The Qualitative Report 2015 Volume 20, Number 7, How To Article 2, 1029-1045
http://www.nova.edu/ssss/QR/QR20/7/harwood2.pdf

Cross-Check for Completeness: Exploring a Novel Use of Leximancer in a Grounded Theory Study

Ian A. Harwood
University of Southampton, Southampton, United Kingdom
Rod P. Gapp and Heather J. Stewart
Griffith University, Gold Coast, Australia

This paper investigates the potential for Leximancer software to actively support the Grounded Theory (GT) analyst in assessing the "completeness" of their study. The case study takes an existing GT study and retrospectively analyzes the data with Leximancer. The Leximancer output showed encouraging similarities to the main themes emerging from the GT analysis; but not sufficiently at the selective coding level to justifiably claim a definitive cross-check for overall theoretical saturation. Whilst Leximancer is not found to be a substitute for the 'hard labor' of GT coding and theory development, it can provide a very useful, efficient and relatively impartial cross-check of completeness/saturation in the open (and possibly axial) coding stage(s) of a GT study. This automated post-analysis check of GT coding is a novel use of a CAQDAS package. Keywords: Grounded Theory, Leximancer, Software, Theoretical Saturation, CAQDAS, Qualitative research

Over the last 50 years, qualitative research has achieved wider acceptance whilst the associated qualitative data analysis techniques have seen significant developments and improvements. From its inception in the 1960s, Grounded Theory (GT) has developed into being one of the most comprehensive qualitative research methodologies available (Haig, 1995). Based on iterations of deduction and induction, GT aims to build substantive theories of complex (mainly social) phenomena through five key tenets: (1) the constant comparative method, (2) theoretical coding, (3) theoretical sampling, (4) theoretical saturation, and (5) theoretical sensitivity (O'Reilly, Paper, & Marx, 2012). There is a rich literature on most of these aspects, starting from the genesis of GT in Glaser and Strauss (1967); although knowing when or how to break out from the iterative loops set up by constant comparison and theoretical sampling remains troublesome (see Jonsen & Jehn, 2009). Through the principle of reaching theoretical saturation and using a range of evaluative criteria (Strauss & Corbin, 1990), the GT analyst needs to make a judgment on when a GT study is "complete." This leads to nagging doubts of how to judge when a GT analysis has reached saturation or how, in a related sense, to cross check for the completeness of the study. As a result, many GT articles (i.e., Harwood, 2006; Partington, 2000; Schwarz & Nandhakumar, 2002) do not explore the specific issue of gaining theoretical saturation or "cross-checking" for completeness of the coding stages in any great depth. Some GT studies do not even mention theoretical saturation or the associated difficulties (i.e., Xiao, Dahya, & Lin, 2004).

Juliet Corbin frankly admits to being troubled by the issue of measuring "quality" in qualitative research and what the evaluative criteria might be (Corbin & Strauss, 2008). Most writers in this area are anxious to avoid terms such as "validity" and "reliability" as they have roots in positivism. Instead, GT studies can be evaluated through concepts such as whether they provide "plausible accounts" (Charmaz, 2006) or their overall "credibility" (Corbin & Strauss, 2008; Glaser & Strauss, 1967; Guba & Lincoln, 1985). Charmaz (2006) presents nineteen self-evaluative questions under the headings of credibility, originality, resonance and usefulness. Corbin & Strauss (2008) are appreciative that Charmaz's criteria cover both the

"scientific" and "creative" aspects of evaluation, but remain cautious about the need for the analyst's ability to critically self-reflect on their own GT output. Silverman (2001) is less reticent in explicitly dealing with validity and goes on to suggest five ways of validating qualitative research: analytic induction, the constant comparative method, deviant-case analysis, comprehensive data treatment, and using appropriate tabulations. Corbin & Strauss (2008) also suggest two overall ways in which to validate the theoretical GT output: (1) go back and compare the scheme against the raw data, doing a kind of high-level comparative analysis, and (2) tell the story to respondents or ask them to read it. A common theme throughout the GT literature appears to be the need for comparative analysis as a means of "validation" or verification and to check for theoretical saturation, but this relies on self-evaluation with the potential for bias and premature closure.

Over recent decades, there have been numerous Computer Aided Qualitative Data AnalysiS (CAQDAS) packages brought to market (see Gibbs, 2011, for a useful overview). These packages offer the promise of a more robust audit trail between data and theory, assisting the analyst to manage the usually high volume of data and in some cases to actively guide the coding process. Corbin and Strauss (2008) used a software package (MAXQDA) in their GT studies but like many analysts, they restricted its use to assisting solely with the theory development process. As mentioned above, there is much debate in the GT literature about how to assess the "completeness" and validity of the analytical output. What we mean by "validity" in this study is the opportunity to cross check the resultant coding and findings with an alternative analytical process (but using the same primary data). We therefore explore whether a CAQDAS package can be used retrospectively to "check" or verify the output of a GT study for theoretical saturation and completeness (rather than the more usual prospective application during the theory development stages). This is a novel use of the software and one which offers a potentially significant development in the GT (and perhaps wider qualitative research) community.

This paper reports on our findings whereby one of the author's existing data set from a significant and completed organizational GT study were independently analyzed by the other authors using Leximancer and the two resultant outputs are compared and contrasted. There are many different CAQDAS packages available but we specifically chose to limit the scope and focus on Leximancer in this study as it: (a) avoids the need for detailed word by word manual coding (i.e., it is efficient and seemingly ideal for exploring this "cross check" application), (b) its lexical analysis is algorithmically generated (thereby avoiding coder bias) and, (c) two of the authors had extensive prior experience in using the package.

There now follows an overview of Grounded Theory and Leximancer processes in general before outlining the methodology of this present comparative study. The original GT study findings are then summarized followed by the Leximancer analysis leading on to a comparison between the two outputs. A final discussion section provides insights into how Leximancer might be used in the GT evaluative process.

Overview of Grounded Theory and Leximancer

Grounded Theory Coding Process

GT is set within an interpretive research paradigm, which aims to "… reveal the underlying processes of what is going on in a substantive area of study" (Lowe, 1998, p. 106). Originally developed in 1967 by Barney Glaser and Anselm Strauss in the sociology discipline, GT has become widely used and accepted in other fields such as accounting, social welfare, psychology, education and management (see Locke, 2001).

