

1 *Anxiety, Online Fora and This Study*

Introduction

The following excerpts are taken from posts made online to a health forum about anxiety:

I know that the evil anxiety wants me to give into it. I try not to. I try my hardest to fight it but it's hard when you feel horrible most the time :(

Do diabetics feel guilty because they need meds? No, they don't, and niether should you. Anxiety is a disease/illness just like any other, just not as well understood as most illnesses. No shame in taking this med if you need it, none at all!

I finally found the love of my life after being single so long and my anxiety is destroying everything I constantly second guess her and myself always paranoid

These excerpts were chosen because they demonstrate a range of different ways that people can make linguistic choices to represent anxiety in different ways. The first poster describes anxiety as *evil* and goes on to characterise it as a sentient being which has its own agenda (*wants me to give in*). The poster describes themselves as trying to *fight* anxiety. The second poster represents anxiety as a *disease* or *illness*, comparing it to diabetes, whereas the third poster describes anxiety as part of them: *my* anxiety, then uses superlative language – it is destroying *everything*, they *constantly* second guess their partner and they are *always* paranoid. The poster also describes anxiety as carrying out an action – *destroying*. Like many subjects, it is possible to talk about anxiety in multiple ways. However, we would argue that the language we use around anxiety is likely to play an extremely important role in the way that people make sense of it, which in turn will impact on the ways that they will try to manage anxious feelings and their chances of success in doing so.

As we will demonstrate later, anxiety can be debilitating and is certainly widespread; if anything, it is becoming more commonly

recognised. In this book we examine anxiety in respect of the language used about it, presenting analyses of hundreds of thousands of posts of the type shown earlier. We aim to gain a clearer understanding about the lived experiences of anxiety, what it means to those who experience it, how they have tried to manage it and how they interact with others who are in a similar position. This book will not outline a set of ‘good’ and ‘bad’ uses of language in relation to anxiety – everyone is different and we do not believe that people will benefit equally from a ‘one-size-fits-all’ set of suggestions. Instead, we aim in this book to illustrate the *range* of ways that different people have used language in relation to anxiety. For people who experience anxiety, this may be useful in terms of helping them to recognise their own linguistic strategies, and in cases where they feel they are not coping well with anxiety, the book should offer the possibility of alternatives. In later chapters of the book, we focus on the kinds of language use that are typical of different types of people – young and old, male and female, American and British, indicating that to an extent, the way we conceptualise anxiety is influenced by aspects of our identity or the culture we live in.

In this chapter we begin by identifying the key types, definitions, symptoms and treatments of anxiety as well as providing estimates on the number of people who experience it. We then review existing work on language and mental health, focusing on studies that have used discourse analysis, computational linguistics and corpus linguistics. This section also helps to familiarise readers with aspects of methods that we will be using in our own analyses.

After that we consider analyses of language use in online forums, particularly those related to mental health, which leads to a description of the forum we focus on in this book, the Anxiety Support forum from the site HealthUnlocked. We discuss how we collected posts from this site and how we approached issues such as spelling variation, use of emoji and graphics and the ethical considerations of working with texts posted online. Finally, we provide an overview of the remaining chapters of the book.

Anxiety Disorders

Anxiety is a feeling of unease, characterised by fear or worry, which can be accompanied by physical symptoms. The experience of anxiety in response to a perceived threat is an adaptive behaviour that has contributed to humankind’s very survival. However, a distinction can be made between adaptive and maladaptive behaviours, the latter

requiring treatment and determined to be pathological. Unfortunately, establishing this distinction is not easy and usually requires clinical judgement. As such, better understanding of experiences of anxiety is required, from both the biophysical standpoint of clinical medicine and the experiential viewpoint that can be conveyed through language.

Descriptions of experiences of what would now be understood as anxiety and mood disorders are found in medical texts from Ancient Greece, though it was with the growth of modern psychiatry in the late nineteenth century that diagnostic classifications relating to anxiety were developed (Crocq, 2015). The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (APA, 2013) groups recognised anxiety disorders into three categories: anxiety; obsessive-compulsive disorder (OCD); and trauma- and stressor-related disorders. Distinctions between anxiety disorders are typically based on the situations or objects that are associated with excessive fear and anxiety. For instance, the four most common anxiety disorders are generalised anxiety disorder (GAD), social anxiety disorder (SAD), panic disorder (PD) and agoraphobia (Bandelow et al., 2017) and are characterised, respectively, in relation to anxiety caused by everyday events, social interactions, sudden attacks of fear or anxiety, and anxiety that arises from a concern about being in a situation or place from which it might be difficult or embarrassing to escape (APA, 2013). Other anxiety disorders are similarly labelled in relation to what is the source of excessive fear, such as separation anxiety disorder, health anxiety and phobias (e.g., animal, natural environmental, blood-injection injury). OCD and related conditions, such as body dysmorphic disorder, hoarding disorder and excoriation (skin picking) disorder, are categorised as distinct from anxiety disorders since they are characterised by obsessions and compulsions, rather than anxiety. Trauma- and stressor-related disorders include acute stress disorder (ASD) and post-traumatic stress disorder (PTSD).

Although anxiety disorders are various, there are some common behavioural responses and physical symptoms that are included in the diagnostic criteria across anxiety disorders. The Rethink Mental Illness factsheet on anxiety disorders (Rethink Mental Illness, 2022) lists the following reported effects:

- racing thoughts
- uncontrollable over-thinking
- difficulties concentrating
- feelings of dread, panic or ‘impending doom’
- feeling irritable
- heightened alertness

- problems with sleep
- changes in appetite
- wanting to escape from the situation you are in
- dissociation (feeling like you are not connected to your own body, or like you are watching things happen around you without feeling it).

