

CHAPTER TWO

Corpus linguistics in illness and healthcare contexts: A case study of diabulimia support groups

GAVIN BROOKES


1 INTRODUCTION

Corpus linguistics is a group of methods that use specialist computer programs to analyse large collections of naturally occurring texts (McEnery and Hardie, 2012). These data sets are known as corpora (singular **corpus**, Latin for ‘body’) and are assembled with the aim of representing a language or linguistic variety on a broad scale. Due to the ease with which they can be analysed using computational methods, corpora are often much larger and more representative than the types of data that tend to be analysed using manual, purely qualitative methods of linguistic analysis. Indeed, corpora typically run into millions, and sometimes billions, of words in size, importantly providing researchers with the opportunity to base their analyses on more representative bodies of text and so to make their findings more generalizable to the social group, context or linguistic variety under study.

The aim of this chapter is to provide a detailed introduction to corpus linguistics, in preparation for the subsequent chapters in this volume, and to

demonstrate how corpus methods can be used to approach language ~~used~~ in particular illness and healthcare contexts, showcasing some of the main benefits of applying corpus methods in this area. These aims will be met by way of a case study examining the language used in messages posted to an online support group concerned with a contested eating disorder known as ‘diabulimia’. The next section of this chapter provides an overview of the topic of diabulimia, detailing its medical status, prevalence and health implications for those who experience it. Following this, a more detailed introduction to corpus linguistics is given, before the focus of the chapter turns to the diabulimia case study, first introducing the data and then the analysis, which makes use of three established corpus techniques (keywords, collocation and concordance). The findings are then discussed in terms of their possible implications for medical practitioners caring for people with ~~experience of~~ diabulimia in the future. The chapter concludes with a reflection on the utility of corpus methods for health(care) communication research, considering both their strengths and drawbacks in relation to this burgeoning field of linguistic inquiry.

2 DIABULIMIA IN CONTEXT

Diabulimia is a contested eating disorder whereby people with type 1 diabetes deliberately reduce their insulin intake in order to shed calories and ~~lose or~~ control their body weight. Although some people who experience diabulimia might conceive of and talk about it as if it were a medical disorder, the reality is that it is not recognized as such by practitioners and medical authorities, who instead tend to view it as an inappropriate compensatory behaviour, as well as a marker of deviance from prescribed diabetes self-management regimen (Sharma, 2013). The *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) (American Psychiatric Association, 2013: online), an authority on the classification of mental disorders, does not recognize diabulimia as a legitimate mental condition, but instead offers the following labels under which it might be classified: ‘inappropriate compensatory purging behaviour’, ‘misuse of medications for weight loss’, ‘bulimia nervosa’ and ‘eating disorders not otherwise specified’. The moniker *diabulimia* is therefore not a medically legitimate label but is a portmanteau of the words ‘diabetes’ and ‘bulimia’, invented by people experiencing diabulimia as a means through which  share their experiences of it and seek advice about it online (Goebel-Fabbri et al., 2008).

Due to its contested status, it is not possible for a person to receive a diabulimia diagnosis from a medical practitioner. This notwithstanding, as many as 30 per cent of people with insulin-dependent diabetes are estimated to have intentionally restricted their insulin to control their body weight at some point in their lives (Goebel-Fabbri et al., 2008), with adolescents (Colton et al., 2009)

and women (Shih, 2011) *recognized* to be the most affected groups. Diabulimia can lead to a number of negative health consequences for its sufferers, including diabetic neuropathy, kidney disease, diabetic retinopathy and increased susceptibility to heart attack and stroke (Mathieu, 2008). Research into the long-term effects of diabulimia suggests that life expectancy in those affected by it could be reducible by as much as thirteen years (Shih, 2011).

As a consequence of its contested status, diabulimia is under-researched from both medical and social scientific perspectives (Hughes, 2010). Of those studies which *have* sought to provide insight into diabulimia, the majority has approached the topic from a decidedly positivistic perspective (Thorne, 1997), seeking to provide prevalence figures and to understand its biological consequences and potential *underpinnings*. Although such studies provide an undoubtedly useful resource for health professionals seeking to gain an understanding of what diabulimia is, they reveal little about individuals' subjective experiences and understandings of this emerging health phenomenon (Balfe, 2007). Meanwhile, that limited body of research which has explored lived experiences of diabulimia has based its insights on anecdotal evidence and researcher-invented accounts (Shih, 2011), meaning that, in empirical terms at least, we know very little about the perspectives of people who have lived experience of diabulimia.

The case study reported in this chapter will help to address this knowledge gap by examining the language that people use to disclose and discursively construct their experiences and understandings of diabulimia in the context of three online diabetes support groups, at the same time demonstrating the usefulness of corpus linguistic methods for examining this type of health-related language data. Although language should not be treated as providing a transparent window into the minds and experiences of writers/speakers, for people can 'design' and even filter their linguistic output to suit their objectives, audiences and particular contexts of interaction (Bell, 1984), it seems fair to assume that the linguistic choices that the support group contributors make when talking about diabulimia do at minimum offer representations of their and others' experiences and understandings of it. There is clear value in analysing such representations, not least because they have been produced by contributors for potentially significant purposes in this context, but also because they are consumed, reproduced and challenged by other members of the support groups under study, and so have the potential to shape those other members' own understandings and experiences of this health issue. For the purposes of this analysis, then, there is clear value in utilizing corpus techniques, which will allow me to base my findings on a large collection of support group messages that is more likely than smaller data sets to represent the potentially diverse range of diabulimia-related perspectives shared by the members of these online communities. These advantages, and more besides, will be explored in more detail in the next section.

3 OVERVIEW OF CORPUS LINGUISTIC METHODOLOGY

The data sets on which corpus linguistic analyses are based, corpora, come in two main types: general and specialized. General corpora, designed to represent language use on a broad scale, tend to be very large. For example, the British National Corpus (BNC), which was designed to represent general British English as used during the late 1980s/early 1990s, consists of 100 million words (90 per cent written and 10 per cent spoken texts) (Aston and Burnard, 1998). Specialized corpora, on the other hand, are designed to represent language in more specific contexts and so tend to be smaller than general corpora, though even specialized corpora can still amount to millions of words. For example, in their study of patient feedback about the National Health Service (NHS) in England, Brookes and Baker (2017) examined a corpus of 228,113 online patient comments (29 million words) posted to the NHS Choices website over a period of 2.5 years. Even as a specialized corpus, this data set afforded a more widely representative picture of patients' concerns than would have been possible (or at least practical) had these researchers adopted a purely manual approach to their analysis. As a result, their findings were more generalizable to the patient population overall. Returning to the case study reported in this chapter, corpus methods therefore provide the option of exploring a greater number and wider range of subjective perspectives on the topic of diabulimia than would likely be apprehensible through the analysis of a small number of interviews or illness accounts, for example.

In terms of its application to health(care) language, following an initial interest in written and then spoken discourse, recent corpus studies of health(care) communication have tended to focus on language produced in digital contexts, with research exploring online platforms for advice-seeking and -giving gaining prominence in recent years. For example, across a series of studies, Harvey and colleagues examined variations of a two-million-word corpus of emails sent by young people to a health website aimed at adolescents (Harvey et al., 2007; Harvey, 2012). Another popular area of focus among corpus studies of online health(care) interaction is online, particularly peer-to-peer, support groups, for example Demmen et al.'s (2015) analysis of metaphors for cancer and end of life care in corpora representing support group interactions involving patients, family carers and healthcare professionals (see also Kinloch and Jaworska, this volume). An emerging area of focus for corpus studies of digital health(care) communication is patient feedback, for example Brookes and Baker's aforementioned (2017) study of key concerns in a **twenty-nine**-million-word corpus of online patient feedback about the NHS in England (see also Loew et al., this volume, exploring GPs' responses to an online questionnaire). This strong focus on digital discourse in corpus studies

of health(care) communication can be interpreted as reflecting the growing influence of digital (~~communicative~~) technologies over the ways that people communicate – and indeed act – in relation to their health (Lupton, 2017). On a practical level, for the purposes of a corpus analysis, texts that have been produced within digital contexts ~~are eminently~~ more convenient to collect than spoken and written texts and require minimal processing to be readied for computational corpus analysis. By exploring the discourse surrounding diabulimia in online support groups, the case study reported in this chapter can thus be viewed as harnessing the advantages of collecting and analysing readily digitized language data, all the while contributing to the growing body of corpus-based online health communication research.

