sub-nodes "What's Changed" (56), "Look Ahead" (51), and "Ratchet in Assumptions (12)". Its significance-score therefore becomes 150 (=31+56+51+12). This linked significance therefore places "Strategic Minimalism" above the sub-nodes below.

Finally, to explain the relatively high positioning of "Holism" and "Ethics", despite their low scores, this relies on strong support from the literature. To acknowledge their lesser support from the database respondents, they have been placed in an "auxiliary" or help-function, similar to what is found in some organisational charts. In this way the contribution of the literature and the research informants are combined into a coherent picture.
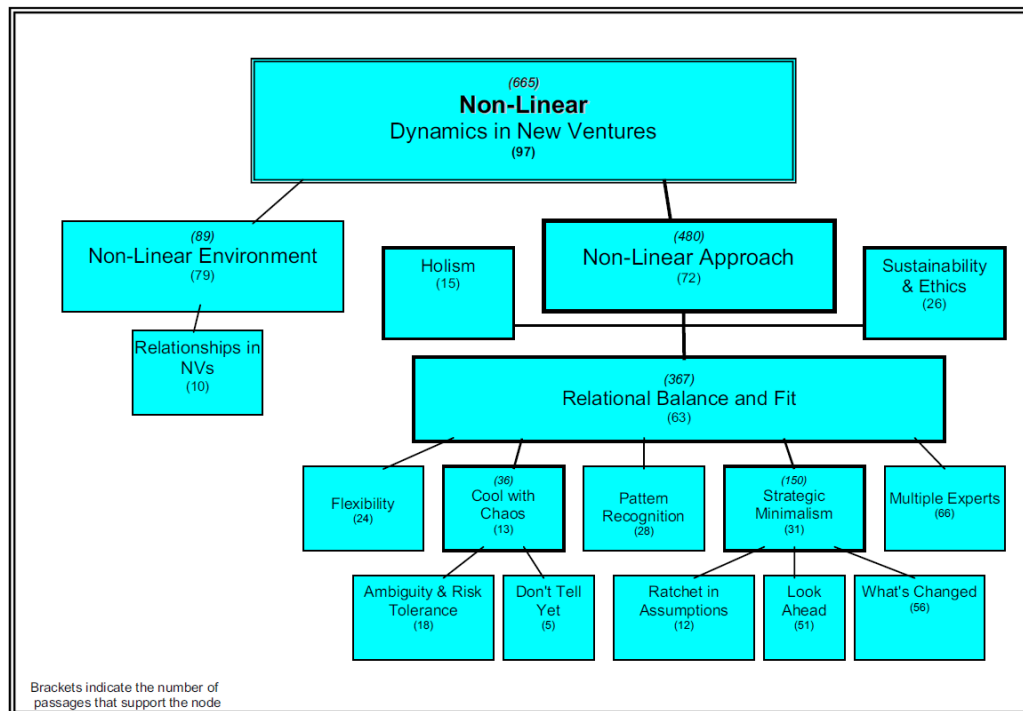
Figure 7.1: The NRLP Node Tree with individual and total citations indicated

2.6 Integrating the Node Hierarchy

As illustrated in the examples above, the focus on thematic meaning and the matrix sorter produce a hierarchy of nodes, and by making this the initial analysis, the data has been given the opportunity to "speak first". These data-based results can then be integrated coherently into the overall context of the research project. This integration would typically include comparison with existing theory frameworks, matching with industry experience, and questions around how this knowledge might be applied in practice. The final illustration above (Holism and Ethics nodes) is an example of where the database significance is integrated with the knowledge in the extant literature.

3. Implications:

Qualitative researchers have many challenges to deal with (Michaud 2017, Yang 2013, Neergaard and Ulhoi 2007) but for the persistent the rewards are many (Mir 2018, Davidsson 2005). To that end, the method and matrix sorter outlined here, is offered as a pathway forward, enabling the qualitative researcher to follow a rigorous and systematic process. This starts with a pile of documents and transcripts, and arrives at a coherent hierarchy of meanings (nodes). All the while this helps the researcher to reconnoitre the valuable content itself – from the research informants – and to reveal potential insights, confirmations and discoveries that are somewhere in the qualitative database (Alasuutari 2010).