They also list the following physical symptoms:

- sweating
- heavy and fast breathing
- hot flushes or blushing
- dry mouth
- shaking
- hair loss
- fast heartbeat
- extreme tiredness or lack of energy
- dizziness and fainting
- stomach aches and sickness.

These effects and symptoms can exacerbate the experience of anxiety, resulting in a vicious circle. Despite awareness of associated symptoms, the *DSM-5* states that a diagnosis of an anxiety disorder occurs ‘only when the symptoms are not attributable to the physiological effects of a substance/medication or to another medical condition or are not better explained by another mental disorder’ (APA, 2013, p. 189). This means that individuals experiencing excessive fear or anxiety are typically subject to a range of diagnostic tests designed to identify other conditions in order to rule them out. It also means that a degree of uncertainty is inherent in the diagnosis of an anxiety disorder, since it is characterised not by the positive identification of markers, but the absence of those indicative of other conditions. This uncertainty can itself be a source of distress, which is further complicated by the fact that a given anxiety disorder often co-occurs with another anxiety disorder, and/or with major depression, personality disorders and substance abuse disorders (Bandelow et al., 2017). Neurobiological research has yet to identify specific biomarkers for anxiety disorders and the aetiology of anxiety disorders is more strictly tied to psychosocial factors such as childhood adversity, stress or trauma (Bandelow et al., 2017). As such, understanding an individual’s personal history and appraisal of events in their life is key to managing their relationship with their anxiety.

Health professionals rely on lay descriptions of experiences of anxiety in order to prescribe treatments and to deliver talking therapies such as cognitive behavioural therapy (CBT). Treatment for anxiety

disorders typically involves psychotherapy, pharmacotherapy or a combination of the two and Bandelow et al. (2017) emphasise the importance of individual factors in determining the appropriate treatment, which could include the patient's preference, their history with previous treatment attempts and comorbidities. CBT has been shown to be the most effective psychotherapeutic treatment for anxiety disorders (Carpenter et al., 2018) and is predicated on changing maladaptive beliefs about the perceived threat, that is, manipulating dysfunctional ways of thinking to change patterns of behaviour and support emotion regulation. As a development of CBT, acceptance and commitment therapy (ACT) has been shown to be effective in treating anxiety disorders (A-Tjak et al., 2015), pursuing psychological health in terms of the ability to consciously experience feelings and thoughts as they are and guiding behaviour change according to patients' goals and values (Hayes et al., 2012). Fundamental to this approach is the acceptance (and the reduced avoidance) of anxiety, rather than prioritising symptom reduction per se. One of the strengths of talking therapies such as CBT and ACT is that they can be tailored to patients, since they focus on each patient's values and beliefs. As such, their potential for maximising adherence and increasing self-efficacy in patients is contingent upon understanding the individual perspective.

The World Health Organization (WHO, 2017) reports that the total estimated number of people living with anxiety disorders in the world is 264 million and this total for 2015 reflects a 14.9% increase since 2005 (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016) as a result of population growth and ageing. Furthermore, anxiety disorders were ranked the sixth largest contributor to non-fatal health loss, globally (WHO, 2017). In addition to the personal hardships felt by individuals experiencing anxiety, a systematic review of studies assessing the cost of illness for anxiety disorders reports that the average direct costs corresponded to 2.1% of healthcare costs and 0.2% of gross domestic product (Konnopka and König, 2020). Despite their documented prevalence, anxiety disorders are severely underdiagnosed and undertreated. In a large-scale study of the prevalence of mood, anxiety and alcohol-related disorders throughout Europe, Alonso et al. (2009) report that only 20.6% of participants with an anxiety disorder sought professional help and of those who did contact healthcare services, 23.2% received no treatment at all. Barriers to receiving treatment include clinical shortage, long wait times, social stigma and high treatment costs (Shim et al., 2017), highlighting the need for alternative support and treatment options.

Domhardt et al. (2019) argue that internet- and mobile-based interventions have value in mitigating some of these barriers, by offering time- and space-independent delivery, potential anonymity, and different degrees of human support involving text-based and asynchronous communication. They also constitute low-cost interventions that are scalable. Furthermore, opportunities for self-management can benefit health services looking to make cost savings in the face of populations living longer with chronic disease, and for patients, improving health knowledge and self-efficacy adds to their expertise derived from lived experience (Armstrong et al., 2012). Online support forums are an increasingly common resource facilitating self-management, increasing patient autonomy while also offering social support, which manifests in information exchange and sharing of experiences. Writing with respect to experiences of depression, Sik (2021, p. 756) describes online forums as ‘narrative sandboxes’, where ‘an attempt is made to express the experience of depression, to re-enter intersubjectivity in a low threshold interaction and find mutual explanations to the suffering’. In this sense, online forums support the elaboration of identity-narratives, which are fundamental to the restoration of a coherent self in a world burdened by illness and which offer an alternative to the biomedical paradigm that is imposed when those with mental health issues engage with health professionals (Sik, 2021). The anonymity of online spaces facilitates the discussion of sensitive topics as well as helping to overcome geographical and social isolation and providing opportunities for gathering practical information and advice (Sik, 2021). In this book, we investigate an online forum offering support for anxiety disorders, focusing on the representations offered by those with experience of anxiety disorders through the language they use to report those experiences, and the communicative exchanges that demonstrate how online communities provide social support. In focusing on language as the conduit through which feelings, beliefs and discussions about anxiety disorders are expressed, we can consider how members of the forum reconfigure lay and expert knowledge about anxiety, which can subsequently inform how those with anxiety are supported, including by health professionals.