Once a corpus has been compiled, or an existing corpus selected, it is then analysed with the help of specialist computer programs, such as *WordSmith Tools* or *AntConc* (see Semino et al., this volume, for examples of other tools). Such packages, which are often freely available or purchasable for a relatively modest license fee, allow the researcher to search for every occurrence of any word or combination of words, generate frequency information about linguistic phenomena of interest (e.g. words, chains of words, grammatical types), perform statistical tests on those frequencies (i.e. to measure the significance or strength of relationships between phenomena) and present the data in ways that render it more amenable to manual analysis (Baker, 2006), meaning that mixed methods of analysis can be used. Indeed, corpus studies of health(care) language tend towards interdisciplinarity – incorporating perspectives from fields as diverse as medicine, psychology and sociology into their analyses (Brookes and Harvey, 2016). As well as reflecting the interdisciplinary nature of health(care) language approaches generally, this methodological and theoretical diversity also mirrors the diversity of corpus linguistics approaches, whose flexibility provides fertile ground for such multiplicity of methods and theoretical approaches (Hardie and McEnery, 2010). Corpus studies of health(care) communication are thus able to provide both quantitative *and* qualitative perspectives on data that represent a wider range of communicative routines, lived experiences and understandings of health(care) and illness than would likely be feasible using purely qualitative approaches, producing – by and large – more generalizable findings in the process (Crawford et al., 2014).

Computer assistance can also be beneficial for the analysis itself, as it can reveal patterns that run counter to human intuition or which feature sparingly in one or two texts but become significant when considered as part of a larger collection of ~~texts~~ (Baker, 2006). ~~An interrelated~~ advantage of corpus methods is their ability to help analysts guide their analytical focus using more objective criteria, such as frequency information and statistical significance, thus reducing (though not eliminating entirely) the influence of human bias on their analyses. This increased objectivity is also supported by corpus linguistics'

commitment to methodological transparency, which is underpinned by two guiding principles: (i) no systematic bias in the selection of texts included in the corpus (i.e. do not exclude a text because it does not fit a pre-existing argument or theory) and (ii) total accountability (all data gathered must be accounted for) (McEnery and Hardie, 2012). Combined, these principles can help analysts to overcome the accusation, often directed at more qualitative approaches to discourse analysis, that analyses are based on convenient texts or examples, cherry-picked by the researcher because they support a preconceived argument or theory (Widdowson, 2004). In this way, by affording the opportunity to examine large quantities of authentic language data, corpus linguistic methods might also be said to go some way towards appeasing the commitment to more objective approaches to large data sets that is commonplace in the domain of empirical health research (Brown et al., 2006), thus having the potential to help researchers to bridge the gap between social scientific and biomedical perspectives on health and illness.

4 DATA AND METHOD

The case study presented in this chapter is based on a specialized, purpose-built corpus of messages posted to three English-speaking diabetes support groups. Support groups were sourced through a search engine query using the search-terms ‘diabetes support group’, ‘diabetes forum’ and ‘diabetes message board’. Of the top 100 search results, only three support groups met the following criteria for inclusion in the corpus:

- (i) English-speaking;
- (ii) Dedicated to diabetes;
- (iii) Hosts peer-to-peer, user-generated content (as opposed to practitioner-directed or practitioner-led content);
- (iv) Not affiliated to a healthcare provider or charity (such sites are typically monitored by practitioners and other specialists);
- (v) Meets ethical criteria (does not require registration to view content, explicitly informs users of the public nature of their contributions and does not explicitly discourage, or state requirement of permission for, the use of content for research purposes).

From the three qualifying support groups, individual threads (chronologically ordered chains of messages) were included in the corpus if they contained mentions of the words ‘diabulimia’ and/or ‘diabulimic’ once in the thread title and/or three or more times across the messages contained within. Qualifying threads were included in the corpus in their entirety. The completed corpus is 119,982 words in size, comprising 81 threads and 1,072 messages posted

between 2007 and 2014. The corpus contains messages from a mixture of contributors who ostensibly do and do not present themselves as having first-hand experience of diabulimia, with the latter group consisting mainly of relatives and advice-givers. With the data available, I have little choice but to accept all the messages at face-value. However, I am conscious of the possibility for contributors to falsely present either as having or not having diabulimia (in the former case, for example, to seek advice on behalf of another (Harvey, 2012)). With these caveats in mind, the corpus is most accurately described as representing the disclosure of experiences and understandings of diabulimia by contributors to three online, English-speaking, diabetes support groups.

Before progressing further, it is perhaps useful to briefly outline some of the ethical considerations that arise when using online data such as the support groups featured in my corpus. Debates concerning the ethics of online data collection began around the turn of the twenty-first century and continue to endure to this day. One area of contestation relates to whether online interactions should be regarded as belonging to the public or private domain (Elgesem, 2002). To this end, I wish to distinguish between those online spaces that can be accessed only by registered members who have ‘logged in’ and those that do not require a login but can be accessed publicly, by anyone. While we might consider websites or support groups which require users to register and log in to view content as constituting private domains of interaction (accessible to the members of those online communities only), those online spaces which do not require people to register and log in to view content might be considered to be more public than private (see also Coulson et al., 2007; Demmen et al., 2015). Thus, as per point (v) of the support group inclusion criteria outlined above, I included in my corpus only those support groups which I regarded as ‘public’ in as much as they did *not* require registration to view content, *did* explicitly inform users of the public nature of their contributions and did *not* explicitly discourage, or state requirement of permission for, the use of the content for research purposes (see also Eysenbach and Till, 2001). Due to the public nature of the messages, I did not seek informed consent from the support group contributors but have removed all mentions of names and any other identifying information (e.g. contact details, locations, etc.) in order to preserve the contributors’ anonymity. Ethical approval was obtained prior to data collection.

As well as being suited to investigating a health phenomenon that was effectively born in cyberspace (Sharma, 2013), online support groups offer two significant methodological advantages for this case study. First, the support group threads were readily available in a digital format, which meant that they required minimal processing and could be imported into the corpus with relative ease. Second, as a relatively anonymized platform, online support groups afford their users opportunities for heightened levels of candour compared to

interactions taking place in offline contexts – what Suler (2004: 322) refers to as the ‘online disinhibition effect’. However, a limitation of sourcing data from anonymizing platforms is that it is not possible to access demographic information about the contributors, which meant that it was not possible to assess the demographic balance of the people who posted the messages included in my corpus. Consequently, my findings cannot be attributed to any specific group. It should also be noted that the contributors to the support groups in my corpus constitute a self-selecting sample, which means that the diabulimia-related perspectives they provide might best represent the perspectives of people who have access to the internet and who choose to disclose their illness experiences online. It is important to acknowledge that there might be important differences between this group’s experiences and understandings of diabulimia and those of people who, for one reason or another, do not access or contribute to such groups.

5 USING CORPUS TECHNIQUES TO ANALYSE ONLINE SUPPORT GROUP MESSAGES

The analysis in this case study was carried out in an exploratory fashion – with the aim of first identifying key themes in the data, before examining how one of those key themes (specifically, insulin) is linguistically constructed across the support group messages. There is no standard approach or set of procedures in corpus linguistic methodology. However, the forthcoming analysis will make use of three well-established techniques in corpus linguistics, namely keywords, collocation and concordance, all of which were accessed using *WordSmith Tools* Version 7 (Scott, 2016).¹ The forthcoming case study will be reported in a way that serves to demonstrate how these techniques can be combined to at first achieve a quantitative overview of key themes in the data before exploring the linguistic construction of those themes in a more qualitative fashion.

5.1 Keywords

My examination of the support group messages begins with keywords. Keywords are words which occur with a significantly higher frequency in the corpus being analysed when it is compared against another corpus (Scott, 1999). To generate keywords, we first have to generate a frequency list for all the words occurring in the corpus under study (obtained through the ‘word list’ function of *WordSmith* and most other corpus software packages). We then compare the word list for the corpus we are analysing – in this case, my corpus of diabulimia support group messages – against the word list for a comparable reference corpus, which typically represents a norm or ‘benchmark’ for the type of language under investigation. Words are deemed to be keywords based

on statistical comparisons of the word frequency information for each corpus, with keywords representing those words that occur considerably more often in the analysis corpus compared to the reference corpus. The computational method of generating keywords is therefore not based on concepts that are subjectively viewed as important to culture (see Williams, 1983), but allows for potentially any word to be a keyword, provided that it occurs frequently enough in the corpus under analysis compared to in the reference corpus.

