Using multiple data sets

Chapter 7: Using multiple data sets

Sylvia Jaworska and Karen Kinloch

7.1 Introduction

The aim of this chapter is to illuminate the exploratory and explanatory power of using multiple data sets, or in other words, data triangulation in a corpus-assisted discourse study (CADS) (Partington; Duguid and Taylor 2013). We would like, in particular, to focus on the benefits of comparisons across contexts, to which data triangulation ultimately lends itself. It was Descartes who famously proclaimed comparison as the only tool of knowledge asserting that 'it is only by way of comparison that we know the truth precisely'.[1]

For the purposes of this Chapter we adopt the critical realist perspective (Harré 2009) of the understanding of knowledge. Therefore, we see knowledge and truth as forms of social practice constituted in and through symbolic means utilised by social agents in accordance with established discursive rules. Seen from this vantage point, knowledge and truth are never static entities waiting to be discovered. They are always partial and changing dependent on social contexts, and thus never universal and impossible to know *precisely*. Nevertheless, echoing Descartes in a more postpositivist vein, we show that comparison of multiple data sets can bring us a little closer to the phenomena we study allowing richer and more comprehensive understandings, while the corpus analytical methods that we use to interrogate

[1] Original citation: 'Ce n'est que par une comparaison que nous connaissons precisement la verite', Rene Descartes, Regulae ad directionem ingenii, Rule XIV, Oeuvres Philosophiques, 9 vols., ed. Ferdinand Alquie, Paris, Garnier, 1963, vol. 1, p. 168.

the data can add more evidence-based precision and rigour to the process of data analysis, and guide interpretation.

Triangulation is not a new territory in corpus linguistics; some researchers have adopted forms of triangulation, specifically investigator triangulation (Marchi and Taylor 2009) and method triangulation (Baker and Egbert 2016) demonstrating their benefits as well as limitations for CADS research. Yet, little attention has been paid to multiple data sets and data triangulation. This chapter sets out to address this gap by presenting a hands-on framework for using data triangulation in multi-contextual CADS research.

We begin first by outlining the rationale for using data triangulation and how it can help identify blind spots and enrich CADS research. Subsequently, we move on to the nitty-gritty of the methodological decision-making involved in selecting appropriate data sources and analytical tools, and outline a hands-on and flexible framework for doing CADS with multiple data sets. How this framework can be used in practice is demonstrated in the case study, which focuses on the discursive constructions of postnatal depression in medical, media and lay accounts.

7.2 Rationale for using multiple data sets

Before we begin with articulating the rationales for using multiple data sets, we need to, at least briefly, indicate the general theoretical understanding (set of ontological and epistemological principles) that has driven our research agenda and the use of data triangulation. The choice of method(s) and analytical procedures must be congruent with and follow from the general ontological and epistemological context in which one formulates research questions.

A substantial bulk of research in (critical) discourse analysis is carried out using the post-structuralist (Williams 2014) or constructivist frameworks (Maturana and Varela 1987), which reject empiricism as the basis of inquiry and presume a relativist stance, for example, the constructivists argue that the reality is a construction of human mind. However, we argue, such research would not be congruent with corpus-linguistic tools and methods that are essentially grounded in the empirical tradition. Our ontological and epistemological position derives from the critical realist stance (Sealey 2010), especially as formulated by Harré (2009) in his notion of critical realism. For Harré (2009), the dominant form of practice of social life is that of *conversation* understood as any kind of meaningful performance (or discourse) spoken and written, produced by social actors and normatively guided by discourse conventions that both constrain and enable what one can do or say. Corpus linguistic tools and methods are well suited to study the prime practice of social life, that is, discourse as seen from the vantage point of social realism (Sealey 2010). They provide important evidence for regularities and patterns in language use through analysing what is frequently said. Equally, they can shed light on the less frequent and unusual patterns that may seem contradictory, but, in fact, show the diversity of choices made by individuals (Sealey 2010).

One important aspect emphasised by realists is that there are no universal rules that guide social practices. Rather, these are contextually dependent. The way we produce discourse varies as we move from one social context to another because each context comes with its own set of rules and conventions that enable and constrain what can be said in given circumstances. However, it needs to be stressed that, at the same time, discourses are rarely confined to a particular context and mostly travel across contexts. There might be therefore several commonalities in the ways in which a discursive phenomenon ‘behaves’ across contexts, but we would not know until we compare this ‘behaviour’ across contexts (cf.

Partington, Duguid and Taylor 2013: 12). This inevitably invites the researcher to collect data produced in different contexts and hence, to use multiple data sets. But what is context?

At least since Malinowski's seminal essay 'On the problem of meaning in primitive languages' (1923), the notion of context and the relevance of taking context into account has become a kind of linguistic truism. Previously context was defined in a narrow sense as a stretch of texts (sentences) that immediately precedes or follows a passage of interest to an analyst (what in corpus linguistic terminology would be called *cotext*). Malinowski insisted on expanding the boundaries of context beyond mere linguistic structures to the wider conditions under which speech is produced. Since then, several attempts have been suggested to conceptualise context, of which the most influential is the model proposed by Halliday based on three dimensions: *tenor*, *field* and *mode* (Halliday 1978). *Tenor* refers to the participants, their roles, goals and relations, and it is sometimes described as a domain. *Field* is understood as a subject matter (topic) and *mode* describes the channel of communication and rhetorical mode (informative, persuasive etc.). Language use is heavily dependent upon such dimensions and will change as each changes. Although the Hallidayan model of context has been critiqued for being rather static and less suitable to study digital contexts (Jones 2004), it can offer a useful heuristic for delineating contexts when compiling multiple data sets. In any case, the researcher needs to consider the key variables of communication in social settings, specifically who speaks to whom, when and for what purpose (function) (Coupland 2016) including the type of texts and the mode (e.g. spoken or written).