Investigating the Language of Anxiety Disorders

Language is crucial to the ways that we convey our experiences of health and illness to health professionals, family members and friends in order to obtain treatment and support. Language also constitutes how we make sense of health and illness as part of a wider life

experience. Language-based studies of the experience of anxiety disorders, like those of many other types of mental illness, have sought to describe patterns in language use that reflect and characterise a broader psychological experience. One of the earliest studies in this area was carried out by Capps and Ochs (1997), who identified a ‘grammar of panic’: a shortlist of language features that they reported were characteristic of experiences of anxiety and more acute panic attacks. This included the use of adverbials such as *all of a sudden* and a high proportion of mental verbs such as *think* and *realise*. Furthermore, they found that participants positioned themselves as *helpless* and that this was realised through:

- Referring to self in semantic roles other than agent or actor
- Using modal verbs that frame action as necessity (i.e., not voluntary)
- Using grammatical forms to imply failure to achieve a goal
- Using grammatical forms that intensify vulnerability and deintensify one’s ability to cope and control.

(Capps and Ochs, 1997, p. 67)

The identification of linguistic patterns, as used in particular contexts and by certain participants, intended towards select communicative ends, is consistent with investigations of discourse. The analysis of discourse is both linguistic and social, in that we discuss the construction of discourses in terms of the aggregation of precise linguistic features, with a view to understanding how people use language to manage interactions, cultivate particular representations of themselves, other people and objects, and express ideas and beliefs – with varying degrees of explicitness. Thus, we can talk about language at the level of word choices, standard grammatical relations and textual features such as orthography. When we consider what is appropriate to a particular mode of communication – such as an online forum – and reflective of particular social conventions, such as the language used by health professionals during a consultation, the awareness that the communicator has for what is typical gives us insights into the discourses shaping that interaction.

Georgaca (2014, p. 55) discusses how discourse analytic approaches draw on principles of social constructionism to investigate ‘systemic ways of speaking’ about mental health and illness across clinical categories, public and policy texts, as well as user experiences. Studies in this area work to destabilise the taken-as-given nature and objective status of, for example, medical descriptions of illness and demonstrate that definitions and representations are socially and historically constructed. For example, medical discourses have historically conceptualised mental distress as illness, located in the human

body while positioning patients as passive sufferers and medical professionals as experts. Alternative discourses work to empower the individuals who experience the distress as active agents and as having expertise derived from their lived experience, which has in turn informed effective therapeutic interventions (Nissling et al., 2021). Studies of representations of illness in the media have also set out to ‘disrupt’ the discourses that have represented those who experience mental disorders as dangerous, criminal, pitiable or diminished, subsequently fuelling the stigmatisation of mental illness (Hannaford, 2017). Georgaca (2014, p. 58) concludes that ‘This field of social constructionist research which critically examines the production and maintenance of dominant clinical categories and attempts to denaturalize and destabilize them through a series of deconstructive strategies has been one of the major contributions of discursive approaches to critical mental health work.’ As mentioned earlier, online forums offer one such space where those with knowledge derived from experience can recontextualise and reconfigure the discourse around anxiety that may come from health professionals or the media. Through a close examination of language produced in this context, we are afforded a view of how social and institutional practices are realised at the micro-level.

Informing our analysis of the language used in online forums is a recognition that there are certain textual features and social conventions associated with the platform as a form of computer-mediated communication (CMC). Researchers have attributed the structural features of CMC to its technical affordances, in that while users cannot utilise the kinds of paralinguistic features associated with face-to-face communication (intonation, gesture, facial expressions), there are options more germane to online communication – such as hyperlinks, emoji and multimodal features such as video. Herring and Dainas (2017) collectively refer to the visual elements of CMC as ‘graphicons’, which would include emoji, stickers, graphics interchange format (GIF) files, images and videos. In addition, there are CMC practices that serve as proxy to audible features, such as CAPITALS to indicate volume (Riordan and Kreuz, 2010). Not only will the data we observe be influenced by what the forum allows users to contribute, at a technical level, but those technical aspects will also shape the interpersonal dynamics of how participants (are able to) interact with one another. For instance, members are afforded a high degree of anonymity and the opportunity to manage which aspects of their disembodied identity are visible to others. This, combined with the fact that participants are more likely to be communicating with peers, means that support groups potentially offer a more deliberative space compared

with, for example, exchanges between patients and healthcare professionals where there are more tangible power asymmetries. As such, we are interested not only in how members of the forum craft their posts using text characters and emoji, for instance, but also how they express affiliation and operate as a peer community. By linking the social practices of an online forum to specific features of language, we can use computer tools and software to help us process the multitudinous forms of expression and interaction between the vast number of users of online forums. Computing these textual features on a large scale supports us in identifying how even simple word choices aggregate to form the discourses that characterise representations of anxiety in this context, as we will now explain.

Computational Linguistics

Recent studies investigating language features associated with communities discussing anxiety and mood disorders have benefitted from approaches in computational linguistics and natural language processing (NLP) to conduct large-scale studies of, in particular, CMC, such as social media and forum posts. Automated procedures that enable us to process lots of data quickly also help us to discern statistically meaningful patterns in language use. For instance, there have been indications that increased use of first-person singular pronouns (*I, me, my*) is a robust indicator of depression, though their use in relation to anxiety is sparser and, at times, contradictory (Haase, 2021). Al-Mosaiwi and Johnstone (2018) report that anxiety, depression and suicidal ideation forums contain more absolutist words than control forums (categorised as general, asthma, diabetes and cancer forums), with these especially favoured in suicidal ideation forums. The researchers generated their own absolutist and non-absolutist dictionaries through introspection, with both categories indicating magnitudes and probabilities but absolutist words doing so without nuance (*always, totally, entire*) compared with no-absolutist words, which carry some nuance (*rather, somewhat, likely*).