When selecting a reference corpus, we usually want one that is similar in size to, or larger than, the corpus being analysed. Ideally, the reference corpus should also represent language belonging to the same genre as the texts in the analysis corpus, so that the keywords will flag up what is *lexically* distinctive about the texts in the corpus compared to others of a similar type. To generate a set of keywords for this case study, I compared the word frequency list for my corpus of diabulimia messages against the word frequency list for the general language BNC (introduced earlier; downloaded from *lexically.net*). The BNC is a suitable reference corpus in this regard, as its combination of spoken and written registers can be said to represent a standard or benchmark for the part-written/part-spoken character of computer-mediated communication (Baron, 1998).² The resulting keywords were therefore those words which occurred with an unusually high frequency in my corpus compared to the BNC (in other words, compared to general ~~written and spoken~~ British English).

Another important decision we have to make when generating keywords regards which statistical test to use. Most corpus analytical software packages offer a number of statistical tests which measure either effect size or statistical significance. Effect size metrics indicate the strength of the difference or relationship we have found (i.e. higher scores indicating stronger differences/relationships), while measures of statistical significance indicate the level of confidence the researcher can have that the difference or relationship observed is dependable and not the result of a sampling error (Gabrielatos, 2018). In other words, effect size indicates strength, while statistical significance indicates confidence. It is recommended that researchers combine metrics of effect size and statistical significance to make their results more robust. Accordingly, for this case study I generated keywords using Log Ratio (McEnery and Baker, 2016: 23), which combines a test of statistical significance (log-likelihood, Dunning, 1993) with a measure of effect size, which quantifies the strength of the difference between the observed frequencies, independent of the sample size. This statistic will produce as keywords those words whose relative (or 'normalized') frequency is significantly higher (beyond the 99.9 per cent threshold level) in my corpus compared to the reference corpus (BNC). Each statistically significant keyword is then assigned a Log Ratio score based on the size of the observed difference in relative frequencies, with bigger differences producing higher scores. This measure therefore uses log-likelihood as a cut-off

to ensure that keywords are significant but has the advantage that it also allows us to rank those keywords according to how unusually high, or ‘marked’, their frequency is in the corpus being analysed. Since Log Ratio privileges low-frequency keywords, I also stipulated that keywords should occur in a minimum of 5 per cent of the texts in the corpus ($n = 53$). The top 20 keywords (an arbitrary cut-off) ranked by Log Ratio are displayed in Table 2.1.

One way of approaching these keywords is to group them into semantic categories reflecting the most characteristic themes in the corpus. Thus, we might interpret the above keywords as indicating themes such as diabulimia (‘diabulimia’, ‘disorder’), insulin (‘lantus’ (a commercial brand of insulin),

TABLE 2.1 Top 20 Keywords, Ranked by Log Ratio

| Rank | Keyword | <i>Diabulimia</i> <i>corpus freq.</i> | | Texts | <i>BNC Freq.</i> | | Log Ratio |
|------|---------------|--|------|-------|------------------|-------|-----------|
| | | N | % | | N | % | |
| 1 | diabulimia | 200 | 0.17 | 146 | 0 | <0.01 | 139.77 |
| 2 | Lantus | 84 | 0.07 | 53 | 0 | <0.01 | 138.52 |
| 3 | Dka | 72 | 0.06 | 61 | 0 | <0.01 | 138.30 |
| 4 | carbs | 112 | 0.09 | 74 | 22 | <0.01 | 12.04 |
| 5 | Carb | 86 | 0.07 | 69 | 52 | <0.01 | 10.42 |
| 6 | insulin | 791 | 0.66 | 372 | 631 | <0.01 | 10.02 |
| 7 | diabetes | 328 | 0.27 | 208 | 652 | <0.01 | 8.70 |
| 8 | sugars | 68 | 0.06 | 57 | 152 | <0.01 | 8.53 |
| 9 | diabetics | 63 | 0.05 | 53 | 233 | <0.01 | 7.81 |
| 10 | diabetic | 117 | 0.10 | 95 | 623 | <0.01 | 7.28 |
| 11 | hi | 104 | 0.09 | 101 | 1011 | <0.01 | 6.41 |
| 12 | complications | 83 | 0.07 | 64 | 838 | <0.01 | 6.36 |
| 13 | diagnosed | 72 | 0.06 | 59 | 730 | <0.01 | 6.35 |
| 14 | disorder | 140 | 0.12 | 102 | 1617 | <0.01 | 6.17 |
| 15 | eating | 316 | 0.26 | 203 | 4169 | <0.01 | 5.97 |
| 16 | weight | 559 | 0.47 | 297 | 8400 | <0.01 | 5.79 |
| 17 | pump | 73 | 0.06 | 59 | 1248 | <0.01 | 5.60 |
| 18 | healthy | 95 | 0.08 | 74 | 3529 | <0.01 | 4.48 |
| 19 | eat | 187 | 0.16 | 127 | 7280 | <0.01 | 4.41 |
| 20 | lose | 161 | 0.13 | 117 | 6302 | <0.01 | 4.40 |

‘insulin’, ‘pump’), diabetes (‘dka’ (diabetic ketoacidosis), ‘diabetes’, ‘sugars’, ‘diabetics’, ‘diabetic’, ‘complications’, ‘diagnosed’), and weight and eating (‘carbs’, ‘carb’, ‘eating’, ‘weight’, ‘healthy’, ‘eat’, ‘lose’), while the keyword ‘hi’ reflects the interactional dynamic of the genre. The keywords measure thus provides a rapid and replicable overview of key themes in the corpus, as these are reflected in its most characteristic words. Compared to manual thematic analysis, computational measures of keyness are closer to accepted methods in medical research because they are not only statistically reliable (keywords are statistically significant and not just intuitively interesting) but also replicable (the keywords procedure can be re-run in future analyses to verify or test hypotheses about the data). However, inspecting words in isolation, as the keyword list forces us to do, can be somewhat limiting, as it provides a relatively decontextualized view of the language in the corpus. Thus, although the keywords reveal key themes in the diabulimia support group messages in my corpus, this list alone cannot tell me why these themes are important or reveal how they contribute to the linguistic construction of diabulimia in these contexts. To gain such insight, it is necessary to examine the keywords within their more expansive contexts of use. We can begin to do this using the collocation technique.

5.2 Collocation

Collocation, as conceptualized in corpus linguistics research, refers to the association between words based on patterns of co-occurrence. Collocation is typically judged to exist using a word association measure that tells us how often two or more words occur within close proximity of one another in the corpus, and whether this association is notable as a sizeable effect (i.e. that the words in question have a measurably strong preference to occur together as opposed to being randomly associated). Following Firth’s (1957: 6) dictum that ‘you shall know a word by the company it keeps’, corpus linguists have long sought to learn about words’ meanings and patterns of use by examining them in terms of the words with which they tend to co-occur, or ‘collocate’. Analysing the collocates of a keyword or other word of interest can therefore provide insight into the textual context that surrounds that word in the corpus. For reasons of space, I will focus here on the collocates surrounding just one keyword, ‘insulin’. The keyword ‘insulin’ is not just a computational keyword but takes on a broader, cultural significance with respect to diabulimia, given that the deliberate restriction of insulin is diabulimia’s most essential practice. Note, however, that a full analysis of this corpus would explore more of the keywords in greater detail, perhaps according to the thematic groupings offered in the previous section.

As with the generation of keywords, generating collocates requires the researcher to make a series of procedural decisions, for instance pertaining to

the span of collocation, the choice of statistical measure and possible use of a minimum frequency threshold, all of which will ultimately shape the amount and type of words that are flagged as collocates by the computer. First, we have to decide on the collocational span – that is, the number of words to the left and/or right of the search-word within which we want to search for candidate collocates. Tighter spans will produce a smaller number of collocates which occur within closer proximity to the search-word, while wider spans will produce a higher number of collocates, some of which won't occur in such close proximity to the search-word. For this case study, I searched for collocates of 'insulin' using a window of five words to the left and right of the search-word (otherwise expressed as $L5 > R5$). This is a fairly standard window in corpus linguistic research, as it is judged to provide a 'good balance between identifying words that actually do have a relationship with each other (longer spans can throw up unrelated cases) and [gives] enough words to analyse (shorter spans result in fewer collocates)' (Baker et al., 2013: 36). Furthermore, I imposed a minimum frequency threshold of five (default in *WordSmith Tools*), which meant that a word had to co-occur within the aforementioned span of five words to the left and right of 'insulin' at least five times to be a candidate collocate. These parameters can be adjusted to focus on words which collocate more or less closely by changing the span and more or less often by changing the minimum frequency threshold.