As any other field of linguistic inquiry, corpus linguistic research has taken context into consideration, but has so far been mostly preoccupied with selected contexts or in Hallidayan terms with one tenor or mode. Extensive research has been conducted on differences in language use across speech and writing (e.g. Biber 1998) or texts produced by, for example, learners vs. proficient users, novice vs. expert writers (e.g. Chen and Baker 2010). In CADS,

most attention has been paid to topics, themes, discourse of X, but this has been investigated predominantly in print media though arguably there is a diversity of text types within this medium. With the exception of work by Baker and McEnery (2005), Demmen et al. (2015) and cross-linguistic comparisons (e.g. Jaworska and Krishanmurthy 2012; Taylor 2014; Vessey 2016), researchers using some forms of CADS rarely venture outside these contextual boundaries. This has some benefits in that it allows the researcher to engage in depth with the chosen context. However, such an approach is limited in several ways. Firstly, findings represent the studied context only with generalisations or more broader views being impossible to formulate. Secondly, the significance of findings might be unconsciously over- or underestimated. There might be ample examples of contradictory results in other contexts, but the researcher would not know and risk overestimations. Equally, other contexts may supply further evidence and thus, strengthen findings obtained from the analysis of one context only. Similar to the blind men from the Indian folktale 'The Blind Men and the Elephant', when using one data set from one context, the researcher might be inclined to believe that the one part or pattern which he or she has found represents the 'whole' thing. Studying how a discursive phenomenon behaves across contexts with multiple data sets can liberate us from the confines of a contextual circumference. It helps the researcher discover differences and commonalties that exists in the ways in which discursive phenomena are constructed and how this depends on the participants, their roles, relations and domain in which they operate as well as the constrains and affordances of the medium. In doing so, the researcher is able to arrive at a much more comprehensive understanding of the phenomena under study; results obtained from multiple sets of data collected from different contexts can carefully guard against over- or under-interpretation. At the same time, the researcher is able to see more clearly how each context and the language used within that context differ. And vice versa, having results from multiple contexts, the researcher can see commonalties

between the contexts. This can illuminate discursive trajectories that a phenomenon leaves when travelling from one domain to another; showing aspects of discourse that are taken up, and equally aspects that are marginalised or silenced altogether. In this way, using multiple data sets can significantly increase our understanding of recontextualisation and intertextuality - that important discursive processes which simultaneously bind and transform texts, making communication possible and meaningful (Fairclough 1992). These processes are never neutral, but always intertwined with ideological positionings. As Bernstein (2000: 32-3) observes: 'every time a discourse moves, there is a space in which ideology can play. No discourse moves without ideology at play.' Investigating how a discursive phenomenon behaves in multiple contexts using multiple data sets can therefore help the researcher not only explore recontextualisations and intertextuality but also uncover the playgrounds of ideologies and help understand the mechanics of ideological work in and through discourse.

7.3 Doing CADS with multiple data sets

We need to highlight at the outset that our understanding of *data set* is consonant with a corpus and a corpus with a specific context. Therefore, each corpus represents a different but relevant context. This inevitably raises the following questions: 1) which contexts are relevant, 2) how much data is needed from each to make the multiple data sets (corpora) representative and appropriate for CADS research, and 3) which analytical corpus tools best serve a comparative inquiry? We begin by answering the first question guided by the Hallidayan model of context.

As with any piece of research, all should start with a research question. Most CADS studies are interested in a particular discourse type (topic, theme) and their representations. This

already delineates research to a specific thematic area. While Halliday's model of context offers a range of possible combinations, in a CADS research, *field* will, in most cases, remain the same, while *tenor* and *mode* can be changed to include different contexts. An ideal scenario would be to collect data from all possible participants who are involved in the production and dissemination of the discourse type in question. But this would be difficult to achieve not only because some discursive domains are huge in scope, but also due to the sometimes very strict rules (e.g. confidentiality) that govern the production and dissemination of texts. Hence, in most cases the researcher needs to make a decision and select the aspects of the domain which are the most *relevant* to his or her research questions and possible to obtain. This is inexorably linked with ethical considerations and in the case of textual data with copyright issues. At any stage of data collection, researchers are urged to check the copyright status and require permission from the copyright holder if necessary.

The question of how much data is intrinsically linked to issues involved in the corpus building, specifically representativeness and balance but these are potentially contentious when building a specialist corpus (Koester 2010). A somewhat pragmatic approach is often needed and preferred. While an ideal scenario would be to include all possible data produced in a given context in order to claim, in a scientific manner, a total accountability, this is in practice rarely possible. With some exceptions, for example, the works of Shakespeare or speeches of a famous politicians, most domains are open systems with language data being produced continuously. Most corpora or data sets are therefore subsets (samples) presenting in most cases a partial representation of a discursive phenomenon. Having said that, some contexts might be more exhaustive than others allowing the researcher to collect a good representation of the phenomenon under study. Given the contextual differences in the production and dissemination of texts, CADS research based on multiple data sets is very likely to be based on corpora of unequal sizes, which might involve the pitfalls of normalised

frequencies. To avoid this drawback, the researcher might want to balance the sizes through, for example, random sampling, but this could also mean a huge topical sacrifice in that the researcher could lose some important data. Although having unequal sizes can have implications for statistical data analysis, for CADS research it is probably more important to have data which is relevant, appropriate and exhaustive enough to address research aims. In sum, how much data to include in a comparative CADS research with multiple data sets should be a question of relevance and appropriateness rather than representativeness and balance. Prior engagements with the contexts and participants can assist the researcher in making an informed decision regarding what kind of data and how much can be collected (see Section 4).

Once appropriate and relevant data sets have been created, the next question is which analytical tools are suitable to interrogate and compare data sets that are very likely to be of unequal sizes. There are two procedures to bear in mind: the first is to consider tools and metrics that do not depend on the total size of a corpus thus allowing for meaningful comparisons across data sets of unequal sizes; the second is to use them consistently on all data sets involved.

If we want to compare the usage of a particular concept or term in corpora of different sizes, then normalised or relative frequencies need to be calculated in any case. However, the researcher needs to be aware that normalised frequencies do not give a true account of the total corpus data because language data is not normally distributed. This is why it is considered good practice in corpus-based research to provide both raw and normalised frequencies when comparing a use of a particular item across corpora (McEnery and Hardie 2012).

Keyword analysis is a useful ‘way in’ to identify salient or distinctive lexical items in multiple data sets and it can be speedily conducted using the commonly employed corpus

linguistic software programmes such as WordSmith Tools, AntConc and Sketch Engine. The way in which keywords are computed in Sketch Engine is especially useful for comparing multiple data sets because it does not rely on significance testing, which in turn depends on the sample size (cf. Gabrielatos and Marchi 2012). Whereas WordSmith Tools and AntConc use cross-tabulation and loglikelihood (LL) to compute statistically significant keywords, Sketch Engine provides a keyword score based on a normalised frequency ratio 'word W is N times as frequent in corpus X versus corpus Y' with a simple math parameter added to account for the zero problem in divisions (Kilgarriff 2005). Kilgarriff (2005) argues that the use of significance testing is problematic in keyword retrieval because all it does is to disprove the null hypothesis - that language is random, which is not. Retrieved keywords can be grouped manually into semantic domains to identify dominant topics and themes in data sets (Baker et al. 2013) and compare them across the sets to see which themes are more salient in which data set.