One of the more commonly used tools in language-based investigations of mental health and illness is the Linguistic Inquiry Word Count (LIWC) program (Tausczik and Pennebaker, 2010), a text analysis program that calculates the percentage of words in a given text that falls into one or more of over 80 linguistic, psychological and topical categories indicating various social, cognitive and affective processes (e.g., 'negative emotion'). Lyons et al. (2018) used the LIWC program in comparing the linguistic content of individuals in online communities for different types of mental distress (generalized anxiety disorder,

borderline personality disorder, major depressive disorder, obsessive-compulsive disorder, and schizophrenia) and a control group (a finance discussion forum). Compared to the control group, people who experience one of the forms of distress displayed a higher frequency of singular personal pronouns and higher frequency of negative emotion word use.

Although frequently deployed to investigate language-in-use, the LIWC program does have certain limitations that inhibit the extent to which we can understand how language is used in context. Hunt and Brookes (2020) argue that in relying upon individual words and/or rudimentary word combinations, the tool is likely to lead to simplistic analytical claims that do not take into account various levels of context (including co-text, register or audience design). As such, Slonim (2014, p. 17) argues that LIWC is not suited to tracking the ‘ebb-and-flow’ of language across time, or even across the course of a text. Slonim (2014) also offers a critical evaluation of the studies that have used LIWC, identifying mixed results on purported claims that, for example, people with depression use the first-person pronoun ‘I’ more than other people, use more negatively valenced emotion words (and fewer positively valenced emotion words) and more cognitive mechanism words (e.g., *think*, *realise*). Slonim (2014, p. 15) argues that one of the limitations of these studies is that they have not identified markers that are specific to depression and in the case of ‘I’, this is acknowledged by Pennebaker (2011, p. 289), who states that:

Depending on the context, using I-words at high rates may signal insecurity, honesty, and depression proneness but also that you aren’t planning on declaring war any time in the near future.¹ Using I-words at low rates, on the other hand, may get you into college and boost your grade-point average but may hurt your chances of making close friends.

Whether using the pre-determined LIWC categories or an alternative, there is a tendency in works relying on automated classification and processing to develop bespoke dictionaries that purportedly tie lexical items to psychological constructs; however, those links remain unsupported. Slonim (2014, p. 16) refers to studies by Vanheule et al. (2009) and Ramirez-Esparza et al. (2008) as examples where it is implied that semantic categories of words (such as ‘positive emotion words’) are indicative of positive or negative affect, but there is no theoretical explanation to support this proposition. As such, while it is not particularly surprising to find that people experiencing mood disorders

¹ Pennebaker (2011) provides the anecdotal example that Harry Truman used far more I-words than Barack Obama.

and anxiety refer to ‘negative emotions’ more frequently, we must also be cautious about inferring that the use of these terms is indicative of mental state.

NLP approaches have been applied as a form of clinical linguistics, identifying metrics such as lexical diversity, readability scores, sentence complexity, negation, uncertainty and degree of repetition as a means of identifying and even predicting specific mental health conditions (Conway et al., 2019, p. 213). However, Perkins (2011, p. 926) argues that identifying language ‘dysfunction’ and ‘atypicality’ when compared with ‘normal’ usage ‘cannot satisfactorily be accounted for in terms of a deficit in a language “module”, and that ‘competence/performance-type explanations are an unhelpful and even misleading way of characterizing these phenomena’. Our approach, then, is not to pursue the identification of ‘linguistic markers’, since as Hunt and Brookes (2020) assert, making any claims as to what (emotional) suffering looks like in relation to health and illness runs contrary to the diversity with which people talk about it. Our approach comes from a discipline that is firmly grounded in linguistic theory and founded on the recognition that language is used in context. A corpus linguistic approach capitalises on the speed and processing power of computers; however, it also raises important questions about how to identify, quantify and analyse features of language in the context in which they are used.

Corpus Linguistics

Corpus linguistics refers to a set of procedures that allow us to process large datasets and make observations of patterns in how language is used, supporting claims about how people discuss experiences of health and illness, for example. We refer to our large dataset as a ‘corpus’ (pl. corpora), which is collected according to principled sampling procedures in order to be representative of a domain of communication and documented in a machine-readable format to allow quick calculations and searches using a computer. The emphasis on computational processing of large datasets in corpus linguistics affords researchers the opportunity to base their observations on more representative bodies of text and, thereby, make their findings more generalisable. Nevertheless, the emphasis on viewing language in its original context and the application of qualitative forms of analysis supports investigations of language that are sensitive to the ways in which language users construct their texts – such as forum posts – with an awareness of structure, sequence and their audience. Corpus linguists therefore caution against the decontextualised analysis of

language and while certain procedures, say, count individual words, this is regarded as a preliminary step in an inquiry that returns to the situated context in which those words appear.