The next decision pertains to how candidate collocates are ranked, scored or 'cut-off'. Collocates can be ranked either according to frequency of co-occurrence or by using a statistical measure. Popular measures in corpus research include Mutual Information (MI), MI^3 , Z-score, log-likelihood, log-log and Log Ratio (see Gablasova et al., 2017 for a review). Again, the choice of statistic will shape the collocates produced. For this case study, I derived collocates using the cubed version of the MI statistic (MI^3). MI determines collocation strength by comparing the observed frequency of each collocational pairing against what would be 'expected' based on the relative frequency of each word and the overall size of the corpus. The difference between the observed and expected frequency of co-occurrence is then converted into a score (MI score) which indicates the strength of collocation, with higher scores assigned to stronger collocational pairings. I selected the cubed version of MI for the present study as, in contrast to traditional MI which tends to emphasize exclusive and unusual word combinations (Evert, 2008), MI^3 favours collocates which have a higher frequency of co-occurrence and so tend to be 'more established in the discourse' (Brezina et al., 2015: 160). This means that MI^3 not only is useful for corpus-assisted discourse analysis but also offers the practical advantage that it helps to control for the presence of typos and non-standard spellings (both of which are common in computer-mediated communication) by pushing them down the list in favour of more frequent collocates. Table 2.2 shows the top 20 collocates of 'insulin', ranked by MI^3 score.

As Table 2.2 shows, MI³ has given a mixture of lexical and grammatical collocates for ‘insulin’. Focussing on the lexical collocates, note the prominence of words which denote the act of taking (or not taking) insulin, including ‘taking’, ‘take’, ‘skipping’ and ‘less’, as well as ‘not’ which, as a collocate of ‘insulin’, tends to precede words like ‘taking’ to form strings such as ‘not taking insulin’. In fact, most of these words focus on the act of not taking – or restricting – insulin, confirming this as the central feature of diabulimia. Inspecting examples containing the words ‘taking’ and ‘take’, we see that these words tend to feature in formulations denoting insulin restriction – for example, as part of formulations like ‘not taking insulin’ ($n = 33$), ‘stopped taking insulin’ ($n = 9$), ‘stop taking insulin’ ($n = 5$), ‘not take insulin’ ($n = 5$) and ‘take less insulin’ ($n = 4$). The other notable pattern in this table is the

TABLE 2.2 Top 20 Collocates of ‘Insulin’ (5L > 5R), Ranked by MI³

| <i>Rank</i> | <i>Collocate</i> | <i>Collocation freq.</i> | <i>Texts</i> | <i>MI³</i> |
|-------------|------------------|--------------------------|--------------|-----------------------|
| 1 | to | 276 | 173 | 19.66 |
| 2 | taking | 93 | 78 | 19.43 |
| 3 | and | 196 | 146 | 18.37 |
| 4 | the | 189 | 124 | 18.27 |
| 5 | weight | 101 | 81 | 18.09 |
| 6 | I | 184 | 120 | 17.61 |
| 7 | of | 134 | 104 | 17.60 |
| 8 | on | 97 | 78 | 17.31 |
| 9 | my | 110 | 75 | 17.14 |
| 10 | take | 59 | 51 | 17.00 |
| 11 | your | 87 | 63 | 16.89 |
| 12 | not | 87 | 71 | 16.79 |
| 13 | that | 102 | 78 | 16.45 |
| 14 | is | 98 | 78 | 16.44 |
| 15 | a | 111 | 86 | 16.33 |
| 16 | for | 82 | 67 | 16.06 |
| 17 | skipping | 23 | 17 | 15.65 |
| 18 | less | 30 | 24 | 15.56 |
| 19 | lose | 37 | 36 | 15.54 |
| 20 | with | 66 | 63 | 15.52 |

collocation of ‘insulin’ with words relating to weight (‘weight’ and ‘lose’), which attests the weight loss motivation that is widely perceived, including by healthcare professionals, to cause people to deliberately restrict their insulin in the first place (Sharma, 2013).

In taking us beyond the solitary linguistic ~~items keyword~~ list, the collocation technique provides a logical ‘next step’ following ~~these~~ inductive procedures. A full analysis of this corpus would consider the relationships between the word ‘insulin’ and all the collocates in Table 2.2 (and possibly more besides). For this case study, however, I am going to focus on the most prominent theme among the lexical collocates; specifically, linguistic constructions of the act of (not) taking insulin. To explore how this theme featured in the contributors’ messages and contributed towards the broader construction of their understandings and experiences of diabulimia, it was necessary to go beyond the relatively narrow, ~~still decontextualized~~ view of the data afforded by collocation, and explore this theme through the prism of concordance.

5.3 Concordance

The concordance technique provides a way of viewing the corpus data that allows us to examine every occurrence of a word or phrase in context, and thus to quickly scan for patterns of use. An example of what a concordance output looks like is given in Figure 2.1, based on the search term and keyword ‘insulin’.

| Context (left) | Search-word | Context (right) |
|--|-------------|--|
| one but I realized that it was enabling me to skip | insulin | since it was pumping just enough to keep |
| to 30% of teenage diabetic girls intentionally skip | insulin | to lose weight. I had never really made the |
| take the great advice given here and do not skip | insulin | to lose weight. There are safer and smarter |
| health at risk or doing the right way. I wouldn’t skip | insulin | to lose weight. If all else failed I would go to |
| Wow! I would never skip | insulin | for the sake of losing weight. Very bad, |
| to add that I am not a “diabulimic”, and don’t skip | insulin | injections, no judgement passed; my |
| have type 1 diabetes 22 years. I used to skip | insulin | and eat what I wanted. I lost 2 stone and felt |
| i have never skipped | insulin | Shots because I when I was younger there |
| , admitted that they skip or have skipped | insulin | for weight management. |
| to find out I am not the only one. I have skipped | insulin | recently, but before this I used to purge/ |
| Hello and welcome to DD! Skipping | insulin | shots is bad. You can very easily get sick |
| for 26 years and figured out about skipping | insulin | and losing weight at the age of 14. I did it on |
| eating disorder but if you are doing it by skipping | insulin | I can say that my pump’s basal rate was |
| I can’t add much more but skipping | insulin | now will catch up with you one day and it |
| bringing –throwing up –tablets she takes – skipping | Insulin | - ... How can I help her, without scaring her |
| I’m told it’s people with T1D skipping | insulin | injections & using less insulin than they |
| to deter eating too much fat.. I googled ‘skipping | insulin | ’ to research and found your post. However, |
| & difficult to shed unwanted pounds by skipping | insulin | & letting excessively high BG’s shed the |
| diabetes. Instead, she focused on this: Skipping | insulin | Equals weight loss. For the next 17 years, |
| , but it’s better than it was. I’ll also say skipping | insulin | was far from the only part of my eating |
| discussions wrote that once you begin skipping | insulin | , it is really difficult to start taking insulin |
| enough to lose weight, so I found that skipping | insulin | Injections would help lose A LOT of weight |
| of ‘diabulimia’ emergeThe practice of skipping | insulin | shots to create weight loss hasn’t been |
| the situation 10-fold by intentionally skipping | insulin | ? The long-term complications are too scary |

FIGURE 2.1: Sample concordance of ‘insulin’.

With the search-word running down the centre, and a few words of context displayed to the left and right, the concordance output can be very useful for spotting patterns that might be less obvious during more linear, left-to-right and top-to-bottom readings of the data. The concordance lines can be displayed in order of occurrence, in random order, or alphabetically according to the words surrounding the search-word (rendering recurrent patterns more visible). The concordance lines displayed in Figure 2.1 have been sorted alphabetically according to the word directly preceding the search-word (i.e. in the L1 position). For a more contextualized view of the data, the analyst can access each original text (in this case, support group message) in its entirety, simply by clicking the highlighted search-word in the centre of each concordance line. Concordancing thus provides the means to adopt a different perspective on the language in the corpus – one which allows the human analyst to undertake closer, more qualitative analysis of words or phrases in the corpus and develop theory-informed observations based on extended patterns of use.