The paper has therefore presented two ways in which to increase rigour while using the familiar method of thematic content analysis, which is especially important in qualitative research.

- Utilising the concept of meaning-units to maintain thematic focus during the search for themes in the qualitative database
- Using software analysis tools during the search-phase to reduce researcher fatigue, and to increase thoroughness and consistency when identifying key meanings in the qualitative database

Also, the paper has developed a simple but powerful matrix which sorts the identified themes into four distinct categories, with associated characteristics that enable the researcher to arrange the themes in a way that reflects the data.

Finally, this framework has been illustrated using actual research examples, to flesh out both the utility and capability of the framework.

Overall, the paper outlines how themes can be coherently identified and then categorised in a clear hierarchy, offering some confidence that the data itself has been "allowed to speak". The power of the 2x2 matrix sorter has been used to explore the unique value in the qualitative database itself, independent of the expectations of theories or experience. Importantly, this has retained "the human in the loop" to ensure that the results still rest on the expertise of the researcher(s), rather than on a simple machine output (Vaismoradi et al 2016). The use of computer-aided searching and counting has been leveraged to enhance the researcher's skills of perception and understanding (Basit 2003). The expert views that have so carefully been gathered in the database, are now analysed and organised in a way that is both scientifically repeatable and academically rigorous (Fabregues and Molina-Azorin 2017).

4.   Conclusion

Many researchers, including those referred to in this article, identify the typical obstacles facing those that choose the qualitative route. These can range from "it's a quant world", to the sometimes overwhelming richness of the data, and even identifying that qualitative approaches are often driven by a particular worldview that questions and challenges conventional group-think.

The method and matrix-sorter is therefore presented as one way to address these obstacles. There are some quantitative aspects to the technique which will hopefully appeal to those with a penchant for that approach, while importantly also being useful in analysing qualitative databases. The attention to meanings-not-words and software tools can assist in maintaining focus through the richness of the data, and the matrix-sorter presents a relatively straightforward yet powerful categorisation of nodes. Finally, for those that have chosen the qualitative approach and worldview, it is hoped that the framework outlined here will help to open the path ahead somewhat, and support the continued exploration of new knowledge and valuable insights.

References

Alasuutari, P., 2010. The rise and relevance of qualitative research, *International Journal of Social Research Methodology*, 13(2), pp.139-155

Basit, T., 2003. Manual or electronic? The role of coding in qualitative data analysis, *Educational Research,* 45(2), pp.143-154

Bengtsson, M., 2016. How to plan and perform a qualitative study using content analysis, *International Journal of Nursing Practice, Research and Education,* 2, pp.8-14

Bohm, D., and Peat, D., 2000. *Science, Order, and Creativity (2nd Edition)* London and New York: Routledge.

Braun, V., and Clarke, V., 2006. Using thematic analysis in psychology, *Qualitative Research in Psychology,* 3(2), pp.77-101

Corbin, J., and Strauss, A., 2015. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA, USA: Sage Publications, Inc.

Crowe, M., Inder, M., and Porter, R., 2015. Conducting qualitative research in mental health: thematic and content analyses, *Australian & New Zealand Journal of Psychiatry* 49(7), pp.616-623

Davidsson, P., 2005. Methodological Approaches to Entrepreneurship: Past and Suggestions for the Future, *Small Enterprise Research* 13, pp.1-21.

DeGregorio, E., Tagliafico, I., and Verde, A., 2018. A comparison of qualitatively and quantitatively driven analytic procedures of psychotherapeutic group sessions with deviant adolescents, *Quality & Quantity: International Journal of Methodology* 52, pp.1731-1760

Denzin, N., and Lincoln, Y., 2013. *Collecting and Interpreting Qualitative Materials*. Thousand Oaks, CA, USA: Sage Publications, Inc.

Dess, G.G., and Lumpkin, G.T., 2005. The role of entrepreneurial orientation in stimulating effective corporate entrepreneurship. *The Academy of Management Executive*, 19(1), pp.147-156.