It needs to be noted that manual classification into semantic domains is a subjective and time consuming process. It cannot just proceed from the lists of keywords because the lists present words as isolated items 'hiding' meanings that they may have in context. Hence, checking corpus evidence by reading concordance lines is an essential procedure. In order to reduce the level of subjective judgment and ensure a better consistency, it is recommended if feasible, to use interraters (multiple judges) and measures of interrater reliability, for example, Cronbach's Alpha.

Another way of performing a keyword analysis is to compare the data sets against each other. This is often preferred by researchers who work with two corpora as it allows them to tease out differences that exist between two data sets by simultaneously avoiding problems associated with a general reference corpus. While this procedure is useful for highlighting differences in a more precise manner, it will overlook what the data sets have in common

'hiding' shared discourses (see also Chapter 2). This would limit the perspectives on how discourses travel across contexts making it impossible to explore recontextualisations and intertextuality. Using one reference corpus as a benchmark can therefore be more insightful for research based on multiple data sets in that it allows the researcher to reveal keywords that are both unique and also shared across data sets. Unique keywords highlight the contextual specificity, while shared keywords can be useful pointers to discursive recontextualizations and intertextuality.

To explore aspects of recontextualizations and intertextuality in more depth, studying collocational patterns of relevant unique and shared keywords can be very helpful. The metric for collocation retrieval offered in Sketch Engine - the Log Dice - seems particularly suitable for comparing collocations across corpora. In contrast to other commonly used statistics such as Mutual Information or T-test, Log Dice is a ratio with a maximum value (theoretically 14, but practically 10 or below) and it does not depend on the total size of the corpus (Rychlý 2008). This allows the researcher to have a consistent comparison measure across multiple data sets.

Summarising the above, we propose a framework for doing CADS with multiple data sets and combining both quantitative corpus techniques and qualitative discourse-analytical procedures (see Table 7.1). The framework is partially modelled on Baker et al. (2008) and consists of methodological procedures and practical steps to guide the researcher through the process of data collection and analysis. It needs to be noted that not all stages and steps are relevant for every project and the researcher might select those that are most suitable to answer his or her research questions.

<INSERT TABLE 7.1 HERE> (see at the end of the file)

7.4 Case Study: discursive constructions of postnatal depressions in medical, media and lay contexts

To demonstrate how the framework can be put in practice, this section presents a case study which explores the discursive constructions of postnatal depression (PND) in medical, media and lay accounts. It does so by outlining the process of data collection, ethical considerations and analytical tools selected for comparisons of multiple data sets.

The case study forms a part of a larger project which investigates public discourses around PND (Jaworska and Kinloch 2016). PND is a type of depression which can occur within one year of childbirth; it is a highly stigmatised condition, which in the UK affects 10-15% of mothers, with suicide due to PND being the leading cause of maternal death (NHS 2016).

7.4.1 Data Collection

The first question which needs to be answered is what are the contexts in which discourses of interest to the researcher are likely to be produced and disseminated. PND is a mental health condition and the obvious answer is the medical domain. But the medical domain is a multi-layered profession and an industry with many sites and participants who have different status, roles and goals. It includes medical researchers, clinicians and practitioners who all are involved in production and 'consumption' of a variety of texts and operate in various modes sometimes simultaneously. The ideal would be to collect data about PND from all possible participants but this would be difficult to achieve not only because of the enormous scope of the domain, but also due to the strict confidentiality that surrounds dissemination of texts in medical contexts.

Our initial interest was in lay discourses of PND and how they are influenced by wider discourses around the condition disseminated in the UK. For the purpose of our study, Context 1 were conversations about PND produced by lay participants in online discussions

on Mumsnet. Mumsnet is the largest online parenting forum in the UK attracting over 6 million unique visitors per month. The data are examples of spontaneous written conversations and the dominant purpose is information and support. Since the lay person was our prime focus, we decided to select medical contexts and texts that a lay person with PND is likely to encounter. This was based on the authors' insider knowledge and experience of being involved with mothers' groups and personal encounters with women who had PND. Through the engagement with the participants, it became clear that mothers learn about PND primarily through consultations with medical professionals (GPs, health visitors, midwifes) and by reading materials produced by medical professionals for lay people, many of which are distributed online and mothers are often directed to consult these resources. In this way, we were able to select a subdomain from the domain of medical profession that was directly relevant to our study and offered pointers to texts that were produced for and used by women with PND. Context 2 was, therefore, written texts about PND produced by medical professionals for lay people. Because most of the texts produced in Context 2 were derived from medical context *per se* and included references to medical, academic and clinical literature, we also decided to include medical literature about PND produced by medical professionals for medical professionals. Since media play a significant role in the dissemination of discourses around health and illness and this significance became apparent when engaging with Context 1, we also decided to include news stories about PND published in the major British national newspapers since 2000. Table 7.2 presents the four contexts that were considered in our study. The appropriateness of each context is ensured by the focus on the topic, the relevance of texts produced in each context and the geographical location (UK).

<INSERT TABLE 7.2 HERE>

Once we settled on the contexts from which to obtain data, the next question was how much data to collect. Again, an ideal scenario would be to include all possible language data about PND produced in the four contexts, but this was not possible, because some of the domains (Mumsnet) are open systems with language data being produced continuously. As far as online data is concerned, we first identified threads that had *postnatal depression* or *PND* mentioned in the thread and downloaded using the all posts from the selected threads. This generated a corpus (Data Set 1) of 4,778,285 words, which we considered large given its 'specialist' status. Other domains were more closed in nature leading to smaller data sets. For example, in Context 2, we included guidelines and information brochures produced by the main health service provider in the UK, the National Health Service (NHS), and affiliated medical organisations or charities to which mothers are directed on the websites of NHS including the Royal College of Psychiatrists (RSP), National Collaborating Centre for Mental Health (NCCMH), Association for Post Natal Illness (APNI), PANDAS Foundation and private healthcare providers such as BUPA. It quickly became apparent that there was a great deal of similarity and repetition in the texts produced in Context 2, which is not surprising given that healthcare providers need to get across consistent information. The data set collected from Context 2 was therefore considerably smaller than Data Set 1 and included 50,113 words, but it was felt to be exhaustive and representative of Context 2. A similar collection method was used in Context 3 which comprises professional guidelines for health care professionals in the field of perinatal mental health. The total size of data collected stands at 187,940 words.