In this case study, I used the concordance technique to carry out more qualitative analysis of the linguistic construction of insulin-restricting practices in the support group data. Although the collocation analysis was useful for first directing my attention to this theme, my analysis does not have to be restricted to these collocates alone. As such, to obtain a more comprehensive picture of the patterns surrounding insulin restriction in my corpus, I searched the list of collocates (L5>R5) of the word ‘insulin’ to look for other terms that could be used to denote the practice of reducing or not taking insulin (collocating with ‘insulin’ at least twice). In addition to the collocates ‘taking’, ‘take’, ‘skipping’ and ‘less’ highlighted in the previous section, this also brought to my attention a number of other collocates which could be useful for exploring constructions of insulin restriction, including (collocation frequencies in brackets) ‘skip’ (18), ‘stop’ (17), ‘reduce’ (15), ‘stopped’ (13), ‘reducing’ (11), ‘cut’ (9), ‘omitting’ (8), ‘manipulating’ (6), ‘skipped’ (6), ‘cutting’ (4), ‘decrease’ (4), ‘lower’ (4), ‘abusing’ (3), ‘adjust’ (3), ‘drop’ (3), ‘manipulation’ (3), ‘omission’ (3), ‘quit’ (3), ‘restrict’ (3), ‘stopping’ (3), ‘underdosing’ (3), ‘withhold’ (3), ‘abuse’ (2), ‘avoid’ (2), ‘limit’ (2), ‘limited’ (2), ‘lowering’ (2), ‘minimizing’ (2), ‘reduction’ (2), ‘restricting’ (2), ‘restriction’ (2), ‘short’ (2) and ‘withholding’ (2). In the ensuing analysis, I use the concordance technique to examine the discourses surrounding insulin restriction in a random sample of 100 support group messages in which ‘insulin’ co-occurred with one or more of the collocates listed above. Although concordancing provided access to these messages, the analysis is not restricted to solitary concordance lines, but is usually based on entire messages, and in some cases threads of messages. For each type of construction discussed, I indicate the percentage of the messages in the sample in which that construction was observed. However, it should be noted that messages could contain more than one type of construction, as I interpreted it. Excerpts are

provided to demonstrate analytical points and were selected because they were deemed to be representatives of the particular pattern being discussed.

The first pattern I want to consider here pertains to the construction of the practice of taking insulin as causing weight gain, which I observed in 62 per cent of the messages in the sample. For example, the contributor of the message in Excerpt 1 describes discovering that insulin intake equates to weight gain ('insulin = weight gain').³

Excerpt 1

I was diagnosed with type 1 diabetes in 2004,,I educated myself very quickly on it through books and internet, thats when i discovered that insulin = weight gain...

Although this was a dominant discourse in my sample, claims that taking insulin induces weight gain could be challenged by other members of the support groups. In fact, this link was problematised in 18 per cent of the messages in my sample, as in Excerpt 2.

Excerpt 2

I think there's the potential for diabetics to gain weight easier than non-diabetics. But that doesn't mean you will. And I don't think being diabetic and taking insulin makes it harder to lose weight. It all depends on your control and lifestyle and how you use the insulin... because you could very easily eat what you want, shoot up for it and then gain 10 pounds.

Although the contributor of this message concedes that there is the 'potential for diabetics to gain weight easier than non-diabetics', they also problematize the link between insulin and weight gain, suggesting that it depends on control of diabetes more generally, as well as wider lifestyle factors. However, it should be borne in mind that this was a minority position in the messages, respective to the comparatively unproblematic linking of insulin intake to weight gain.

The next construction I want to consider is that of insulin restriction as helping to induce weight loss, which could be observed in 49 per cent of the messages in the sample and manifested in contributors' descriptions of their own or others' insulin-restricting practices, or in more general discussions about diabulimia. Given that such messages recall diabulimia's most essential practice, their prominence in my sample is hardly surprising. Perhaps a more interesting pattern was the tendency for the contributors to describe (ostensibly their own) insulin restriction in conceptually vague and distancing terms. Even in cases where a lexical verb appeared to explicitly denote the act of insulin restriction, the precise details regarding *how much* insulin was restricted were rarely given –

specifically in just 2 per cent of cases. This vagueness is exemplified by the message reproduced in Excerpt 3.

Excerpt 3

8 years ago, before hearing the term “diabulimia” I decided to reduce my insulin intake as well in order to lose some weight. I maintained my basal injections, but hardly ever gave a bolus (regardless of correction bolus or food bolus). I did this for probably 4-5 months and ended up losing 40 pounds. It got to a point though where I started getting so thirsty that I would consistently eat more as well, and my portion sizes became outrageous. At 125 pounds, I could sit down and eat an entire large pizza and not think twice. Kind of crazy looking back.

The vagueness with which the practice of insulin restriction tended to be constructed in the corpus is a curious trend given the high level of specialist diabetes knowledge that is often exhibited by people with diabetes when communicating about their condition online (Fox et al., 2005). Indeed, this vagueness is rendered striking when compared with the tendency for the contributors to provide quite precise details pertaining to other aspects of their insulin restriction. For example, this contributor gives precise details regarding their body weight (‘at 125 pounds’), how much weight they lost (‘ended up losing 40 pounds’) and even when and for how long their insulin-restricting practices took place (‘8 years ago’, ‘I did this for probably 4-5 months’). The precision with which this contributor and others constructed the bodily and temporal dimensions of their insulin restriction is consistent with existing research into eating disorder disclosure, which reports the tendency for individuals to exhibit acute awareness and knowledge of information such as calorie consumption, body mass information and so forth, in the ways in which they talk about their condition (Day and Keys, 2008).

One interpretation of the conceptual vagueness concerning the amounts of insulin restricted is that it might constitute a strategy by which contributors can downgrade the perceived seriousness of their actions (Adolphs et al., 2004) by shifting focus *away* from the amounts of insulin being restricted. This interpretation is further supported by the tendency for most of the contributors to linguistically distance themselves from their insulin restriction. For example, returning again to Excerpt 3, this contributor temporally distances their self from their insulin-restricting actions by situating these firmly in the past through the expression ‘kind of crazy looking back’; a negative evaluative statement which also casts the contributor in the less stigmatizing role of rational and disapproving observer in relation to their previous actions (Galasiński, 2008). This contributor also distances their self from the label diabulimia itself, making it clear that their insulin restriction predates their learning of the label when

they write: ‘before hearing the term “diabulimia” I decided to reduce my insulin intake as well in order to lose some weight.’ This might be an attempt by this contributor to sidestep the accusation that they are engaging in something faddish or ‘jumping on the bandwagon’, as is often the perception of diabulimia among people who have limited knowledge or experience of it (Sharma, 2013).

Vague and distancing constructions of insulin-restricting practices might therefore be interpreted as a means for contributors to sidestep potentially negative and critical responses from fellow contributors (Armstrong, Koteyko and Powell, 2011). Indeed, there is ample evidence of such responses in the corpus: more than any other type of response, disclosures of insulin-restricting practices elicit warnings from other support group members about the dangerous, risky and even irresponsible nature of the practices attested (68 per cent; Excerpt 4).

Excerpt 4

As others have said omitting insulin to loose weight is highly dangerous and is really not worth the risk.

As well as being perceived as dangerous and risky, insulin restriction could also be constructed as a form of deviance. Following Clinard and Meier (2011: 6), I interpret deviance to be ‘a collection of conditions, persons, or acts that society disvalues ..., finds offensive ... , or condemns’. As part of the self-management of their condition, people with diabetes are required to exercise a high degree of control over their blood sugar levels and calorie intake (Peel et al., 2005). In 28 per cent of the messages in the sample, insulin restriction was constructed as a form of deviance from the responsibilities associated with diabetes self-management. The most explicit manifestation of this discourse could be found in cases where insulin restriction was lexicalized through verbs which, I argue, have an inherently negative evaluative prosody, such as ‘abus(ing)’ and ‘misus(ing)’, as in Excerpts 5 and 6.

Excerpt 5

I misuse my insulin for years, had poor control and never tested. It can get you into a cycle called diabulimia, without realizing it. Not saying that is what you are do or have, but that is what I did. For me, it last from my teens into my thirties. Now am 35, and for the first time have been healthy for the last 4 years, with my diabetes in control, since my diagnosis age 8.

Excerpt 6

My daughter was diagnosed with diabetes at 12 (not a great age with hormones kicking in :() and developed “disordered eating” pretty soon

after. She is nearly 18 and has been abusing insulin ever since. How she has managed to survive until now is beyond belief and is currently taking basal insulin to try and retain some stability. However, she is still abusing and has now confided she is taking it 3 days a week Mon, Tue, Wed, and then stopping. She is binge eating thur, fri, sat, sun and running deliberately with high ketones to get rid of the weight she has put on in those 4 days. She can lose anything up to a stone in a week. :(She has been staying with her father who is in complete denial there is a problem, but we have spent the weekend in tears where I have managed to get her to admit exactly the extent of what she is doing so will obviously be acting immediately. I am currently researching as much as I can to find the support she needs.