Applications of corpus-based approaches to different types of health communication have afforded insights that help us to understand how health professionals and lay communities alike report aspects of health and illness with a view to optimising healthcare outcomes. Focusing on the communication strategies of health professionals during consultations with patients, researchers have identified patterns in how consultants use interpersonal strategies (Thomas and Wilson, 1996) and how health professionals involve patients in decision-making (Skelton et al., 1999). Quantitative methods can identify general patterns, but these patterns exist in a complex context that can only partly be described quantitatively and so quantitative statements ‘should always be accompanied by detailed qualitative analysis’ (Skelton et al., 1999, p. 621). One of the fundamental procedures of corpus linguistics is keyness analysis, which helps to identify features that are particularly frequent in one dataset compared with another. Adolphs et al. (2004) used this approach to identify salient features in NHS Direct exchanges between professionals and patient callers, finding that in addition to a highly involved, interpersonal style (indicated, for example, in the regular use of the second-person pronoun, *you*), health professionals balanced direct instructions and imperatives (such as *try*, *take* and *avoid*) with vague language, which provided optionality for the caller and performed important politeness work.

News media are also an important source of information regarding health issues and corpus linguistics has supported researchers in conducting longitudinal studies of health coverage. Price (2022) explored representations of mental illness in the UK press between 1984 and 2014 to consider their contribution to stigmatisation and found that such reports used identity-first forms (*a schizophrenic*) to refer to people with mental illness, more than person-first forms (*a person with schizophrenia*). Price (2022) also found that the press tended to represent the process of having mental illness as *suffering*, but first-person accounts from people with mental illness are more likely to refer to *experiencing* mental illness. Hannaford (2017) similarly investigated UK press coverage between 1995 and 2014, which was examined alongside the UK National Attitudes to Mental Illness Survey (AMIS) (TNS BMRB, 2015). Hannaford (2017) reports an increase in destigmatising coverage of mental illness over the period, as well as a decline in coverage of people with mental illness as dangerous, which correlated with a decrease in negative attitudes towards people with mental illness among the public.

Atkins and Harvey (2010) applied a keyness analysis (described in more detail later) to explore the lay perspective as conveyed through health enquiries submitted anonymously to a doctor-led health advice website. This approach supported them in identifying some of the common misconceptions about sexually transmitted infections and the online format seemed to facilitate an uninhibited discussion of sexual health issues. Harvey (2012) has investigated the same collection of adolescent health queries to explore questions and disclosures around depression. He reports two important recurring constructs: *I am depressed* and *I have depression*, which ‘encode different perspectives and meaning making with regard to the conceptualisation of depression, the former describing depressive experiences as a reaction to negative life events, the latter portraying depression as a pathology originating within the individual’ (Harvey, 2012, p. 349). Harvey (2012) thus argues that the language choices made by those submitting queries to the website reflect broader attitudes towards mental health that are informed by the language of ‘psychiatrization’ and the construction of sadness as a clinical problem.

There are also studies investigating online forums using corpus linguistic methods, relating to various health concerns. Demmen et al. (2015) investigated an online support group for cancer and end of life care involving patients, family carers and healthcare professionals and reported a wide range of violence metaphors (*battling* cancer, for example) used to various effects, but with greater frequency in the online forum context, compared with face-to-face interactions. Kinloch and Jaworska (2020) explored the lived experience of postnatal depression (PND) through a comparative analysis of discourses about the condition produced by mothers in an online discussion forum, the medical profession and the UK print media. They were particularly interested in how mothers positioned themselves in relation to the social stigma cultivated in media discourses around PND and found that ‘secrecy’ – that is, hiding symptoms and experiences of PND – was a key theme in the data that could be directly linked to stigma.

The studies reported here, which use corpus linguistic methods to investigate health communication in various contexts, each report one or more of a handful of techniques that have become recognised as the fundamentals of corpus linguistics (Brookes and McEnery, 2020) and which we rely on in the analyses reported in this book. One simple procedure for initiating a quantitative analysis is to extract frequency information, which is typically readily available through corpus analysis tools (software programs) in the form of wordlists. This tabulated presentation of the most frequent items (typically, words) in the corpus

can provide a very quick yet illustrative overview of a dataset and is often enabled for sequences of words, or word categories. However, in order to contextualise the raw frequencies, corpus linguists typically provide a reference point for comparison. This might be reported as a relative frequency: a normalised value that converts the actual number of occurrences to a standardised measure, such as x words-per-million (wpm), in order to make it easier to compare the rate of occurrence of a feature in corpora of different sizes.²

The keyness approach mentioned earlier is an extension of this principle in that it is based on a comparison of the proportional frequency of a feature (e.g., a word) in two (or more) corpora and combined with a statistical test to determine if differences in the relative frequencies are statistically meaningful. Through identifying what is 'key' in one text compared with another, we begin to establish what is salient to the data we are investigating, as a matter or convention or an area of particular interest to communicators in that context. Conventionally, this baseline is determined from a larger reference corpus, though the selection of the reference data has implications for what emerges through our comparison. Our Anxiety Support forum data represents content that is not only characterised by the topic of anxiety disorders but also the mode, as a form of CMC. If we were to compare this to a corpus of general English (such as the BNC2014), the results would likely highlight features that are characteristic of online forms of communication, as distinct from spoken and written forms. As such, we can select a reference corpus that is more closely matched in terms of register, that is, one containing examples of online communication to 'neutralise' – or normalise – the frequent occurrence of features indicative of the online mode. In order to facilitate a comparison that tells us more about the nature of the forum and the topic of anxiety, we can refer to the enTenTen20 Corpus of the English Web (hereafter, the English Web 2020 corpus): a corpus of 36 billion words of 'linguistically valuable web content' collected between 2019 and 2021, including blogs and documents covering technology, sports, business and so on from a range of English-language domains (e.g., UK, US, Australia) (Jakubíček et al., 2013). We refer to the English Web 2020 corpus to consider how the Anxiety Support forum compares to general web-based English, though in later chapters we also divide the forum data into sub-corpora in order to conduct statistical comparisons between parts of

² For instance, if a feature occurs 100 times in a corpus of one million words (100 wpm), this is comparatively more often than occurring 150 times in a corpus of 10 million words (15 wpm).

the data defined by contextual factors, namely whether the contributors identified as female or male (Chapter 5); as from the UK or the USA (Chapter 6); and how messages from different timepoints compare (Chapter 7). As such, we will be reporting the results of different keyness analyses as we investigate different segments and populations within our study of an online Anxiety Support forum.