As far as Context 4 is concerned, the procedure was straightforward; we built on previous CADS research and utilised the newspaper database LexisUK to obtain relevant data since 2000. We settled on the year 2000 because some of the important national newspapers with the highest number of readers (e.g. *Daily Mail*, *Sun)* were only added in the late 1990s or

2000. We wanted to make sure that our corpus represented a variety of sources and not only broadsheets. While the choice of the start point might appear arbitrary, it was justified (and constrained) by the availability of data. We used again the terms *postnatal depression* and *PND + depression* to retrieve topical articles with the research terms occurring 3 times or more in the text. Newspapers differ in terms of coverage (regional vs national), style (e.g. broadsheet vs tabloid), and despite supposed neutrality, always have some kind of political leaning. These are important factors that impact dissemination, audiences, style and persuasion and need to be taken into consideration when collecting and analysing media data. Since we were interested in wider discourses around PND, the decision was made to include UK national newspapers only and as our project was not concerned with strictly political matters, the political orientation of a newspaper was a lesser concern to us. We made a distinction between broadsheets, tabloids and middle-range tabloids to account for the degree of formality and sensationalism. This led to the creation of a media corpus (Data Set 4), which consisted of 845 articles with 1,585,954 words. Table 7.3 shows the size of each data set.

<INSERT TABLE 7.3 HERE>

Throughout the process of data collection, we engaged with ethical matters concerning the data. Medical guidelines were the least problematic as they are in the public domain designed for public consumption. Newspaper articles are protected by copyright laws but newspaper data can be collected for non-commercial research purposes and if single articles are not distributed as a whole, permission is not normally required. [2] Online conversations produced

[2] The British Library offers a useful guide on copyright issues regarding newspaper articles, see http://www.bl.uk/reshelp/findhelprestype/news/copynews/

by human subjects, even if anonymously, present more of a grey area and researchers are divided by the opinion how to treat them. Some argue vehemently that informed consent from online participants should be obtained in any case, whereas others insist that by posting anonymously participants automatically give their consent (Roberts 2015). No agreement has been yet reached, although guidelines in this area have been produced by the Association of Internet Researchers (AOIR) and the British Psychological Society[3] and we consulted these.

Our online data was collected from a discussion form on Mumsnet called *Talk*. *Talk* is a public forum that can be browsed by members and non-members, but only registered members can post. The terms and conditions of Mumsnet stipulate that *Talk* is a public space and users are made aware that anyone can view their posts. Following procedures adopted in previous research using posts from Mumsnet (e.g. Pedersen 2016), consent was not sought from the participants because the material used was not directly elicited from them and only obtained after it was spontaneously generated. Yet, the terms and conditions of Mumsnet state that all content published on its site including *Talk* are the sole property of Mumsnet and reproduction of any parts without approval is prohibited. Consent was therefore sought from Mumsnet to use *Talk* data and approval was granted. The approval stated that usernames or any other potentially identifying details must be removed to protect posters' anonymity and this procedure was adopted throughout.

7.4.2 Analysis

This section shows how the retrieval of keywords and a subsequent classification of keywords into semantic domains can provide a useful way in to multiple datasets and how the further interrogation of selected keywords using Word Sketch can give insights into how

[3] https://aoir.org/reports/ethics2.pdf and https://beta.bps.org.uk/news-and-policy/ethics-guidelines-internet-mediated-research-2017

discourses are taken up and potentially contested across contexts. To showcase the rather neglected area of similarities (Taylor 2013) and recontextualisations, we focus on shared keywords only.

Sketch Engine was used to produce and compare keyword lists from our data sets using the BNC as the reference corpus. Subsequently, we selected the 100 most distinctive content keywords in each data set and grouped them manually into semantic categories, a procedure adopted from previous CADS research (Baker et al. 2013).

It is worth clarifying at this point what we gloss as 'semantic domains' are thematic categories developed inductively and reiteratively from studying the keyword data, as distinct from automated classification through tools such as Wmatrix (Rayson 2008). This inductive process means that the coding for some categories is more finegrained than others, as appropriate to the particular topic. In the case of PND, we first began with identifying general categories, for example, Actors, Medical Actions, Emotions etc., but quickly noticed that there exist subcategories within each general category. For example, the general category Actor included a variety of actors that could be further grouped into subcategories depending on their role in the process. This approach reflects the multiplicity of discourses and practices around the biomedical model of perinatal mental health. In Table 7.4 below we show the set of semantic categories developed and examples of the keywords for each domain, whilst Figure 7.1 illustrates the comparison of their normalised frequency across the four corpora.

<INSERT TABLE 7.4 HERE>

<INSERT FIGURE 7.1 HERE>

The comparison of semantic domains across corpora is one method for eliciting the commonalities and differences on a particular topic, highlighting the specificity of each context. For example, it is not surprising to see that Medical Labelling plays a much more important role in MEDICAL and MEDLAY, while Emotions, particularly Negative emotions, are prevalent in MEDIA and MUMSNET. However, while Figure 7.1 provides a broad-brush view of the topics for closer analysis, we turn to the keywords which are found in all of the corpora. While we acknowledge the usefulness of looking at keywords unique to a particular tenor or mode, for example, in the Mumsnet corpus the use of acronyms such as DD (dear daughter) and FF (formula feeding) is a stimulating topic for stylistic investigation; this is not central to this particular cate study.

A calculation of shared terms in the top 100 keywords for each corpus elicited 21 words for investigation of patterns and paths of recontextualisations including *women*, *NHS*, *GP*, *midwife*, *baby*, *mothers*, *mother*, *babies*, *child*, *birth*, *pregnancy*, *health*, *illness*, *breastfeeding*, *anxiety*, *help*, *depression*, *postnatal*, *mental*, *PND*, *parenting*. In the example analysis, we address one keyword from the Medical Labelling semantic domain, *depression*, and one from the Personal Actors domain, *mother*, in order to illuminate the discursive constructions and recontextualisations of the medical condition itself and the key social actor in these texts. We begin with the keyword *depression* and its Word Sketches across the corpora.

The most frequent collocational pattern of *depression* identified using the Word Sketch function is modifier + depression and the significant collocates are shown in Table 7.5 below. We consider collocations with the Log Dice value of 7 or above, which points to very strong associations (Rychlý 2008) and the minimum frequency of 3.

<INSERT TABLE 7.5 HERE>

The modifiers of *depression* indicate the extent to which the biomedical model of depression is accepted and recontextualised from medical texts across media and lay accounts of PND. The use of diagnostic modifiers, such as *postnatal*, *antenatal*, and medical gradation markers, such as *severe*, *major*, *mild* are used even in lay accounts, demonstrating both acceptance of this model which while validating the lived experience through the adoption of biomedical explanation potentially destigmatises those who have PND. Interestingly however, the MUMSNET corpus shows use of two modifiers that were not found in the other data sets, that is, *reactive* and *chronic*. The term *reactive* is a scientific term from the domain of chemistry, which is sometimes used in psychiatry to refer to the recurrence of mental illness (Oxford English Dictionary, OED). Similarly, *chronic* is a medical term which is used to describe long lasting and intensive illness. Both emphasise the temporality of PND as an ongoing or recurring event in the lived experience. This aspect of PND seems unaccounted in MEDICAL and MEDLAY which emphasise depression as a stable entity to be diagnosed and treated (see below). There are other conspicuous differences that may warrant further investigation. For example, the use of *manic* and *terrible* in MEDIA is striking and not matched by other domains potentially hinting at sensationalist media attitudes towards PND.