The contributor of Excerpt 5 groups together in a single sentence their insulin misuse and purportedly poor diabetes self-management when they write, 'I misuse my insulin for years, had poor control and never tested', and goes on to equate recovery and being 'healthy' with having diabetes 'in control'. Similarly, the contributor of Excerpt 6 expresses their surprise that their daughter, who 'abuses' her insulin, is still alive, such is the severity and longevity of her diabetes mismanagement. This contributor also notes the deliberate nature of their daughter's insulin abuse and draws on a discourse of secrecy (Rhodes et al., 2010), evidenced through lexical choices such as 'confided' and 'admit', to construct their daughter's actions as not only abusive but also secretive and perhaps even shameful. This can be interpreted as reflecting the broader penchant for diabetes control to be framed as a moral duty (Balfe, 2007: 146), failure at which can result in one being negatively evaluated as 'bad' or 'deviant' (Gomersall et al., 2011: 13). However, it is also important to bear in mind that this negative evaluative discourse was not the preserve of those who claimed to manage their diabetes 'properly' in accordance with practitioner advice but was so pervasive that it was also drawn upon by contributors to negatively evaluate *themselves* for restricting their insulin.

Insulin-restricting practices were not just constructed as a form of deviance but could be related to the responsibilities of diabetes self-management in other ways, too. In 15 per cent of the messages in the sample, insulin restriction was actually represented as a normal way of effectively and autonomously managing diabetes, on the seeming proviso that it occurs in conjunction with the appropriate reduction of calorie intake (i.e. not restricting insulin dosage to compensate for excessive calorie intake, or 'bingeing', Littlefield et al., 1992), for example in Excerpt 7.

Excerpt 7

As long as he is taking enough insulin to cover what he eats then there is really nothing wrong with "abusing" your insulin to "eat things you

shouldn't". Personally, I don't have any forbidden food or drink list (I drink regular soda) and I keep my control perfect all the time. Your concern should be that he does take enough insulin when he eats higher carb/sugar foods. It will not harm him at all if he takes the right amount of insulin to cover it. I think your husband probably knows what he's doing, there is nothing wrong with what he is doing.

The insulin-restricting practices attested in this message can be interpreted as autonomous in as much as this contributor recommends balancing insulin dosage with calorie intake. However, it should be noted that constructions of insulin restriction as a potentially normal, even medically acceptable part of diabetes self-management constituted a minority discourse in this context, for, as we saw earlier, the majority of the contributors tended to talk about diabulimia in ways that suggested they considered it to be a deviant and even dangerous practice. Nevertheless, there does appear to be potential overlap in the ways that at least some people with diabulimia (and others) conceive of pathologized and therapeutically intended regimes of diabetes self-management. A similar conflation of disordered and normative bodily practices has been observed with respect to the discourse surrounding eating disorders, for instance by Malson (2008: 35–6), who argues that the ideals of weight monitoring and maintenance are 'enacted par excellence' by anorexia nervosa, which is characterized by a hyper-disciplined micro-management of the body and bodyweight. Such overlap between so-called normalized and pathologized bodily practices has also been observed in relation to diabetes by Paterson et al. (1998), who argue that accomplishing or at least performing autonomy and expertise, as per the requirements of neoliberal frameworks of diabetes self-management, can involve active experimentation with diet, activity and medication. However, it must be noted that although these messages serve to normalize insulin restriction, the inclusion of the condition that insulin restriction should take place in conjunction with measures such as closely monitoring calorie intake implicitly marks insulin restriction that does *not* meet such conditions as deviant or even pathological. Indeed, as Malson (2008) observes in relation to weight loss practices, it is through the very labelling of certain practices as 'normal' that others come to be thought of as pathological.

Finally, rather than signalling either the management or mismanagement of diabetes, in 7 per cent of the messages in the sample, insulin restriction was constructed as something of a means for relinquishing the feelings of powerlessness, obedience, restraint and self-denial that are often felt by people who are required to live with and manage chronic illnesses, including diabetes (Paterson, 2001).

Excerpt 8

Anyway, since I am taking a bunch of premed classes now, I not only get to hear about all of the terrible things that happen with chronic high sugars, but I also learn a lot about cell bio and can use this info to further manipulate my insulin. All of this started out as just wanting to drop a few pounds, but now that I have tasted life without diabetes, I just don't want to go back. When I was diagnosed, I just accepted that this was the way it was going to be. I didn't ever consider just not doing it. It feels so good to not obsess over every carb that I put in my body, to not worry about testing, to not have to calculate corrections or meal boluses.

Such is the promise of insulin restriction for overcoming feelings of powerlessness that this contributor equates it to relinquishing diabetes itself, when they write, 'now that I have tasted life without diabetes, I just don't want to go back.' Yet, this contributor also seems keen to stress to other members of the support group that their actions are not careless or neglectful. They present their insulin restriction as deliberate and even calculated, lexicalizing it as insulin 'manipulati[on]', and qualifying their actions by foregrounding their knowledge of biomedical concepts like 'cell bio', acquired through 'premed classes'. This presentation of the self as knowledgeable and empowered in relation to insulin restriction can thus be sharply contrasted against the relative powerlessness otherwise attested by this contributor and others in relation to the requirements of them to self-manage their diabetes.

6 IMPLICATIONS OF FINDINGS FOR HEALTHCARE PRACTICE

The analysis reported in this case study has revealed a wide range of ways in which the practice of insulin restriction is linguistically constructed in the context of three English-speaking online diabetes support groups. Rather unsurprisingly, insulin was frequently constructed as inducing unwanted weight gain, while the practice of restricting insulin, by extension, was commonly constructed as the (almost logical) solution to this quandary. Yet, in a smaller percentage of cases, this position was also challenged by messages in which contributors took issue with the very notion that taking insulin necessarily causes weight gain, citing instead the influence of other lifestyle factors, such as diet, exercise and diabetes management. Disclosures of insulin-restricting practices were most commonly met with negative evaluative responses from other group members who warned of the dangers and risks associated with insulin restriction and implored fellow contributors to stop engaging in such practices. Perhaps to avoid such – largely negative – evaluative responses, most

of the contributors who described restricting their insulin did so in ways which omitted the precise details of such practices, for example pertaining to how much insulin or how many doses were ‘skipped’, and how often.

Another finding emerging from this study is that insulin restriction – and, by extension, diabulimia – is likely to be about more than *just* losing and controlling body weight, for central to many of the accounts examined in the foregoing analysis were ideas relating to autonomous diabetes management. In recent times, Western societies have witnessed an increased orientation towards a neoliberal model of public health (Lupton, 1995), according to which individuals are responsabilized into accessing relevant expert health information, proactively managing their health risks and practising self-care, all to reduce the demands that ill-health places on the state (Brown and Baker, 2012). The neoliberal imperatives that imbue notions of health have quite profound implications for people who experience chronic illnesses, including diabetes, who are accordingly implored to take responsibility for and actively self-manage their condition, ideally in accordance with medical advice (Naemiratch and Manderson, 2006). The people writing the support group messages I analysed drew upon this neoliberal framework of diabetes self-care to construct their understandings and experiences of insulin-restricting practices in numerous, sometimes conflicting, ways. For most contributors, insulin restriction constituted a deviation from this neoliberal imperative. On the other hand, a minority of contributors constructed their insulin restriction as a way of relinquishing the demands (and associated powerlessness) of diabetes self-management, while others actually conceived of insulin restriction as a normal part of the ways in which they (and others) responsibly managed their diabetes, signalling a potential overlap between the meaning of so-called ‘normal’ and pathological diabetes-related practices.

Taken together, these various ways of constructing experiences and understandings of insulin restriction provide evidence for a wide range of (potentially conflicting) views and, likely, beliefs, about diabetes, insulin and diabulimia. Healthcare practitioners should therefore be sensitive to the possibility that people with diabetes who restrict their insulin will likely conceive of, and thus possibly talk about, their insulin-restricting practices in various ways, some of which might index potentially disordered bodily practices in subtler, ~~less-obvious~~ ways, than others. The inconsistent and – from a biomedical perspective – potentially problematic ways in which the support group contributors seemed to interact with notions of autonomous diabetes management give cause to question the usefulness of such a neoliberal model of disease management for promoting in people with diabetes more contented attitudes towards their condition, their bodies and their health more generally. Recent discussions on this topic have put forward the possibility of a model of diabetes management which has at its centre the notion of partnership between

patient and practitioner in diabetes care (Anderson and Funnell, 2000). At this point, it should be borne in mind that the contributors to my corpus do not represent *all* people with diabetes or indeed diabulimia, and it is beyond the scope of this study to make firm recommendations pertaining to best practices in healthcare. However, based on the evidence of the support group interactions in my corpus, it would seem that at least some people with diabetes might benefit from such a therapeutic alliance, with management plans designed by both patients and practitioners together. Such a collaborative approach could help to ensure that people with diabetes understand the requirements of their self-care role, leading to reduced inconsistency in what is regarded as ‘good’ and ‘bad’ diabetes self-management in online spaces such as the ones examined here. Moreover, for some people with diabetes, the notion of being part of a therapeutic alliance – as opposed to sole ‘owner’ of their condition – could help them to feel more comfortable and capable with respect to the demands of self-care that are inevitably placed upon them.