Due to divergence between the GT originators following their original text (Glaser & Strauss, 1967), there are now different approaches to GT analysis. The reasons for this split are beyond the bounds of the present article however interested readers could read Rennie (1998a, 1998b), Corbin (1998) or Babchuk (1997) for an overview. With common roots though, the different GT coding approaches have a similar end goal in mind, that of generating substantive theory which is *grounded* in the data. Heath and Cowley (2004) develop an in-depth discussion between the two main approaches of GT analysis and conclude that the novice researcher should choose based on their own cognitive styles. In the present case, the data were analyzed using Strauss and Corbin's (1990, 1998) methodology due mainly to the more structured approach and availability of analytical tools and frameworks.

In GT, data are usually collected through discussions with participants who are centrally and actively involved in the phenomenon of interest. This quite often starts from a convenience sample (Saunders, Lewis, & Thornhill, 2012) and then further individuals are drawn into the study through theoretical sampling which, "rather than being predetermined before beginning the research, evolved during the process and is based on concepts that emerged from analysis and that appear to have relevance to the evolving theory" (Strauss & Corbin, 1998, p. 202).

Strauss and Corbin's approach to GT analysis involves three interlinked stages. Firstly, in open coding, data are fragmented into discrete parts, closely examined, and compared for similarities and differences. In doing so, open coding aims to "uncover, name, and develop concepts, [through] opening up the text and exposing the thoughts, ideas, and meanings contained therein" (Strauss & Corbin, 1998, p. 102). There are various terminologies associated with the GT coding processes, the first of which is a "concept" which can be described as a "labelled phenomenon," and can take the form of an event, happening, object and/or action/interaction (Strauss & Corbin, 1998). As the analyst scans the data and identifies concepts, any that are conceptually similar can be grouped together and assigned a new over-arching label referred to as a category. Categories are the major "building blocks" in answering the question of "what is going on here?" (Strauss & Corbin, 1998). Further fragmentation of the data takes place by asking questions such as when, where, why, how and so on of each category (Strauss & Corbin, 1998). Two further terms used in open coding are "properties" and "dimensions." Strauss and Corbin (1998) define properties as "the general or specific characteristics or attributes of a category" (p. 117). They go on to define dimensions as "representing the location of a property along a continuum or range." In other words, the same object, concept or category can be differentiated through the dimensional range of its properties.

Having fragmented the data into detailed open codes (along with their related properties and dimensions), axial coding aims to reassemble the data through establishing relationships between categories. In axial coding, the analyst is moving away from description and into a more abstract, conceptual level (Strauss & Corbin, 1998). Various analytic tools are available to assist in axial coding with the paradigm model providing an important "canvas" on which to integrate the conditional context with the key actions and interactions of respondents. Throughout both open and axial coding, the category labels are constantly being reviewed, grouped, regrouped and in some instances re-labelled, to take into account the emerging themes from the data. The analyst keeps a written track record of theoretical and coding developments through the use of memos.

As the GT study progresses, the analyst is constantly reviewing for theoretical saturation, that is, "the point in category development at which no new properties, dimensions, or relationships emerge" (Strauss & Corbin, 1998, p. 143). At this stage, selective coding is employed in order to integrate and refine categories through the use of relational statements resulting in the "isolation" of a core category. The core category is a word or phrase that is representative of the central phenomenon emerging from the study. This should have sufficient

explanatory power to integrate all lower level categories into the theoretical framework. Strauss & Corbin (1998) provide various tools and frameworks to assist in selective coding such as writing a *storyline*, *mini-frameworks*, reviewing and sorting memos and developing integrative diagrams. A set of criteria for establishing the core category is also provided (Strauss & Corbin, 1998, p. 147).

The aim of coding is to allow the analyst to constantly compare original data with an emerging theoretical framework before informing the need for further data collection and so on until a robust and integrated theory results. Far from being linear, the data collection and coding stages in GT proceed in iterative cycles, one informing the other until theoretical saturation is achieved (Corbin & Strauss, 2008).

Background to Leximancer

Already actively used as a research tool in published literature, Leximancer is a text analysis software program that employs proximity values to conduct analysis of transcript data (Cretchley, Rooney, & Gallois, 2010a; Cretchley, Gallois, Chenery, & Smith, 2010b; Smith & Humphreys, 2006). In contrast to manual coding of data, Leximancer automatically analyses text to create concepts and themes from the uploaded data (Cretchley et al., 2010b). Computer generated analysis tools create rigor in the analysis in several ways. Leximancer limits any possible predispositions to prior knowledge gained in the data collection phases. It is also argued that Leximancer provides reliability through stability and reproducibility (Cretchley et al., 2010a).

Based on statistical algorithms, Leximancer is a software tool designed for the analysis of natural language text data. While the details of the algorithms are beyond the scope of this paper, in summary they extract semantic (meaning) and relational information. The extraction of concepts is achieved through word association by investigating terms that indicate meaning around a word. At this point a lexical concept is formed when a group of related words are found to travel together through the text. From synonyms, adjectives, proper nouns and compounds, seed concepts are formed as a starting point and a thesaurus is built. Semantic meaning created through conceptual analysis is achieved by the presence and frequency of words, phrases and co-occurrence of words leading to the generation of a concept. Both explicit and implicit concepts are identified as an outcome. Through a dictionary and learning thesaurus Leximancer generates syntax then word, name and term resulting in implicit coding, for example *customer* linked to *client*, *subscriber* and the plurals of these words. The software is interactive and allows the researcher to directly search, add, remove, and merge terms as well as the above automatic process. Themes are extracted from the collection of related concepts in close proximity. The theme name is the most prominent concept (in terms of the concept most central in meaning and/or most highly connect to the other identified concepts), not the most frequently occurring concept. Of importance are the detailed logs that the software produces, for each concept these logs are indicated in the actual text with each concept located. From this point, the researcher can be guided through the text with all concepts highlighted allowing a level of human reflexivity in the analysis process (Leximancer Manual, 2011; Middleton, Liesch, & Steen, 2011; Rooney, 2005; Weaven, Frazer, & Giddings, 2010).