The second most frequent collocational pattern is VERB with depression as an object and it is to this we turn in the next part of our analysis to unpick how postnatal depression is acted upon (see Table 7.6).

<INSERT TABLE 7.6 HERE>

In comparison with the modifiers of *depression*, the collocational patterns in Table 7.6 indicate greater variation across the corpora. Whilst the lexis of biomedical action, for example *diagnose, treat, prevent*, occurs across all four corpora, the lexis of the experience of PND varies. While more neutral terms such as *have* and *experience* show high significance in all texts, in the media and lay corpora *suffer* is the strongest association, potentially passivizing those who experience PND. Interestingly, the linguistic choices in the media contexts draw on well-documented, war related metaphors for illness (Semino et al. 2017) as indicated in the prominence of *battle, beat, fight, tackle,* and *combat* in MEDIA. While previous research suggests the prominence of war and fight metaphors in the experience of illness, our findings confirm it only to the domain of media suggesting that the use of this type of metaphors might be condition- and context-specific.

Drawing on the idea of an explanatory model for destigmatizing the experience of PND, the patterns *cause*/*trigger depression* are prevalent in MEDIA and MUMSNET but less so in texts produced by medical organisations. Interestingly, when *depression* is an object of *cause* in MEDLAY, 4 out of the 8 occurrences point to unknown aetiology and 4 are carefully formulated references to social and bodily factors accompanied by question marks or hedges (see Figure 7.2).

<INSERT FIGURE 7.2 HERE>

In both MEDIA and MUMSNET the patterns *cause*/*trigger depression* elicit examples of a desire to explain PND either through external social factors, as in the reactive or chronic depression explanatory model, or hormonal/chemical imbalances. Interestingly, the latter themes occur only in MEDIA and MUMSNET suggesting that women possibly draw on MEDIA discourses to explain PND (see Figure 7.3 and 7.4).

<INSERT FIGURE 7.3 HERE>

<INSERT FIGURE 7.4 HERE>

The pressure of the expectations of parenthood versus the imperfect reality and the strain of living up to the public image of the 'ideal' mother are often cited as possible causes of PND (e.g. Kantrowitz-Gordon 2013). In order to explore this further, as the second part of this example we look at the representation of the key social actor in question, that is, the keyword *mother*.

The representation of *mother* is integral to any study of the discursive construction of postnatal depression as the conditions affects mostly mothers. We focus on the pre-modification of *mother* using Word Sketch to show salient ways in to how mothers are characterised and evaluated across the contexts.

<INSERT TABLE 7.7 HERE>

As seen in Table 7.7, the most striking points of the pre-modification of *mother* are the foregrounding of *new* mothers and the evaluative extremes which are present in the MEDIA and MUMSNET datasets. The highlighting of inexperience in the terms *new* and *first-time* in co-occurrence with PND constructs new mothers as a vulnerable group in need of medical attention. This construction occurs in the MEDICAL corpus:

> (1) All **new mothers** and their partners would benefit from sensitive and supportive care from consistent professionals during the perinatal period. (MEDICAL)

Also, the expectation that some level of depression or mental distress is ‘normal’ for new mothers is ‘rehearsed’ in the other domains:

(1) Remember that some these things can also be a normal part of being a **new mother**, such as disturbed sleep or lack of energy. (MEDLAY)

(2) The gruelling “graveyard shift" can reduce even the most level-headed **new mother** to a stressed-out zombie. (MEDIA)

(3) Many **new mothers** are misdiagnosed as having PND when really they need to be told that what they're experiencing is pretty normal. (MUMSNET)

This is potentially problematic for women experiencing PND as they may resist help-seeking due to the expectation of the problematic emotional experience of being a ‘new mother’ and the feeling that as they become more experienced at childcare this will dissipate.

The societal expectations of mothers and related judgemental attitudes towards this group clearly manifest in the wide range of evaluative and polarising lexicon including *perfect*, *good*, *bad*, *terrible*, before we even begin to approach the problematic constructions of *working*, *stay at home*, *single*, *FF* (formula feeding) and *BF* (breastfeeding) which also modify *mother*. The use of this evaluative language is again less frequent in the medical texts (both MEDICAL and MEDLAY) and is used exclusively in the context of reassurance that help-seeking for PND does not equate to being a ‘bad’ mother and pressure to be a good mother is unhelpful. Indeed, the MEDLAY corpus highlights mother guilt or fear of being a ‘bad’ mother as a potential symptom of perinatal mental health problems. The MEDIA and MUMSNET corpora also show the construction of *bad mother* is most commonly used in the

context of fear and stigmatisation around a diagnosis of PND, and seek to provide reassurance:

(1) Many are afraid to tell their health visitors how they feel for fear of having their children taken away, or being seen as **bad mothers**. (MEDIA)

(2) You are being the best **mother** you can be and seeking help when you need it. You are not a **bad mother** and guilt will not help either you or your dd. (MUMSNET)

While the non-medical corpora show a self-reflexive awareness of how expectations around mothering can be problematic, it is also clear in the self and other construction of mothers that fear of stigma and possible aversion to seeking help are critical issues. But why, if it is acknowledged in this range of texts that motherhood is difficult, poorly supported and potentially stigmatised is this problematic construction of 'perfect' mother still perpetuated? The fact that MEDIA and MUMSNET refer to this wide range of evaluative and polarising lexis shows the perpetuation of normative ideologies surrounding motherhood; as soon as a woman gives birth, she is automatically subject to powerful moral judgments that deem her either 'good' or 'bad' and there is very little in between.