7 CONCLUSIONS

This chapter has introduced corpus linguistics and demonstrated how some of its staple techniques, namely keywords, collocation and concordance, can be used to carry out in-depth, quantitative and qualitative examinations of large collections of health-related language data. The case study reported in this chapter, which explored individuals’ subjective constructions of their own (and others’) lived experiences and understandings of diabulimia in the context of online support groups, has been able to provide a novel set of insights revealing the complex and inconsistent ways in which diabulimia’s most essential practice of insulin restriction is discursively constructed, and likely conceptualized, both by people who ostensibly do and do not have first-hand experience of this emerging health phenomenon. Notably, many of the contributors drew upon a neoliberal framework of diabetes self-management, in various ways, to criticize but also justify their own and others’ insulin-restricting practices. While this case study has illuminated issues surrounding the self-management of one chronic condition (diabetes), similar complexities are also likely to exist in relation to issues like self-management and compliance respecting other chronic conditions too.

Corpus methods can be advantageous for health(care) communication research because, compared to purely manual approaches like discourse analysis, they allow researchers to examine larger and more representative bodies of data that are more likely to account for the variation (linguistic and extra-linguistic) that makes up the communicative genre or context under study. Indeed, the corpus examined in this chapter, which contained over a thousand support group messages posted over a seven-year period, was able

to reveal a range of perspectives on the practice of insulin restriction that was much more diverse than would likely emerge from just a single or handful of illness accounts. By basing findings on larger and more representative data sets (i.e. corpora), health(care) communication researchers can provide insights that are more likely to be generalizable to the groups or contexts being researched (Crawford et al., 2014). Meanwhile, quantitative analytical techniques, such as keywords and collocation, offer more statistically robust and replicable starting points for subsequent qualitative explorations, helping researchers to make their analyses more rigorous and objective by basing their findings on patterns and features of the data that are not (just) intuitively interesting, but also frequent and/or statistically salient. However, such a discussion risks implying what is – in the case of corpus approaches to discourse analysis – a false dichotomy between quantitative and qualitative approaches, for whichever technique or combination of techniques are adopted, crucial to any corpus linguistic study is the interplay between quantitative, statistical measures and qualitative, human-led analysis of the data. The computer programs used in corpus linguistics are useful for pinpointing frequent and statistically interesting patterns, but it is up to the human analyst to dig a little deeper (and often wider – drawing on corpus-external sources) to interpret and explain those patterns in a socially and theoretically informed way.

Though undoubtedly rewarding, the corpus approach introduced in this chapter also presented some challenges and limitations. Even as a specialized and relatively small corpus, the collection of support group messages I analysed constitutes a vast data set in discourse analytical terms, and it was not possible within the remit of this chapter for me to pursue every line of enquiry and emergent area of interest. Thus, I restricted the qualitative stage of my analysis to a random sample of 100 support group messages. However, there are likely to be other significant ways of talking about insulin restriction (and diabulimia generally) that my analysis will have missed. Another challenge associated with the size of the corpus is that the data it contains is decontextualized. When converting texts into a text-only format in preparation for corpus analysis, they are stripped of their non-linguistic content, including imagery, fonts, sounds, ~~layout~~ and so on. While this does not present much of an issue for the support group messages I have analysed here, which tend to convey meaning primarily (and in most cases, exclusively) through verbal language, it is likely to constitute a more significant obstacle for researchers wishing to examine characteristically multimodal texts, such as newspaper articles or public health campaigns. One aspect of contextual stripping that *did* cause an issue for the case study in this chapter pertains to the lack of demographic metadata concerning the contributors of the messages in my corpus, which meant that it was not possible for me to ensure that my data was demographically representative of the wider population or make confident generalizations about the diabulimia-related

experiences or perspectives of any particular group. Because of this, the corpus analysed in this chapter would likely fall considerably short of the standards required of data used, for example, in a randomized controlled medical trial. Sampling such a corpus would likely require a significant commitment of time and money.

Although this chapter has focused on language in online support groups, the analytical techniques employed are flexible and can, theoretically speaking, be adapted to analyse language produced as part of any type of text or within any illness or health(care) context (see Jaworska and Kinloch, Loew et al., and Semino et al., in this volume for further examples). In terms of future areas of application, with increasing emphasis placed on evidence-based practice in healthcare training (Brown et al., 2006), corpus linguistics methods could provide practitioners and instructors with substantial evidence bases from which to both teach and learn about the communicative routines of particular (clinical) groups. Given the significant contribution that corpus linguistics has already made to practices in language teaching, including increasingly in the teaching of languages for healthcare purposes (Crawford and Brown, 2010), it is not hard to envisage a role for corpora – as vast repositories of authentic health(care) language – in more general clinical communication training. In the spirit of the case study presented here, a further application of corpora in this context could be to aid (trainee) practitioners when learning about people's subjective experiences and understandings of health-related issues, including emerging health phenomena like diabulimia, by drawing on corpora containing large volumes of naturally occurring illness accounts and narratives. This is an exhilarating period in the history of corpus linguistics, with new corpora and analytical tools emerging at a frenetic pace, constantly pushing the boundaries of what is possible in both corpus linguistics and corpus-aided discourse studies. These developments, coupled with the growing recognition of what the analysis of language can offer to disciplines *outside* linguistics, mean that there has never been a more exciting or opportune time for corpus linguistics research to address real-world health concerns.

ACKNOWLEDGEMENTS

This research was supported by the Economic and Social Research Council (ESRC) (grant number: ES/J500100/1).

NOTES

1. *WordSmith Tools* is a popular corpus analysis software package which offers a range of staple corpus techniques. Users are required to purchase a license (individual or institutional) for a relatively modest fee, though, as alluded earlier, other packages are available for free (e.g. *AntConc*).

2. One difference between my corpus and the BNC that had to be considered was the greater propensity for the latter, as a corpus of online, user-generated content, to exhibit non-standard spellings (Herring, 2001). However, recent research by Smith et al. (2014) has highlighted the robustness of the keywords procedure in this regard, showing spelling variation in computer-mediated communication to have a limited impact on the type and rank of the computational keywords that are generated when genres exhibiting standard versus non-standard spellings are compared.
3. Note that messages appear in their original, unedited form, with instances of non-standard spelling, grammar and punctuation retained.

REFERENCES

- Adolphs, S., Brown, B., Carter, R., Crawford, P. and Sahota, O. (2004), 'Applying corpus linguistics in a health care context', *Journal of Applied Linguistics*, 1 (1): 9–28.
- American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, Washington DC: American Psychiatric Association. Available online: <https://dsm.psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596>.
- Anderson, R. M. and Funnell, M. M. (2000), 'Compliance and adherence are dysfunctional concepts in diabetes care', *The Diabetes Educator*, 26 (4): 597–604.
- Armstrong, N., Koteyko, N. and Powell, J. (2011), "'Oh dear, should I really be saying that on here?'" Issues of identity and authority in an online diabetes community', *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, 16 (4): 347–65.
- Aston, G. and Burnard, L. (1998), *The BNC Handbook: Exploring the British National Corpus with SARA*, Edinburgh: Edinburgh University Press.
- Baker, P. (2006), *Using Corpora in Discourse Analysis*, London: Continuum.
- Baker, P., Gabrielatos, C. and McEnery, T. (2013), *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*, Cambridge: Cambridge University Press.
- Balfe, M. (2007), 'Diets and discipline: The narratives of practice of University students with Type 1 diabetes', *Sociology of Health Illness*, 29: 136–53.
- Baron, N. (1998), 'Letters by phone or speech by other means: The linguistics of email', *Language and Communication*, 18: 133–70.
- Bell, A. (1984), 'Language style as audience design', in N. Coupland and A. Jaworski (eds), *Sociolinguistics: A Reader and Coursebook*, 240–50, New York: St Martin's Press Inc.
- Brezina, V., McEnery, T. and Wattam, S. (2015), 'Collocations in context: A new perspective on collocation networks', *International Journal of Corpus Linguistics*, 20 (2): 139–73.
- Brookes, G. and Baker, P. (2017), 'What does patient feedback reveal about the NHS? A mixed methods study of comments posted to the NHS Choices online service', *BMJ Open*, 7: e013821.
- Brookes, G. and Harvey, K. (2016), 'Examining the discourse of mental illness in a corpus of online advice-seeking messages,' in L. Pickering, E. Friginal and S. Staples (eds), *Talking at Work: Corpus-based Explorations of Workplace Discourse*, 209–34, Basingstoke: Palgrave Macmillan.