Developments of qualitative analysis software packages such as Leximancer facilitate the process of coding, particularly when a considerable amount of information is gathered and transcribed (Easterby-Smith, Thorpe, & Lowe, 2002). Speed and rigor are other advantages that transpire from utilizing computer programs for data analysis (Illia, Sonpar, & Bauer, 2014; Seale, 2003). Although there is not a single or precise way to analyze data, Leximancer is designed to take a global perspective of the themes and concepts to increase the objectivity hence decreasing the preconceptions of manual content analysis (Smith & Humphreys, 2006).

As indicated, Leximancer groups words by concepts to identify the underlying themes (Gapp, Fisher, & Kobayashi, 2008). The non-selective programmed process explores the transcribed text in order to group words by synonyms, associations and frequencies. This process elicits emergent concepts from the data with themes clustered by conceptual words (Cretchley et al., 2010b; Martschinke, Waugh, Beamish, & Davies, 2004). From here the program develops a two-dimensional map displaying the relationships between the themes and concepts as well as assisting the researcher in the interpretation. Several runs of Leximancer stabilize the map to increase the validity of the output. Frequently occurring words along with their relationships become concepts and clusters of concepts are grouped into themes. These themes are summarized into circles that are named after the dominant concept within that group. The theme is also identified by the largest "dot" within that group. Through this visual demonstration of the concepts, their strengths, along with the semantic formation of the data, are provided by the map (Cretchley et al., 2010a). In their review of Leximancer's capabilities, Angus, Rintel and Wiles, (2013) highlight the value of the "visuality" of these data displays from Leximancer but they still stress the importance of the analyst having to make sense of the relationships between themes.

Leximancer software provides both a quantifying and conceptual structure of the qualitative data built from the occurrence of related iterations and concepts that identifies a centrality of similar contexts to build the resulting themes (Gapp & Merrilees, 2006). Increasing use of Leximancer, as an alternative to manual coding, is evident across a diverse range of research and practitioner based disciplines including business, education, politics, information technology, transport, psychology and health care (Cretchley et al., 2010b; Crofts & Bisman, 2010; Gapp & Fisher, 2012; Grech, Horberry, & Smith, 2002; Martin & Rice, 2007; Stewart & Gapp, 2014). A number of studies have investigated the use of CAQDAS in qualitative data coding (i.e., Davidson & Skinner, 2010; Illia et al., 2014; Rettie, Robinson, Radke & Ye, 2008) but none to our knowledge on the specific use of Leximancer in the context of cross-checking manual GT coding output.

Research Methodology and Context

Our overall research question in this study is: "Can Leximancer be used to cross-check for saturation and *completeness* of a GT-derived substantive theory?" To address this question, we are presenting a single exploratory case study (Yin, 2003) to provide "an extensive examination of a single instance of a phenomenon of interest" (Hussey & Hussey, 1997, p. 65). For clarity, we must stress that the unit of analysis and point of interest in this paper is the potential for retrospective checking of a GT study, rather than the actual topic of the GT study itself. We provide a short contextual background to the original study below in order to make the comparison more meaningful, however readers wishing to learn more about the actual GT study can find details in Harwood and Chapman (2009); Harwood, Ward, and Chapman (2009); Harwood (2006, 2001); and Harwood and Ashleigh (2005).

The original study was set within a £240m organizational change program in a multinational healthcare corporation. Following a review of the global supply chain, various scenarios (each with its own business case) were developed in order to streamline supply following various company and product acquisitions and divestments. As a result, more than 2,000 product variants were moved to new supply locations resulting in 16 manufacturing plants being divested. Throughout the four-year change program there were conflicting objectives, resistance and great anxiety amongst the workforce in general; although a good deal of effort was put into minimizing the negative impacts of the changes. Given the significant uncertainties at the start of the rationalization project (and the historically poor levels of "success" in merger and acquisition integration projects in general), the GT study originally

focused on how risk was perceived and managed within the teams involved in developing the business case options. However, as GT is a "voyage of discovery" (Denscombe, 2010) and the substantive theory is "emergent" (Glaser & Strauss, 1967; Glaser, 1978) the actual results showed how risk was being used as a "currency" amongst the individuals in order to shape the outcome of the business cases.

In total, there were 22 individuals involved in the interviews (some being interviewed on more than one occasion). The first round of interview questions were sensitizing and/or theoretical (Corbin & Strauss, 2008), with the aim of enabling respondents to explain and explore their actions within the change process:

- What were the risk management techniques used in the integration program?
- What were the "drivers" behind the usage of these techniques?
- How accurate were the risk management techniques being used?
- What do you think the impact was of the risk management techniques?
- Were there any key learnings with respect to the risk management?
- What was the culture like with respect to risk and risk taking in the program?
- Do you think that the project has been a success?
- What were the success measures used?

Following early open and axial coding various "gaps" were emerging, especially around the impact of confidentiality, and so on the basis of theoretical sampling (Strauss & Corbin, 1998). The following questions were asked in further interviews:

- What does confidentiality *mean* to you?
- How is confidentiality maintained or monitored?
- How do you know who is signed up, and to what level?
- How does confidentiality impact your work?
- How does confidentiality affect *you*?
- What happens if confidentiality is *breached*?
- Have confidentiality constraints given rise to any ethical issues?

The original interview transcripts from the GT study (comprising 212 pages of nearly 110,000 words) were emailed to the "Lexi-authors" to independently carry out the Leximancer analysis. Only then did all authors meet in person to discuss the Leximancer output and for the "GT-author" to explain the GT study for the first time.