Fabregues, S., and Molina-Azorin, J., 2017. Addressing quality in mixed methods research: a review and recommendations for a future agenda, *Quality & Quantity: International Journal of Methodology,* 51, pp.2847-2863

Fine, G., 2014 The Possibility of Inquiry: Meno's Paradox from Socrates to Sextus. Oxford, UK: Oxford University Press.

Giorgi, A., 1985   Sketch of a Phenomenological Method in *Phenomenology and Psychological Research*. Pittsburgh, PA: Duquesne University Press.

Giorgi, A., 2009 *The descriptive phenomenological method in psychology: A modified Husserlian approach.* Pittsburgh, PA: Duquesne University Press.

Glaser, B., and Strauss, A., 1967 *The Discovery of Grounded Theory: Strategies for Qualitative Research* New York: Aldine Publishing Company.

Henderson, Bruce 1970 https://www.bcg.com/publications/1970/strategy-the-product-portfolio.aspx accessed 31 May 2018.

Lakshman, C., 2007. Organizational knowledge leadership: a grounded theory approach, *Leadership & Organization Development Journal*, 28(1), pp. 51-75.

Lowy, A., and Hood, P., 2011. *The Power of the 2x2 Matrix – using 2x2 Thinking to Solve Business Problems and Make Better Decisions.* San Francisco, CA: John Wiley & Sons, Inc.

Lumpkin, G.T., and Dess, G.G., 1996. Clarifying the entrepreneurial orientation construct and linking it to performance. *Academy of management Review*, 21(1), pp.135-172.

Lumpkin, G.T., and Dess, G.G., 2001. Linking two dimensions of entrepreneurial orientation to firm performance: The moderating role of environment and industry life cycle. *Journal of business venturing*, 16(5), pp.429-451.

Madsen, D., 2017. Not dead yet: the rise, fall and persistence of the BCG Matrix, *Problems and Perspectives in Management*, 15(1), pp.19-34.

Michaud, V., 2017. Words fly away, writings remain – paradoxes in and around documents: a methodological proposition, *Qualitative Research in Organizations and Management: An International Journal,* 12(1), pp.35-52

Miles, M. and Huberman, A., 1994. *Qualitative Data Analysis.* Thousand Oaks (CA), USA: Sage Publications.

Mir, R., 2018. Embracing qualitative research: an act of strategic essentialism, *Qualitative Research in Organizations and Management: An International Journal*, 13(4), pp.306-314

Morgan, G., 1997. *Images of Organisation.* Thousand Oaks, CA, USA: Sage Publications.

Neergaard, H., and Ulhoi, J., 2007. *Handbook of Qualitative Research Methods in Entrepreneurship* London and New York: Edward Elgar.

Nowell, L., Norris, J., White, D., and Moules, N., 2017. Thematic Analysis: Striving to Meet the Trustworthiness Criteria, *International Journal of Qualitative Methods,* 16, pp.1-13

Onwuegbuzie, A., and Leech, N., 2005. Taking the "Q" Out of Research: Teaching Research Methodology Courses Without the Divide Between Quantitative and Qualitative Paradigms, *Quality & Quantity: International Journal of Methodology,* 39, pp.267-296

Punch, K., 1998. *Introduction to Social Science: Quantitative and Qualitative Approaches,* London: Sage Publications.

Strauss, A., and Corbin, J., 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* Thousand Oaks (CA), USA: Sage Publications.

Vaismoradi, M., Jones, J., Turunen, H., and Snelgrove, S., 2016. Theme development in qualitative content analysis and thematic analysis, *Journal of Nursing Education and Practice* 6(5).

Wheatley, M., 1992. *Leadership and the New Science* San Francisco, USA: Berrett-Koehler Publishers.

Yang, S., 2013. Surviving as a qualitative researcher in a quantitative world: a personal reflection, *International Journal of Social Research Methodology,* 16(1), pp. 81-85

Yin, R., 1993. *Applications of Case Study Research.* Thousand Oaks (CA), USA: Sage Publications..

Yin, R., 2014. *Case Study Research: Design and Methods.* Thousand Oaks (CA), USA: Sage Publications.

Yin, R., 2018. *Case Study Research and Applications: Design and Methods.* Thousand Oaks (CA), USA: Sage Publications.