We have also stressed that to look at words in isolation offers only a restricted view of how language is being used and so an essential aspect of corpus linguistics is to look at language in context. Corpus analysis tools offer a range of ways through which we can investigate the relationships between words and patterns of word combinations. One such technique is called collocation, which highlights statistically meaningful word combinations in a corpus. This often involves identifying a relevant search term (which could itself be identified through frequency or keyness analysis) and then using a corpus analysis tool to identify words which more often occur in combination with the search term than without. This can highlight meaningful associations and comparisons between corpora from different communicative contexts. For example, Kinloch and Jaworska (2020) found that in an online forum, *woman* collocated with their search term *PND* but in media texts, the term *sufferer* was more likely to be used to describe someone experiencing PND.

Techniques such as frequency, keyness and collocation analysis are often referred to as a ‘point of entry’ (Adolphs et al., 2004) or a ‘way in’ (Kinloch and Jaworska, 2020) to the data, typically in order to allow for more qualitative investigations. This qualitative stage of the analytical process is often informed by other language-based methods, such as discourse analysis, and requires more manual input from the researcher. A common feature of many corpus analysis tools is concordancing: a presentation format of the data which allows the researcher to observe sequential occurrences of a search term as a series of rows, with a snippet of the preceding and succeeding text either side. This is alternatively labelled a ‘key word in context (KWIC)’ view and offers a glimpse of the original text in each instance that is often sufficient for helping the researcher to discern patterns in how the term is used. Of course, for a more informed view, the researcher should familiarise themselves with the full text and concordance tools are often hyperlinked in a way that allows the researcher to select one of the occurrences and activate a view of the complete text (in our case a single forum post or the sequence of messages the post occurs in).

The move to concordance lines is usually the point at which the researcher is tasked with demonstrating that they are a ‘socially- and

linguistically-informed human analyst' (Baker et al., 2019, p. 224) in order to 'address the subtler ways in which meaning is created' (Baker et al., 2019, p. 223). Accordingly, McDonald and Woodward-Kron (2016) augmented their corpus-based investigation of online support groups concerned with bipolar disorder with systemic functional linguistics (SFL) as a theoretical framework, focusing on interpersonal and experiential meanings in their investigation of member roles and identities in the forum. Following Harvey's (2012) observations of the formulations *I am depressed* and *I have depression*, McDonald and Woodward-Kron (2016, p. 169) use this framework to articulate their observations of how members of the bipolar disorder support forum foregrounded possession of the disorder (*I have bipolar disorder*) over an intensive attributive construction (*I am bipolar*). They explain that the process of ownership conveys a degree of control over their possessed condition (McDonald and Woodward-Kron, 2016).

In the documentation of socially relevant themes, researchers might also draw on other approaches germane to the social sciences, such as content analysis, in order to bring together observations of comparable linguistic features. For example, Hunt and Brookes (2020, p. 88) explain their categorisation of keywords derived from a corpus of posts to a forum concerned with anorexia into themes that represent 'central areas of meaning that characterize the forum users' discussions of anorexia'. As a result, Hunt and Brookes (2020) are able to account for a larger proportion of their keyword list than would be possible if they simply took each word in turn, which are collated around themes such as 'recovery', 'feelings and emotional responses' and 'forum-related' aspects. We similarly set out not to simply describe the formal, linguistic features of the data but to consider how the language features identified through procedures of corpus analysis reflect more meaningful aspects of the lived experiences of the users of the forum. Thus, utilising corpus linguistics in combination with other (qualitative) approaches has afforded insights that direct us to larger patterns but also the local, contextualised usage of important terms, text composition and interpersonal aspects of the forum.

Our approach therefore combines the computational identification of frequent and salient patterns via corpus linguistics with a close reading of text that takes into consideration context, via discourse analysis. Such a combination is sometimes referred to under the paradigm Corpus-Assisted Discourse Studies (CADS), as described by Partington et al. (2013). One aim of CADS is to avoid carrying out a politically motivated analysis (Partington et al., 2013, p. 339), such as those sometimes associated with Critical Discourse Studies (CDS), which characteristically focuses on cases of manipulative language and

abuse of power. For the research in this book, while we largely align with the methodological techniques associated with CADS, we also feel that at times it is worth viewing the data through a more critical lens, especially if it articulates a discourse that might be seen to be impeding someone's recovery; for example, that men should not talk about their feelings because it is seen as unmanly.