Leximancer Analysis Process

The subsequent process was used to undertake the lexical analysis, which highlights the role of the qualitative researcher in creating meaning rather than solely reporting the output of a software tool. Following a review of the interview transcripts to ensure the data was in a usable format, a *project* (Leximancer terminology) was created. Interviews for this analysis had been merged into a single document that was uploaded into the project. Leximancer interpreted this whole document line by line, however in this case the software was set to compare two sentence blocks which is the default setting on a one to three sentence option. Other fundamental checks of the settings were carried out with defaults being opted for, and then a full run of the project was completed. From interpreting the document, a thesaurus is developed to locate synonyms so that patterns of language can be identified and reduced back to categories. This is normally done over a thousand iterations, which is equivalent to reading

the document and looking for meaning a thousand times if done manually. During this stage, the program is looking for significance rather than frequency and it is these substantial findings that create the output of a concept map. This concept map proposes categories that are placed into clusters under identified themes with the themes recognized by their relevancy in the conversation rather than purely on their frequency of use. This first run highlighted issues with words such as *ummh*, *Resp 1* and the interviewer's name being displayed as concepts. By reviewing the instances of these word occurrences in the data, it was ascertained that they had no significant meaning and they were therefore added to the *stop list*. After the second run, interaction with the concept map is needed and is part of the process. For a strict interpretation of the concept map, the clustering was run several times (*randomized* and *re-clustered*) with the map being inspected on each occasion. Once the relative positioning of the concepts is similar between runs, the cluster map is considered to be representative (Leximancer Manual, 2011).

Part of the analytical interaction is in setting the thematic size. This was initially taken to 50% to produce six themes, then 55% with four themes and finally at 100% that resulted in the two highly connected themes of *risk* and *confidentiality*. All initial concepts remain visible yet collapse under the *risk* and *confidentiality* themes. From this, the software utilizes colors to identify the dominant themes with red denoting the primary theme (hot spot) then grading to cooler colors for decreasing significance. The findings are presented in the lexical map where the dominant themes are color highlighted and surrounded by the concepts used to generate the theme. Overlapping themes depict connectivity however there can be isolated themes that sit outside the dominant conversation. Further to this, the conversation was analyzed through a pathway to demonstrate the specific example of connectivity. All identified word categories are presented on the Leximancer screen in both a map and tabular form.

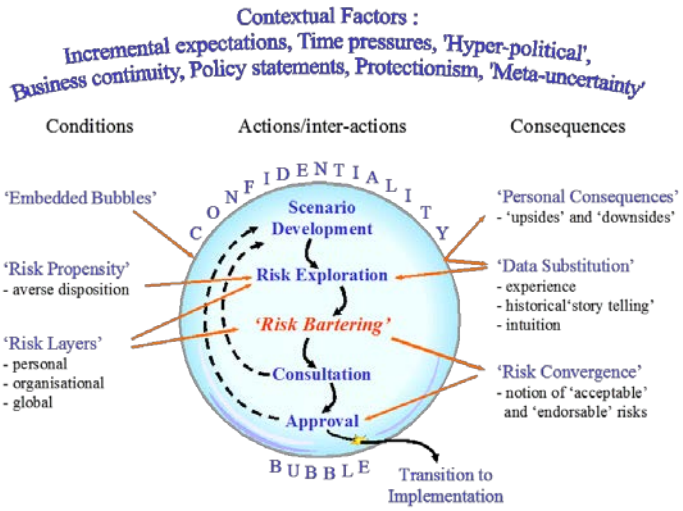
From the list of all concepts generated, including the highlighted themes, Leximancer provides key direct quotations from within the dialogue it has analyzed. This guided us (the "Lexi-authors") to read all of the relevant text within the newly defined context. These relevant paragraphs were exported into a word document allowing us to understand the program's interpretation of each concept. At this point, we assessed the value and trustworthiness of the Leximancer outcomes. This part of the analysis is essentially the reverse of a traditional qualitative process where we built the story from behind the concepts rather than first reading the dialogue to identify concepts and subsequent themes. This is consistent with others' application of Leximancer (Hansson, Carey, & Kjartansson, 2010; Middleton et al., 2011; Rooney, 2005; Weaven et al., 2010).

There now follows a brief summary of the outputs from the GT study and Leximancer analysis. This then forms the basis of the comparisons to crosscheck for theoretical saturation and completeness of the GT study.

Presenting the Outputs of GT and Leximancer Analyses

There were hundreds of categories (and sub-categories) to emerge in the GT study, each with their own dimensions and dimensional ranges. The key themes that became recurrent and therefore *survived* to be considered for the core category were the interplay between *risk*, *confidentiality* and *business case approval.* Under the phenomena of risk, there were aspects such as *risk exploration*, *experience*, *risk culture*, *risk hierarchy*, *normalization* and *pressure.* Under confidentiality, there were concepts such as *personal motives*, *risk filter*, *interpretation* and *inside/outside*. The main issues in developing the business cases were the *pressure of timescales*, *hierarchy of approvals*, *scenario development*, *peer review* and *past successes.* These concepts are shown in Figure 1.

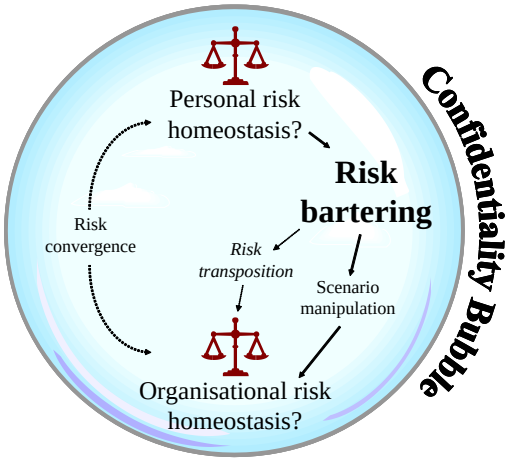
Figure 1: The "Risk Bartering" Paradigm Model



Source: Harwood and Chapman (2009)

During constant comparison between theoretical concepts and data, linkages began to emerge and further codes developed. Essentially, the GT study findings were uncovering a complex web of actions and interactions relating to the role of risk, and how individuals used risk on a personal level during the business case scenario development process. Figure 2 shows the final integrative diagram whereby the core category of "risk bartering" is shown, together with the balance between personal and organizational risk all taking place within confidentiality constraints, which are portrayed through the metaphor of a bubble.