7.5 Evaluation

Using multiple data sets allowed us to gain a much more profound understanding of discourses around PND by reducing some of the blind spots that often lurk in discourse analysis based on a single set of data. It allowed us to see the contexts with fresh perspectives and to notice much more clearly similarities and differences among the data sets. Investigating keywords shared across contexts revealed aspects of recontextualisation by

showing how specific discourses are appropriated in multiple ways to fulfil distinctive purposes. An example of this is the way in which women with PND recontextualise the voice of biomedicine via a pseudo-scientific explanatory model of chemical imbalances to legitimise the otherwise stigmatised condition (Kantrowitz-Gordon 2013). Further comparisons of keywords enabled us to discover discursive specificities and absences in the contexts we studied, which brought to light a number of more concrete and evidence-based implications of our research. A good example of this is the significance of the temporality in the lived experience of PND so prominent in the lay accounts but clearly absent in the medical understanding of the condition. Similarly, results from the multiple data sets illuminated how discourses around PND are intrinsically interwoven with powerful societal ideologies about motherhood and how these are reproduced in the media and lay accounts and again are absent from medical texts. Given the power of the biomedical model (also demonstrated here), engagement on the part of medical professions with ideologies of 'good' motherhood and specifically the unrealistic expectations impressed on mothers could be a possible way forward to help reduce the stigma surrounding PND.

We hope that the many benefits of using multiple data sets in CADS research are now evident. Having said that, there are some caveats that need to be kept in mind before embarking on this kind of research journey. Firstly, the approach can be time-consuming, especially the task of data collection. Secondly, CADS with multiple data sets can benefit from the knowledge of and exposure to contexts under study and this may not always be possible. Thirdly, we focused here on the synchronic perspectives of a discursive phenomenon leaving out the historical or (modern) diachronic dimensions that could add yet other valuable insights. Fourthly, our research was positioned in one geographical and cultural context and the results cannot in any way be generalised beyond it. Replicating the same procedures in a different cultural and linguistic environment might lead to very

different results. Despite these limitations, we feel this method has much to add to the study of discursive constructions of social phenomena, recognizing and interrogating discursive mobilities and recontextualisations which are retained across contexts.

References

Baker, P. 2006. *Using corpora in discourse analysis*. London: Continuum.

Baker, P. and Egbert, J. eds. 2016. *Triangulating methodological approaches in corpus-linguistic research*. London: Routledge.

Baker, P. and McEnery, T. 2005. A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics* 4(2), 97-226.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M. McEnery, T. and Wodak, R. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society* 19(3), 273-306.

Baker, P., Gabrielatos, C. and McEnery, T. 2013. *Discourse analysis and media attitudes.* Cambridge: Cambridge University Press.

Bernstein, B. 2000. *Pedagogy, symbolic control, and identity*. Oxford: Rowman & Littlefield.

Biber, D. 1998. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Chen, Y. H. and Baker, P. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14(2), 30-49.

Coupland, N. 2016. Sociolinguistic theory and the practice of sociolinguistics. In: N. Coupland, ed. *Sociolinguistics: Theoretical debates*. Cambridge: Cambridge University Press, 1-34.

Demmen, J., Semino, E., Demjen, Z., Koller, V., Hardie, A., Rayson, P. and Payne, S. 2015. A computer-assisted study of the use of violence metaphors for cancer and end of life by patients, family carers and health professionals. *International Journal of Corpus Linguistics* 20(2), 205-231.

Fairclough, N. 1992. *Discourse and social change.* Oxford: Polity Press.

Gabrielatos, C. and Marchi, A. 2012. Keyness: Appropriate metrics and practical issues. CADS International Conference 2012. 13-14 September, University of Bologna, Italy, http://repository.edgehill.ac.uk/4196/1/Gabrielatos%26Marchi-Keyness-CADS2012.pdf (Accessed: 28 June 2017).

Halliday, M. A. K. 1978. *Language as a social semiotic: the social interpretation of language and meaning.* London: Edward Arnold.

Harré, R. 2009. Saving critical realism. *Journal for the Theory of Social Behaviour* 39(2), 129-143.

Jaworska, S. and Krishnamurthy, R. 2012. On the F-word: a corpus-based analysis of the media representation of feminism in British and German press discourse, 1990-2009. *Discourse & Society* 23(4), 1-31.

Jaworska, S. and Kinloch, K. 2016. On 'bad' mothers and hormonal imbalances: Comparing discursive constructions of postnatal depression in lay, media and medical accounts. BAAL SIG Health and Science Communication: Experiences of illness and death. 28 November, Open University, UK, https://www.researchgate.net/publication/317225754_On_'bad'_mothers_and_hormonal_imbalances_Comparing_discursive_constructions_of_postnatal_depression_in_lay_media_and_medical_accounts (Accessed: 28 June 2017).

Jones, R. 2004. The problem of context in computer-mediated communication. In: P. LeVine and R. Scollon, eds. *Discourse and technology: Multimodal discourse analysis*. Washington DC: Georgetown University Press, 20-33.

Jones, R. 2013. *Health and risk communication*. London and New York: Routledge.

Kantrowitz-Gordon, I. 2013. Internet confessions of postpartum depression. *Issues in Mental Health Nursing* 34(12), 874-882.

Kilgarriff, A. 2005. Language is never ever ever random. *Corpus Linguistics and Linguistic Theory* 1(2), 263-276.

Kim, J., La Porte, L., Corcoran, M., Magasi, S., Batza, J., and Silver, R.K. 2010. Barriers to mental health treatment among obstetric patients at risk for depression. *American Journal of Obstetrics and Genecology* 202(3), 1-5.

Koester, A. 2010. Building small specialised corpora. In: A. O'Keeffe, and M. McCarthy, eds. *The Routledge handbook of corpus linguistics.* London: Routledge, 66-79.

Malinowski, B. 1923. The problem of meaning in primitive languages. In: C. K. Ogden and I.A. Richards, eds. *The meaning of meaning. A study of the influence of language upon thought and of the science of symbolism.* Reissue Edition. London: Kegan Paul.

McEnery, T. and Hardie, A. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Marchi, A. and Taylor, C. 2009. 'If on a winter's night two researchers… A challenge to assumptions of soundness of interpretation'. *Critical Approaches to Discourse Analysis across Disciplines* 3(1), 1-20.

Maturana, H. R. and Varela, F. J. 1987. *The tree of knowledge: The biological roots of human understanding*. Boston: Shambhala Publications.

National Health Service (NHS). 2016. Postnatal depression. http://www.nhs.uk/Conditions/Postnataldepression/Pages/Introduction.aspx (Accessed: 28 June 2017).

Partington, A., Duguid, A. and Taylor, C. 2013. *Patterns and meanings in discourse. Theory and practice in corpus-assisted discourse studies (CADS).* Amsterdam: Benjamins.

Pedersen, S. 2016. The good, the bad and the 'good enough' mother on the UK parenting forum Mumsnet. *Women's Studies International Forum* 59, 32-38.

Rayson, P. 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4), 519-549.

Roberts, L. D. 2015. Ethical issues in conducting qualitative research in online communities. *Qualitative Research in Psychology* 12(3), 314-325.