- Brown, B. and Baker, S. (2012), *Responsible Citizens: Individuals, Health and Policy under Neoliberalism*, London: Anthem Press.
- Brown, B., Crawford, P. and Carter, R. (2006), *Evidence-Based Health Communication*, Buckingham: Open University Press.
- Clinard, M. B. and Meier, R. F. (2011), *Sociology of Deviant Behaviour*, 14th edn, California: Wadsworth Cengage Learning.
- Colton, P., Rodin, G., Bergenstal, R. and Parkin, C. (2009), 'Eating disorders and diabetes: Introduction and overview', *Diabetes Spectrum*, 22: 138–42.
- Coulson, N. S., Malik, S. and Mo, P. K. H. (2007), 'Researching virtual communities: Some methodological and ethical considerations', *Health Psychology Update*, 16 (3): 3–7.
- Crawford, P. and Brown, B. (2010), 'Health communication: Corpus linguistics, data driven learning and education for health professionals', *International English for Specific Purposes Journal*, 2 (1): 1–25.
- Crawford, P., Brown, B. and Harvey, K. (2014), 'Corpus linguistics and evidence-based health communication', in H. E. Hamilton and W. S. Chou (eds), *The Routledge Handbook of Language and Health Communication*, 75–90, London: Routledge.
- Day, K. and Keys, T. (2008), 'Starving in Cyberspace: A discourse analysis of pro-eating disorder websites', *Journal of Gender Studies*, 17 (1): 1–15.
- Demmen, J. E., Semino, E., Demjén, 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), 205–31.
- Dunning, T. (1993), 'Accurate methods for the statistics of surprise and coincidence', *Computational Linguistics*, 19 (1): 61–74.
- Elgesem, D. (2002), 'What is special about the ethical issues in online research?' *Ethics and Information Technology*, 4 (3): 195–203.
- Evert, S. (2008), 'Corpora and collocations', in A. Lüdeling and M. Kytö (eds), *Corpus Linguistics: An International Handbook*, 1212–48, Berlin: Mouton de Gruyter.
- Eysenbach, G. and Till, J. (2001), 'Ethical issues in qualitative research on internet communities', *BMJ*, 323: 1103–5.
- Firth, J. R. (1957), *Papers in Linguistics 1934–1951*, Oxford: Oxford University Press.
- Fox, N., Ward, K. and O'Rourke, A. (2005), "'Expert patients," pharmaceuticals and the medical model of disease: The case of weight loss drugs and the internet', *Social Science & Medicine*, 16 (6): 1299–309.
- Gablasova, D., Brezina, V. and McEnery, T. (2017), 'Collocations in corpus-based language learning research: Identifying, comparing and interpreting the evidence', *Language Learning*, 67 (1): 155–79.
- Gabrielatos, C. (2018), 'Keyness analysis: Nature, metrics and techniques', in C. Taylor and A. Marchi (eds), *Corpus Approaches to Discourse: A Critical Review*, 225–58, London: Routledge.
- Galasiński, D. (2008), *Men's Discourses of Depression*, Basingstoke: Palgrave MacMillan.
- Goebel-Fabbri, A., Fikkan, J., Franko, D., Pearson, K., Anderson, B. and Weinger, K. (2008), 'Insulin restriction and associated morbidity and mortality in women with type 1 diabetes', *Diabetes Care*, 31 (3): 415–9.
- Gomersall, T., Madill, A. and Summers, L. K. M. (2011), 'A metasynthesis of the self-management of type 2 diabetes', *Qualitative Health Research*, 21 (6): 853–71.
- Hardie, A. and McEnery, T. (2010), 'On two traditions in corpus linguistics, and what they have in common', *International Journal of Corpus Linguistics*, 15 (3): 384–94.

- Harvey, K. (2012), 'Disclosures of depression: Using corpus linguistics methods to interrogate young people's online health concerns', *International Journal of Corpus Linguistics*, 17 (3): 349–79.
- Harvey, K., Brown, B., Macfarlane, A. and McPherson, A. (2007), "Am I normal?" Teenagers, sexual health and the internet', *Social Science and Medicine*, 65 (4): 771–81.
- Herring, S. C. (2001), 'Computer-mediated discourse', in D. Schiffrin, D. Tannen and H. Hamilton (eds), *The Handbook of Discourse Analysis*, 612–34, Oxford: Blackwell.
- Hughes, J. (2010), 'Addressing the new teen trend of "diabulimia": Moral quandaries of a pediatric endocrinologist', *University of Western Ontario Medical Journal*, 79 (1): 11–13.
- Littlefield, C. H., Craven, J. L., Rodin, G. M., Daneman, D., Murray, M. A. and Rydall, A. C. (1992), 'Relationship of self-efficacy and bingeing to adherence to diabetes regimen among adolescents', *Diabetes Care*, 15 (1): 90–4.
- Lupton, D. (1995), *The Imperative of Health: Public Health and the Regulated Body*, London: Sage.
- Lupton, D. (2017), *Digital Health: Critical and Cross-Disciplinary Perspectives*, London: Routledge.
- Malson, H. (2008), 'Deconstructing body weight and weight management', in S. Riley, M. Burns, H. Frith, S. Wiggins and P. Markula (eds), *Critical Bodies: Representation, Practices and Identities of Weight and Body Management*, 27–42, Basingstoke: Palgrave Macmillan.
- Mathieu, J. (2008), 'What is diabulimia?', *Journal of the Academy of Nutrition and Dietetics*, 108 (5): 769–70.
- McEnery, T. and Baker, H. (2016), *Corpus Linguistics and 17th-Century Prostitution: Computational Linguistics and History*, London: Bloomsbury.
- McEnery, T. and Hardie, A. (2012), *Corpus Linguistics: Method, Theory and Practice*, Cambridge: Cambridge University Press.
- Naemiratch, B. and Manderson, L. (2006), 'Control and adherence: Living with diabetes in Bangkok, Thailand', *Social Science & Medicine*, 63 (5): 1147–57.
- Paterson, B. (2001), 'Myth of empowerment in chronic illness', *Journal of Advanced Nursing*, 34 (5): 574–81.
- Paterson, B., Thorne, S. and Dewis, M. (1998), 'Adapting to and managing diabetes', *Journal of Nursing Scholarship*, 30 (1): 57–62.
- Peel, E., Parry, O., Douglas, M. and Lawton, J. (2005), 'Taking the biscuit? A discursive approach to managing diet in type 2 diabetes', *Journal of Health Psychology*, 10 (6): 779–91.
- Rhodes, T., Bernays, S. and Houmoller, K. (2010), 'Parents who use drugs: Accounting for damage and its limitation', *Social Science & Medicine*, 71 (8): 1489–97.
- Scott, M. (1999), *WordSmith Tools Help Manual. Version 3.0*, Oxford: Mike Scott and Oxford University Press.
- Scott, M. (2016), *WordSmith Tools Version 7.0*, Stroud: Lexical Analysis Software.
- Sharma, A. (2013), *Diabulimia: Towards Understanding, Recognition, and Healing*, South Carolina: CreateSpace.
- Shih, G. (2011), *Diabulimia: Diabetes + Eating Disorders; What It Is and How to Treat It*, Lexington: Grace Hui Feng Shih.
- Smith, C., Adolphs, S., Harvey K. and Mullany, L. (2014), 'Spelling errors and keywords in born digital data: A case study using the Teenage Health Freak Corpus', *Corpora*, 9 (2): 137–54.

- Suler, J. (2004), 'The online disinhibition effect', *CyberPsychology and Behavior*, 7: 321–6.
- Thorne, S. (1997), 'Phenomenological positivism and other problematic trends in health science research', *Qualitative Health Research*, 7 (2): 287–93.
- Widdowson H. G. (2004), *Text, Context, Pretext: Critical Issues in Discourse Analysis*, Oxford: Blackwell.
- Williams, R. (1983), *Keywords*, London: Fontana.