Figure 2: The "Risk Bartering" Integrative Diagram



Source: Harwood and Chapman (2009)

The substantive theory of "risk bartering" (Harwood & Chapman, 2009) is written as:

> When faced with a particular integration scenario (i.e., a future change), individuals who are positioned inside the confidentiality bubble make an assessment of their personal risks. If the perceived risks are not acceptable (i.e., an imbalance exists, either too high or too low), the individuals will enter into rounds of risk bartering. The bartering process involves the individuals attempting to transpose their personal risks by over or understating organizational (or global) risks, thereby manipulating the scenarios to move

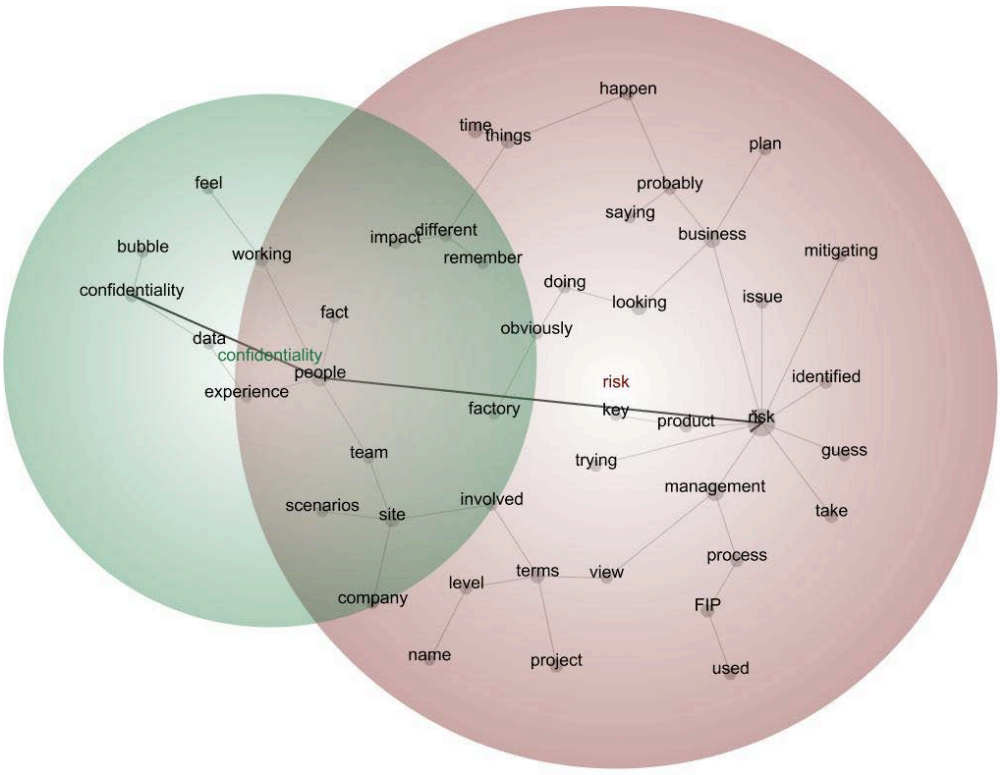
> their personal risks towards a balanced state. Once the scenarios are altered however, other individuals may well perceive an imbalance in their personal risks and so the risk bartering continues, until a compromise situation is achieved (through risk convergence). The confidentiality bubble limits the verification of transposed risks throughout the bartering process, with team members having to rely on data substitutes. (p. 172)

It was clear from the GT work that risk, confidentiality, business case scenarios and individuals with personal consequences were key recurring themes throughout the analysis. The core category (Strauss & Corbin, 1998) of "risk bartering" provided some unique insights into this post-merger and acquisition integration study whereby risk was being used as a currency to barter and negotiate between individuals during the business case development process, all of which was bound by tight confidentiality requirements.

Leximancer Output

Pathways described by Leximancer represent the relationship between themes, which describe the most plausible navigation across the concept map. At a thematic size of 100%, *risk* and *confidentiality* were the most salient themes identified which could not be separated or increased, stressing the connectivity between these themes. The pathways mode within Leximancer views both the frequency and weighting of each word in a sentence and in this case, in a block of two sentences that presents a contribution to the accumulated evidence for the concept. In this scenario, *confidentiality* contributes 64% and *people* 36% to the theme of *risk*. Figure 3 depicts the direct pathway relationship of *confidentiality* through the concept of *people* to *risk* with the most plausible link developed.

Figure 3: Leximancer map at 100% with the dominant themes of *"risk"* and *"confidentiality"* demonstrating their connectivity through the concept of *"people."*



Leximancer displays the primary theme by a red or hot color, in this case *risk* which is dominant to *confidentiality* with the two themes connected through the concept of *people*. Drawing conclusions from this pathway indicates that *confidentiality* for the people involved is a priority and must be kept. This secrecy creates issues in terms of *risk* and *risk* assessment. Concern over the lack of or inability to disclose creates a predicament where decision-making occurs because of a lack of accurate information, or the ability to seek clarification and feedback. The participants concerned with this decision-making process work within a team environment on projects that affect the livelihoods of other *people* in the workplace and can disadvantage those other *people* within the organization. Given that *risk* is based on probabilities in the absence of realities, this reduction in dependable information and data due to the *confidentiality* required indicates a state of stress and confusion for those involved in this process.

Respondents expressed confusion and stress generated by the inability to understand what is really at *risk* in terms of *confidentiality*. The scale of the impact was proportional to their position within the organization. Outcomes indicated here were feelings of inequity, seclusion and fear of diverging interests, which result in negative impacts on individuals and teams. As a result, the lack of information distributed is attributed to the need for *confidentiality* however Respondent 8 confirmed the need for transparency:

> *RESPONDENT 8:*
> *"Hierarchy of confidentiality - need to know basis, some people know more than others, may be a link to hierarchy within the organization. Some people felt that that drove them to almost a feeling of isolation within the teams and that they couldn't actually reflect back to where they had come from, if the teams had come from various parts of the world to come together, and also a conflict of interests."*

This lack of information and transparency led to a reduction in the quality of decision-making. Respondent 7 used descriptors of *isolation* and *conflict of interests* in association with team members feeling undermined as a *team*. In ideal situations, competent teams need to be working together on common objectives. In the case under investigation a combination of data substitution and missing data led to misinformation with decisions made without foundation. Respondent 10 highlighted that *risk* assessment is constrained by the source of information and this is restricted to the *people* included in the *in group* or as labeled within the study, the *confidentiality bubble*.