Rychlý, P. 2008. A lexicographer-friendly association score. In: P. Sojka and A. Horák, eds. *Proceedings of Recent Advances in Slavonic Natural Language Processing*, RASLAN 2008. Brno: Masaryk University, 6-9.

Scott, M. 2010. Problems in investigating keyness, or clearing the undergrowth and marking out trails. In: M Bondi and M. Scott, eds. *Keyness in texts*. Amsterdam: Benjamins, 43-58.

Sealey, A. 2010. Probabilities and surprises: a realist approach to identifying linguistic and social patterns, with reference to an oral history corpus. *Applied Linguistics* 31(2), 215-235.

Semino E, Demjén Z, Demmen J, Koller, V, Payne, S, Hardie, A. and Rayson, P. 2017. The online use of Violence and Journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study. *BMJ Supportive & Palliative Care* 7, 60-66.

Taylor, C. 2013. Searching for similarity using corpus-assisted discourse studies. *Corpora* 8(1), 81-113.

Taylor, C. 2014. Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics* 19(3), 368-400.

Vessey, R. 2016. *Language and Canadian media: Representations, ideologies, policies.* London: Palgrave.

Williams, J. 2014. *Understanding poststructuralism*. London: Routledge.

Tables

Table 7.1: Framework for CADS with multiple data sets

1. Settle on research aims and questions and ground them in a relevant theoretical, ontological and epistemological model; bear in mind that it may not be appropriate or necessary to use corpus-based techniques within some models; 2. Consider contexts in which the topic or a type of discourse is produced and disseminated; 2a. if necessary narrow down the contexts to key domains, to make the analysis feasible; relevance to the research questions should guide the selection; 2b. delineate contexts demonstrating their relevance to the research aims; the Hallidayan model of context and consideration of who speaks to whom, when and for what purpose offer a useful heuristic. 3. Identify data sources within the contexts that are most relevant and appropriate to address the research questions; prior engagement with contexts and participants, if possible, can offer relevant pointers;

4. Collect data bearing in mind the access, availability, amount, ethical considerations and copyright issues;

5. Interrogate the data sets using quantitative corpus-based techniques to identify lexico-grammatical and semantic/thematic patterns, and to select texts or potential sites for further qualitative analyses (step 6);

5a. keywords are a useful way in; identify keywords of interest and worth investigating further;

5b. classify the most distinctive keywords into semantic domains to explore dominant themes; the use of interraters and statistical measures of consistency can help to produce a more robust classification;

5c. study collocations of selected keywords bearing in mind the differences in outputs that the commonly used metrics produce; collocations of shared keywords can point to paths of recontextualization and shed light on intertextuality and interdiscursivity; unique keywords highlight the distinctiveness of a given data set;

5d. study frequencies and collocations of selected lexical items (not keywords) that name and reflect the studied discourse;

6. Interrogate subsections of the data sets using (critical) discourse-analytical techniques.

6a. study concordance lines to identify specific usage of an item or a collocation paying attention to devices that were not accounted for by corpus interrogation (for example, pragmatic markers and pragmatic patterns, metaphors)

6b. go into the text to explore additional discourses and strategies

7. Consider sources outside the data sets (dictionaries, manuals, historical records, statistical/demographic data etc.) to further explore and contextualise the

studied discourse/lexical item (this step should be integrated at every stage of research).

8. All steps could be replicated in another linguistic context adding a cross-linguistic and cross-cultural dimension to the original research.

Table 7.2: Contexts of PND in the UK

Context	Context 1	Context 2	Context 3	Context 4
Topic	PND discourse	PND discourse	PND discourse	PND discourse
Participants	Lay participants with experience of PND	Medical professionals writing for lay audience	Medical professionals writing for medical professionals	News media
Mode	written as spoken	written	written	written
Purpose (function)	informative, support	informative	informative, legislative, clinical	Informative, persuasive

Table 7.3: Sizes of Data Sets

Data Set	**Corpus Name**	**Words**
Context 1 => Data Set 1	MUMSNET	4,778,285
Context 2 => Data Set 2	MEDLAY	50,113
Context 3 => Data Set 3	MEDICAL	187,940
Context 4 => Data Set 4	MEDIA	1,585,954

Table 7.4: Semantic categories

Semantic Category	Examples of keywords
A ACTOR GENERAL	woman, women
AI Institutions as Actors	NICE, NHS, PANDAS, Hospital
AM Medical Actors (Roles)	Midwife, GP, health visitor
AP Personal Actors (Roles)	mum, mother, baby, husband, child, family
B BODILY EXPERIENCE	birth, breastfeeding, pregnancy, tiredness
E EMOTIONS	feel, feeling
EN Negative emotions	worry, stressed, upset, hate
EP Positive emotions	happy, love, lucky, hopefully
MA MEDICAL ACTION	healthcare, admission
MA_A Alternative therapies	CBT, counselling, therapy
MA_M Medical management	screening, identification, refer
MA_P Pharmacological intervention	antidepressants, medication, drug
ML MEDICAL LABELLING	depression, disorder, psychiatric, postnatal

Table 7.5: Modifiers of *depression*

MEDICAL		MEDLAY		MEDIA		MUMSNET	
Collocate	LD	Collocate	LD	Collocate	LD	Collocate	LD
postnatal	12.82	postnatal	13.53	postnatal	13.74	postnatal	12.16
major	10.46	severe	10.47	severe	10.41	post-natal	11.27
post-natal	10.21	post-natal	9.86	post-natal	10.40	antenatal	10.84
antenatal	10.03	moderate	9.34	antenatal	9.15	severe	10.27
severe	9.99	mild	8.99	clinical	9.06	ante-natal	9.72
perinatal	9.80	antenatal	8.51	manic	8.56	mild	8.68
minor	9.69	postpartum	8.48	maternal	7.61	reactive	8.47
maternal	9.25	clinical	8.19	paternal	7.48	chronic	7.76

moderate	9.22	major	8.16	serious	7.34	serious	7.72
paternal	8.71	maternal	8.16	prenatal	7.18	previous	7.32
mild	8.70	untreated	7.80	mild	7.18	bad	7.30
untreated	8.29	disorder	7.78	terrible	7.15	untreated	7.28
possible	8.17	previous	7.75				
non-remitted	7.32						
blue	7.32						