Online Forums

In our review of the literature on applications of (corpus) linguistic approaches to experiences of healthcare, we have shown that online forums are one of many sites in which we can investigate discussions of health conditions and representations of those experiencing illness. Internet forums have emerged as a venue for individuals to come together, share experiences and provide mutual support (Eysenbach et al., 2004). They are often the first port of call when health issues emerge (Fox and Duggan, 2013), since they are typically free and devoted to specific health topics. Furthermore, from a psychological perspective, there is value in knowing that others have comparable experiences. Joinson (2003) describes a process of social comparison that is directed both downwards – whereby self-esteem derives from knowing that others are worse off than you – and upwards – in finding that there are others who can guide you. Indeed, the reported benefits of online support forums are typically characterised in terms of information and emotional support (Yip, 2020). Furthermore, online spaces may offer something that is not available to individuals in their offline world. The accessibility of online resources can mitigate the spatial, temporal and other institutional barriers of offline healthcare services. In addition, users of online forums have reported feeling more comfortable discussing sensitive and personal issues, particularly those subject to stigmatisation, in the relative anonymity afforded by the online context (Webb et al., 2008). This has been shown to be of particular importance in terms of engaging younger people in healthcare services, offering 'a secure platform from which to ask awkward, sensitive or detailed questions without the fear of being judged or stigmatized' (Harvey, 2012, p. 351).

Forums are not the only type of online peer support available and researchers have also investigated how individuals with serious mental illness use social media to share their experiences and seek advice. The reported benefits, in relation to the various kinds of online peer support, include greater social connectedness, feelings of group belonging, gaining insights into important healthcare decisions that can promote healthcare-seeking behaviours and promotion of

treatment engagement (Naslund et al., 2016). Eysenbach et al. (2004, p. 5) have remarked that ‘virtual communities are in fact the single most important aspect of the web with the biggest impact on health outcomes’. The potential risks of online spaces, particularly if they are not moderated and/or do not involve quality information from health professionals, can include misleading information, hostile or derogatory comments from others and creating feelings of more uncertainty about one’s mental health condition. However, Naslund et al. (2016) conclude that based on the evidence, the benefits of online peer-to-peer support outweigh the potential risks and they play an important role in motivating untreated and undiagnosed community members to seek professional help (Powell et al., 2003). Indeed, McDonald and Woodward-Kron (2016) observe that advice provided by veteran members to newcomers is often aligned with mainstream biomedical norms and research has shown that rather than operating in conflict with professional health advice, participation in online forums has led to users feeling better informed, having more confidence in their relationship with their physician, with better knowledge of the treatment options available, and gaining a greater sense of optimism about their prognosis (van Uden-Kraan et al., 2008).

For the purposes of research, online support forums provide a readily accessible platform for text-based data that is, to a large extent, unmoderated by health professionals. They thereby provide the opportunity for us to investigate naturalistic, unsolicited language that principally conveys lay descriptions of, in our case, anxiety disorders. The influence of media and institutionalised discourse are, nevertheless, relevant in these accounts and Kinloch and Jaworska (2020) show that members of a forum related to postnatal depression (PND) draw on biomedical perspectives of PND. While there are criticisms of the reductionist portrayal of depression as a biological condition independent of cultural and social factors, the authors argue that it can help people ‘recognize PND as a real illness in need of treatment and potentially encourage them to seek medical attention’ (Kinloch and Jaworska, 2020, p. 95). This shows that any given discourse can have value in some instances and limitations in another, demonstrating that such discourses can be resources used to various effects. Furthermore, Schofield et al. (2020) show that the various discourses around anxiety disorders directly influence perceptions, contributing to the degree of ‘blame’, self-efficacy and, ultimately, stigma, associated with those who experience social anxiety disorder (SAD).

Given that online forums are purportedly valued largely in terms of the informational and emotional support they offer, of key interest to us as researchers of health discourses are the ways in which members

of the forum interact with one another and how they influence each other's language when discussing experiences of anxiety. Kouper (2010) has shown not only that advice exchange is common among members of an online (motherhood) community but also that messages that solicit and provide advice have distinctive structural and pragmatic features. These often extend into longer units of language that follow a conventional story-telling structure (Lindholm, 2017). We can also observe how particular members assume certain identities, which may evolve over the course of their participation in the forum as their knowledge is established and their need for guidance from others diminishes. Indeed, Sillence (2010) identified a set of users within a cancer support group who collectively conveyed advice-giving as one of their key functions and had developed mechanisms for portraying their competence and trustworthiness. In addition to providing information and offering empathy, more established members also have a role to play in determining the conventions for how messages are formulated in the forum. McDonald and Woodward-Kron (2016), for example, have demonstrated through lexicogrammatical and discourse-semantic choices how new members of a bipolar disorder online support group are socialised into health discourses (i.e., by 'senior' members), thus establishing what it means to communicate as a senior member as well as moderating the wider communicative practices of the forum. Ultimately, as Brookes (2020, p. 46) argues, there is clear value in analysing the representations of health illness that members of online support forums provide through their posts 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'.

HealthUnlocked: the Anxiety Support Forum

The data analysed in this book comes from an online support forum, collected over a period of approximately nine years. The website that hosts the forum is managed by HealthUnlocked, a private company that:

aims to transform individual health experiences into support, insight and understanding for others. We do this by enabling people to share personal health experiences and information online using our site ("Our Site"). In turn this provides support, aids self-management, and improves interactions with professionals, with the aim of improving day-to-day health and well-being. (<https://support.healthunlocked.com/article/148-privacy>)

The site hosts over 300 ‘communities’: distinct forums that are characterised by their focus on a particular health concern. Our data comes from one of these ‘communities’ labelled Anxiety Support, which is the largest of the communities explicitly marked as focusing on anxiety disorders.³ When members register for HealthUnlocked they are asked whether they are happy for their contributions to be used for research purposes. Thus, the data we obtained with permission from HealthUnlocked only comprises posts made by members who consented for their contributions to be used. Posters make use of usernames such as Charles365 rather than their real names. The data given to us by HealthUnlocked have been run through an automatic system to identify names and anonymise them, although this was not 100% effective and we have redacted names which were not anonymised along with any other identifying information, including usernames, when quoting any forum posts.