> *RESPONDENT 7:*
> *"Confidentiality drives **isolation** and/or conflict of interests, also data substitution. Risk culture – people spoken to in [company] are typically risk averse."*

> *RESPONDENT 10:*
> *"The issue there is that it's actually only… your assessment of the risk is only as good as the people you've actually drawn into that confidentiality bubble."*

The respondents take *confidentiality* seriously however the lack of transparency and information impacts on the ability to create appropriate strategies for the effective management of *risk*. *People* and teams involved in the study want to achieve the best outcomes by gathering the most beneficial information to carry out tasks at hand. Guessing and filling in the missing

pieces of data as a result of maintaining the *confidentiality* impinges on the desired performances of those involved resulting in more cautious rather than effective approaches.

> *RESPONDENT 8:*
> *"So… you've got a conflict there between minimizing… the need to minimize the number of people from the confidentiality point of view but maximize the experience and the inputs and the facts and data etc. that you put in. And I think there's no doubt that in the absence of specific…ummh… knowledge of a site or a scenario then you do tend to be more conservative…"*

Over time, as the need for *confidentiality* became more relaxed, more *people* became involved. With this involvement, the quality and quantity of information from more experienced *people* in relevant business areas was accessed. These knowledgeable individuals were able to inject ideas and issues previously unavailable due to the secrecy and isolation experienced as a result of the emphasis on *confidentiality* within the process. Broadening involvement by decreasing the *confidentiality* increased the amount and relevance of the information available to decision makers.

> *RESPONDENT 12:*
> *"…when sort of broadens out to bring a wider group of people on site, because usually it starts off as quite a narrow … err… you know, group, due to confidentiality, when that sort of broadens out and you're getting more people involved, I think you … you do kind of go back through that exercise again, because you're getting to the guys who are close to the coal face, if you like, and they're…they're saying* ***well, what if*** *and* ***what about*** *or* ***have you considered****?"*

Having presented the outputs from both the GT study and Leximancer analysis, the following section evaluates and discusses the usefulness of using Leximancer to *cross-check* for completeness in a GT study.

Discussion

GT analysis is often an individualistic endeavor meaning that the final stages of theory development might well end up at a *false summit*, to use a mountaineering metaphor. In this uncharted territory, there are no maps to show the surrounding *contours* or *paths* to the *true* summit and, more often than not, no experienced guide with whom to seek advice or direction. In a related sense, theoretical saturation as defined by Strauss and Corbin (1998) is problematic as it is almost impossible to imagine a time whereby categories are *fully* saturated and where there are literally *no* further properties to be found in what is most likely to be a dynamic and developing social setting. O'Reilly et al. (2012) state that after several months in the field and collecting data through observations and interviews for their GT study, "no new material aspects or dimensions of the core category… were being discovered" at which point they brought the data collection to a close. Verification of the data saturation involved self-evaluating the substantive theory in terms of whether it was a "reasonably accurate statement of the problem explored" (p. 254). The judgmental nature of this endpoint leaves the GT analyst open to constructive criticism and any additional support is to be welcomed.

The original GT study in the current paper took place over a number of years and included participant observation as a means of data collection to supplement the interviews (as encouraged by Grounded Theory). The richness gained through this additional medium was not available directly to the Leximancer analysis (although it did guide some of the questions

asked in the interviews). It is therefore important when comparing the GT and Leximancer outputs in this case to appreciate that it would be highly improbable for there to be an exact match. Instead, we were looking for any similarity of key themes emerging from the data. In this instance, we found that risk, confidentiality and people/individuals were common to both approaches. The Leximancer *storyline* also makes reference to the impacts of *confidentiality* and how conflicts changed over time. It was reassuring to note that Leximancer did not raise any significant concepts that the original GT study had missed and that the underlying themes were evident.

Whilst Leximancer did not combine the concepts *confidentially* and *bubble* as seen in the construct *confidentially bubble*, the two words are side by side in the lexical analysis indicating a very strong interaction. Other specific terms found in the GT analysis such as *risk homeostasis*, *risk convergence*, *risk transposition* or the core category of *risk bartering* where not directly identified by Leximancer. This higher order understanding would require a human extension of the computer analysis. What were also identified in the lexical analysis though were two concept chains. The first, from *risk* through *management*, *process*, to *used*; and the second, *issues* associated with *risk* and *mitigation* suggesting the need for the researcher to delve deeper into these two aspects of the analysis. This did not create the same language but did indicate similar matters of importance to the GT study's findings. Given the blind nature of the lexical analysis perhaps greater involvement in the data collection process may have led to a stronger synergy between the GT and lexical analysis. This does leave us reticent to claim that Leximancer can unquestionably be used as a final crosscheck of theoretical saturation of a GT study (especially based on just this one case study). With that said, the Leximancer analysis has provided a reassuring account and many of the key themes identified in the GT open coding (and GT paradigm model) were present, enabling a very useful cross-check of the open (and possibly axial) codes.

Utilizing Leximancer as a tool to help verify the *completeness* of the GT open coding provides the analyst with a much needed source of confidence at a crucial time in their work (and an additional form of triangulation to those presented by Jonsen & Jehn, 2009). Further positives of CAQDAS tools and specifically Leximancer include the potential minimization of human bias creating rigor, stability and an efficient approach when analyzing a significant amount of data. Corbin and Strauss (2008) though state that "computer programs don't do the thinking…" (p. 310). We counter-argue that some of the main GT texts are at risk of becoming outdated in terms of advancements in CAQDAS packages. Leximancer certainly explores relationships in the data enabling the analyst to probe for any missed avenues of theoretical exploration. We therefore propose a new application for software-enhanced GT analysis. Instead of using CAQDAS as a repository throughout the coding process (which can arguably restrain the creative process), packages such as Leximancer can be used post-coding in a confirmatory way to help the analyst assess for saturation and the overall *completeness* of the open coding stage of a GT study.