Table 7.6: Verbs with *depression* as an object

MEDICAL		MEDLAY		MEDIA		MUMSNET	
Collocate	LD	Collocate	LD	Collocate	LD	Collocate	LD
experience	11.50	have	11.75	suffer	12.46	suffer	10.52
treat	10.72	experience	10.88	develop	10.55	cause	9.70
detect	10.11	treat	10.80	experience	10.15	develop	8.86
diagnose	9.91	prevent	10.40	battle	9.83	treat	8.80
prevent	9.64	develop	10.30	diagnose	9.74	trigger	8.79
associate	9.58	cause	10.25	have	9.72	experience	8.78
compare	9.57	make	9.83	treat	9.25	diagnose	8.24
target	9.42	diagnose	9.79	get	9.20	have	8.06
identify	9.34	be	9.48	cause	9.09	lift	7.85
assume	9.02	include	9.12	trigger	8.93	cure	7.75
get	8.94	understand	9.04	prevent	8.71	underlie	7.21
have	8.75	recognise	9.04	beat	8.05	understand	7.20
develop	8.74			tackle	7.86	prevent	7.12
address	8.67			fight	7.86		
include	8.40			combat	7.65		

Table 7.7: Modifiers of *mother*

MEDICAL		MEDLAY		MEDIA		MUMSNET	
Collocate	LD	Collocate	LD	Collocate	LD	Collocate	LD
new	11.95	new	12.46	new	11.91	new	9.97
in-patient	10.42	depressed	11.58	single	10.47	bad	9.73
depressed	10.42	many	11.17	bad	9.91	single	9.03
specialist	9.94	bad	10.76	depressed	9.73	other	8.93
many	9.30	non-depressed	10.32	good	9.66	good	8.71
expectant	9.15	most	9.91	young	9.56	perfect	8.65
group	8.60	other	9.69	other	9.54	most	8.49

first-time	9.52	bf	8.46
many	9.45	own	8.35
perfect	9.38	many	8.27
		time	8.22
		terrible	8.21
		young	8.16
		ff	8.14

Figures

Figure 7.1: Semantic categories across contexts

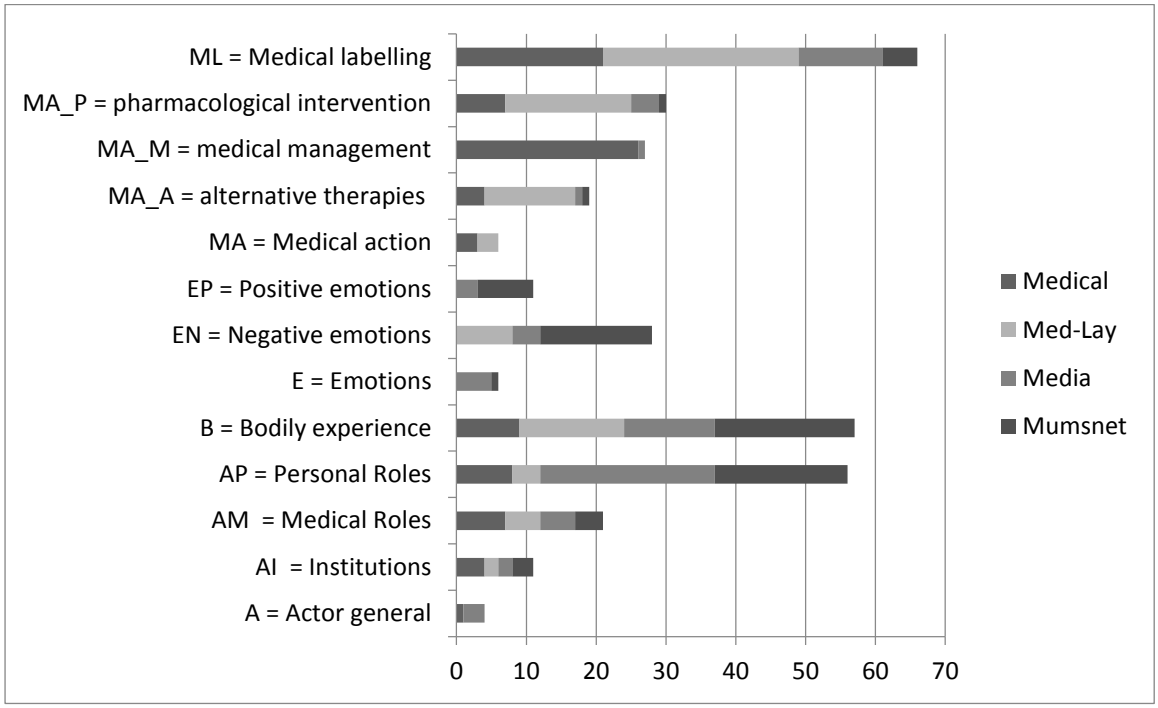

Figure 7.2: Concordance lines of the pattern *cause* and *depression* in MEDLAY

if PND is suspected. What *causes* postnatal **depression** ? PND can affect a new mum regard
postnatal depression? The *cause* of postnatal **depression** isn't completely clear. Some of the fa
your baby and family. " The *cause* of postnatal **depression** isn't clear, but it's thought to be the r
vulnerable to infections. What *causes* postnatal **depression** ? No one really knows for sure;
help if you are like this. What *causes* postnatal **depression** ? The exact cause is not clear.

Figure 7.3: Concordance lines of the pattern *cause* and *depression* in MEDIA

with antidepressants to no effect. Their **depression** is caused by a hormone imbalance that is
resented itself to the naked eye as postnatal **depression** *caused* by postpartum hormonal flux, it
helpline was part of a wider study into **depression** *caused* by pregnancy and the arrival of
these negative emotions *causes* postnatal **depression** . Here's the good news children for
it? Hormone imbalance a role in *causing* **depression** . Feeling overwhelmed of having a baby
serotonin and dopamine. Low levels *cause* **depression** but the boffins say drugs could control
baby when she sank into severe postnatal **depression** *caused* by her husband cheating on her,
compounds in the brain that may *cause* **depression** . Previous studies have suggested that
, both to understand mechanisms that *cause* **depression** and to find a new treatment for the one
whether changes in brain chemicals *cause* **depression** , or result from it. SUZI'S ILLNESS

Figure 7.4: Concordance lines of the pattern *cause* and *depression* in MUMSNET

the chemical vs social argument Some **depression** is *caused* by chemical imbalances- I never
tested for thyroid function as it *causes* **depression** as well as other symptoms (weight gain,
that it was the anxiety that *caused* the **depression** . I'm still suffering badly with anxiety
which let's face it a baby is) can *cause* **depression** , if you add in lack of sleep too it can
where you are. Ironically, a big part of my **depression** was *caused* by my not being able to bf.
issues in my life that have been *causing* my **depression** /general low mood. That's another reason
their babies. I think that postnatal **depression** is *caused* by chemistry of the brain, amount
medication sometimes forever whereas other **depression** is *caused* by something such as grief and
because of stress & depression. The stress & **depression** were *caused* by crap in my life. If I'd
having a baby in our society that *causes* **depression** in many and instead plants the idea that