We must be cautious though. Stewart and Chakraborty (2010) state, "Leximancer is a software tool used to support lexical-text analysis, consistent with grounded theory methodology" (p. 3). We would not go so far as to claim *consistency* as we believe that GT is significantly more than lexical analysis. We also share Jones and Noble's (2007) concerns that GT is in danger of losing its integrity when such claims are made. GT is a rich and creative process and the use of Leximancer is not proposed to shortcut this creativity. Indeed we believe it would be counter-productive to theory development if the analyst were to look at the Leximancer output too early in the process. A reasonable question that might arise from this paper (and especially perhaps by doctoral research students) is: "if Leximancer is so efficient at analyzing the data to arrive at a similar output, why not just bypass GT and save many months of analytical angst and hardship?" Such a question would indicate a serious lack of

understanding of the GT process, constant comparison, theoretical sensitivity and the role of the analyst in creative and robust theory development. We are not suggesting that Leximancer could or should replace the GT analytic process. Instead, we see Leximancer as a complimentary tool to provide a useful *cross check* of the analysis, especially after open coding, to ensure that the grounded theorist has explored all analytical avenues within the available data; a *coding companion* and *critical friend* if you will.

Conclusion

Corbin & Strauss (2008) propose that one of the ways to *validate* the theoretical output of a GT study is to go back and compare the scheme against the raw data, doing a form of high-level comparative analysis. The analyst will never be able to measure (in a statistical sense) the validity of their GT (or any other notionally qualitative) study; to think so shows a misunderstanding of the approach and paradigm. What Leximancer can provide is another form of comparative analysis via an independent (context free) second opinion with which to make an improved judgment. Leximancer is not a panacea, it still requires analytical sensitivity and judgment in its interpretation, but it is straightforward to probe the data and cross-check via the resultant maps. In contrast to many other CAQDAS packages, Leximancer enables the analyst to make sense of large narrative data sets with minimal manual coding. The result is an efficient and impartial second opinion on open codes (concepts, categories and dimensions) and potential links between them with which the GT analyst can then use to inform their later decision on theoretical saturation and completeness, in parallel to the existing tried and tested methods. This is a novel retrospective use of Leximancer, an additional contribution to the use of CAQDAS systems (see Atherton & Elsmore, 2007), and we encourage others to replicate our approach with their own GT data (and alternative CAQDAS packages) for comparison.

References

Angus, A., Rintel, S., & Wiles, J. (2013). Making sense of big text: A visual-first approach for analysing text data using Leximancer and Discursis. *International Journal of Social Research Methodology*, *16*(3), 261-267.

Atherton, A., & Elsmore, P. (2007). Structuring qualitative enquiry in management and organization research: A dialogue on the merits of using software for qualitative data analysis. *Qualitative Research in Organizations and Management: An International Journal*, *2*(1), 62-77.

Babchuk, W. (1997). *Glaser or Strauss? Grounded theory and adult education*. Proceedings of the 15th Annual Midwest Research-to-Practice Conference in Adult, Continuing and Community Education, October 15-17, Michigan State University, East Lansing, Michigan.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London, UK: Sage.

Corbin, J. (1998). Alternative interpretations valid or not? *Theory & Psychology*, *8*(1), 121-128.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Cretchley, J., Rooney, D., & Gallois, C. (2010a). Mapping a forty-year history with Leximancer: Themes and concepts in JCCP. *Journal of Cross-Cultural Psychology, 41*(3), 318-328.

Cretchley, J., Gallois, C., Chenery, H., & Smith, A. (2010b). Conversations between carers and people with schizophrenia: A qualitative analysis using Leximancer. *Qualitative Health Research, 20*(12), 1611-1628. doi: 10.1177/1049732310378297

Crofts, K., & Bisman, J. (2010). Interrogating accountability: An illustration of the use of Leximancer software for qualitative data analysis. *Qualitative Research in Accounting & Management*, *7*(2), 180-207.

Davidson, L., & Skinner, H. (2010). I spy with my little eye: A comparison of manual versus computer-aided analysis of data gathered by projective techniques. *Qualitative Market Research: An International Journal*, *13*(4), 441-459.

Denscombe, M. (2010). *The good research guide for small-scale social research projects* (2nd ed.). Buckingham, UK: Open University Press.

Easterby-Smith, M., Thorpe, R., & Lowe, A. (2002). *Management research* (2nd ed.). London, UK: Sage.

Gapp, R., & Fisher, R. (2012). Undergraduate management students' perceptions of what makes a successful virtual group. *Education+ Training*, *54*(2/3), 7-7.

Gapp, R., Fisher, R., & Kobayashi, K. (2008). Implementing 5S within a Japanese context: An integrated management system. *Management Decision*, *46*(4), 565-579.

Gapp, R., & Merrilees, B. (2006). Important factors to consider when using internal branding as a management strategy: A healthcare case study. *Journal of Brand Management, 14*(1), 162-176.

Gibbs, G. (2011). Comparative reviews of software. Retrieved from http://onlineqda.hud.ac.uk/Intro_CAQDAS/reviews-of-sw.php

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

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

Grech, M., Horberry, T., & Smith, A. (2002). *Human error in maritime operations: Analyses of accident reports using the leximancer tool.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Baltimore, MD.

Guba, E., & Lincoln, Y. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Haig, B. (1995). Grounded theory as scientific method. *Philosophy of Education*, *28*(1), 1-11.

Hansson, T., Carey, G., & Kjartansson, R. (2010). A multiple software approach to understanding values. *Journal of Beliefs and Values*, *31*(3), 283-298.

Harwood, I. A. (2001). *Developing scenarios for post-merger and acquisition integration: A grounded theory of 'risk bartering'* (Unpublished doctoral thesis). University of Southampton, United Kingdom.

Harwood, I. A. (2006). Confidentiality constraints within mergers and acquisitions: Gaining insights through a 'bubble' metaphor. *British Journal of Management, 17*(4), 347-359.

Harwood, I. A., & Ashleigh, M. J. (2005). The impact of trust and confidentiality on strategic organisational change programmes: A case study of post-acquisition integration. *Strategic Change*, *14*(2), 63-75.

Harwood, I. A., & Chapman, C. B. (2009). Risk bartering: What post-M&A integration teams really do. *Strategic Change, 18*, 157-178.

Harwood, I. A., Ward, S. C., & Chapman, C. B. (2009). A grounded exploration of organisational risk propensity. *Journal of Risk Research*, *12*(5), 563-579.

Heath, H., & Cowley, S. (2004). Developing a grounded theory approach: A comparison of Glaser and Strauss. *International journal of nursing studies*, *41*(2), 141-150.

Hussey, J., & Hussey, R. (1997). *Business research: A practical guide for undergraduate and postgraduate students*. Basingstoke, UK: Palgrave Macmillan.

Illia, L., Sonpar, K., & Bauer, M. (2014). Applying co-occurrence text analysis with ALCESTE to studies of impression management. *British Journal of Management*, *25*, 352–372.

Jones, R., & Noble, G. (2007). Grounded theory and management research: A lack of integrity? *Qualitative Research in Organizations and Management: An International Journal*, *2*(2), 84-103.

Jonsen, K., & Jehn, K. A. (2009). Using triangulation to validate themes in qualitative studies. *Qualitative Research in Organizations and Management: An International Journal*, *4*(4**)**, 123-150.

Leximancer Manual. (2011). Version 4. Retrieved from https://www.leximancer.com/support/

Locke, K. D. (2001). *Grounded theory in management research*. London, UK: Sage.

Lowe, A. (1998). Managing the post-merger aftermath by default remodelling. *Management Decision*, *36*(2), 102-110.

Martin, N. J., & Rice, J. L. (2007). Profiling enterprise risks in large computer companies using the Leximancer software tool. *Risk Management*, *9*(3), 188-206.

Martschinke, H., Waugh, C., Beamish, W., & Davies, M. (2004). Layering the reflective process: An approach used to induct new staff into Griffith's practicum model for special education. In B. Bartlett & D. Roebuck (Eds.), *Educating: Weaving research into practice* (Vol. 2, pp. 260-269). Brisbane, Australia: Griffith University, School of Cognition, Language, and Special Education.

Middleton, S., Liesch, P. W., & Steen, J. (2011). Organizing time: Internationalization narratives of executive managers. *International Business Review*, *20*(2), 136-150.

O'Reilly, K., Paper, D., & Marx, S. (2012). Demystifying grounded theory for business research. *Organizational Research Methods*, *15*(2), 247-262. doi: 10.1177/1094428111434559

Partington, D. (2000). Building grounded theories of management action. *British Journal of Management*, *11*, 91–102.

Rennie, D. L. (1998a). Reply to Corbin from one interpreter to another. *Theory & Psychology*, *8*(1), 129-135.

Rennie, D. L. (1998b). Grounded theory methodology: The pressing need for a coherent logic of justification. *Theory & Psychology*, *8*(1), 101-119.

Rettie, R., Robinson, H., Radke, A., & Ye, X. (2008). CAQDAS: A supplementary tool for qualitative market research. *Qualitative Market Research: An International Journal*, *11*(1), 76-88.

Rooney, D. (2005). Knowledge, economy, technology and society: The politics of discourse. *Telematics and Informatics*, *22*(4), 405-422.

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2012). *Research methods for business students* (6th ed.). Harlow, UK: Pearson Education.

Schwarz, M., & Nandhakumar, J. (2002). Conceptualizing the development of strategic ideas: A grounded theory analysis. *British Journal of Management*, *13*, 67–82.

Seale, C. (2003). Computer-assisted analysis of qualitative interview data. In J. Gubrium & J. Holstein (Eds.), *Inside interviewing: New lenses, new concerns* (pp. 289-308). London, UK: Sage.

Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction* (2nd ed.). London, UK: Sage.

Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behavior Research Methods*, *38*(2), 262-279.

Stewart, G. & Chakraborty, A. (2010). *Strategy content analysis for service identification: A case study on government agencies*. In 5th Conference on Qualitative Research in IT, November 29-30, Queensland University of Technology, Brisbane, Qld.

Stewart, H., & Gapp, R. (2014). Achieving effective sustainable management: A small-medium enterprise case study. *Corporate Social Responsibility and Environmental Management*, *21*(1), 52-64.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing gruonded theory* (2nd ed.). Newbury Park, CA: Sage.

Weaven, S., Frazer, L., & Giddings, J. (2010). New perspectives on the causes of franchising conflict in Australia. *Asia Pacific Journal of Marketing and Logistics*, *22*(2), 135-155.

Xiao, J. Z., Dahya J., & Lin, L. Z. (2004). A grounded theory exposition of the role of the supervisory board in China. *British Journal of Management, 15*, 39-55.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Author Note

Ian Harwood, PhD, is an Associate Professor in Management. With an industrial background in managing strategic change programs and process improvement projects, his doctoral thesis explored the role of risk in post-merger and acquisition integration. Ian's continuing research interests are broadly in project management and qualitative research methods (especially Grounded Theory). He has published in the *British Journal of Management*, *Journal of Business Ethics*, *Journal of Risk Research*, *European Management Journal*, *International Journal of Project Management* and *International Small Business Journal* amongst others. He may be contacted at Ian A. Harwood at, iah@soton.ac.uk.

Rod Gapp, PhD, is a senior lecturer at Griffith University's Business School, Australia. He has over 25 years of experience in organizational development and action science focused on enhancing outcomes through individual and group development in areas of innovation, change, quality and sustainable management practices. His research focus is predominately interpretivist with extensive experience in CAQDAS and also in mixed methods. Rod has published in high ranked journals in the above areas. He may be contacted at Rod P. Gapp at, r.gapp@griffith.edu.au.

Heather Stewart holds Honors in management from Griffith University, Australia. After more than 15 years of management experience in both corporate and small business she is currently a PhD student with Griffith University's Business School, focusing on qualitative research of collaboration, continual learning and sustainable management practices in the SME context. Heather is a researcher working in qualitative data collection and analysis and is establishing a publishing profile in high level CSR and sustainable business journals. She may be contacted at Heather J. Stewart at, h.stewart@griffith.edu.au.

Article Citation

Harwood, I. A., Gapp, R. P., & Stewart, H. J. (2015). Cross-check for completeness: Exploring a novel use of Leximancer in a grounded theory study. *The Qualitative Report, 20*(7), 1029-1045. Retrieved from http://www.nova.edu/ssss/QR/QR20/7/harwood2.pdf