The corpus comprises 294,082 comments from the period 20 March 2012 to 14 September 2020, which represents approximately 21 million words. There are 17,770 different contributors posting messages to the forum, who identify as being from 141 different countries around the world. The size of the dataset demonstrates how an approach supported by computer processes is suitable for attempting to account for so many posts and the coverage allows us to investigate longitudinal aspects of the dynamics of the forum and to explore different subsets of contributors according to characteristics such as age, gender and nationality.

Much like many other online forums, members of the Anxiety Support forum can post messages either in response to contributions made by other members, or by initiating a new discussion thread. Messages are then presented as a chronological sequence of responses. The messages on the Anxiety Support forum are largely text-based, although they do include emoji and users can also post images and image macros overlain with text, which were removed from the data transferred to us. For the kinds of corpus procedures we have described earlier, the data needs to be machine-readable (McEnery and Hardie, 2012) and the inclusion of image content would necessitate a form of annotation to document salient features of the images in text. While researchers have demonstrated how a coding scheme can be applied to document information about images in a multimodal corpus analysis, such as Bednarek and Caple’s (2017) investigation of newsworthiness in news texts, there is no standard form of annotation

³ HealthUnlocked also hosts an Anxiety and Depression Support forum and a Living with Anxiety forum.

that can readily be applied across datasets and which can capture the multitudinous variety of what can be represented in graphical format, which is particularly difficult when we consider both the denotative elements (what can be described as literally represented) and the connotative elements (the interpretation of what is represented). Furthermore, for us to have incorporated a record of such images, we would have had to have access to them and to evaluate and code them. This was not available to us, so while these are a feature of the forum, they do not appear in the data and, subsequently, our analysis.

Hunt and Brookes (2020, p. 70) discuss the loss of graphicon material in their construction of a corpora based on online forums concerned with eating disorders and determine that this material performed 'a largely supplementary function to the linguistic exchange of information in the fora, often repeating visually what is explicit in the adjacent text rather than conveying new information'. They also acknowledge that graphicons more evidently relate to the negotiation of interpersonal relationships online, which was not a research priority for their work (Hunt and Brookes, 2020) and Zappavigna (2012, p. 101) similarly states that memes 'are deployed for social bonding rather than for sharing information'. These interpersonal functions are relevant to our exploration of the Anxiety Support forum, however, they were not available to us once the data was converted to text by HealthUnlocked.

What we can comment on, however, are modifications to orthography that similarly operate in interpersonal ways, such as abbreviations (*lol*), elongation (*thaaaanks*) and vocal spellings (*wot*). With a dataset involving participants from a range of countries, we can also expect orthographic inconsistency as a result of different spelling conventions (e.g., *humor/humour*). This has implications for the automated word counts that are the basis for many corpus analysis procedures and there are tools that have been developed to navigate spelling variations according to historical or social norms. The Variant Detector (VARD) software tool, for example, was developed as a pre-processing tool to regularise variant spellings in a corpus prior to other linguistic processing (Baron and Rayson, 2008). The tool enables the researcher to collate and count these variants, while also allowing them to view and, thereby, study those variations. Nevertheless, with a corpus of approximately 21 million words, manual processing of candidate variants would still be required to account for all the possible variations. In a sense, people who do not spell words correctly are likely to have their contributions downplayed in the analysis. So a concordance analysis of how people use the word *anxiety* (which occurred 147,850 times) will miss the 144 cases of

anxiety, 59 instances of *anxioty* and 40 cases of *anxiey* in the corpus. On reflection, we have not corrected mistakes as the amount of spelling variation was quite small and does not impact on the typicality of patterns that were found.

The Anxiety Support forum is facilitated by the USA-based men's cancer survivor support charity Malecare and its executive director serves as a moderator for the forum. Among its community guidelines, the Anxiety Support forum states that it 'is not a substitute for 1-to-1 medical consultations or acute care' (<https://healthunlocked.com/anxiety-support/about>), echoing HealthUnlocked's own policy that any content on the site is 'never a substitute for professional medical advice', that members should 'always speak to a doctor or other health practitioner about your condition and/or treatment or changes to your condition and/or treatment' and, ultimately, 'Never delay seeking advice or dialling emergency services because of something that you have read on Our Site' (<https://support.healthunlocked.com/article/147-terms>). A similar message is pinned alongside the discussion threads in which members take part.

The community guidelines also model how members should formulate their messages, such as using 'For me, this worked' rather than 'You should do this' (<https://healthunlocked.com/anxiety-support/about>). This shows that community moderators recognise the significance of language used to proffer advice with varying degrees of directness, which we will be discussing in our own work. It also shows that alongside concerns for the quality of health information exchanged on the site, moderators are concerned with the communicative style adopted by users in their exchanges. Ordinary members also have the capacity to moderate the content that they see. In addition to replying to messages, users can 'Like' a post, save it to create a personal archive, or report it if they feel it is in breach of the community standards. It is the responsibility of the community moderators to deal with reported posts, although the HealthUnlocked team also manage posts that do not conform to their Terms of Service. Though we are not privy to the messages that have been removed – nor can we know how many that is – the functions for reporting, the establishing of community standards and the roles of moderators remind us that online support forums are policed by various stakeholders and that this can have implications for the content that we, and users of the forum, subsequently see.

The 'disinhibition' of online spaces (Suler, 2004) that facilitates personal disclosures at the same time can encourage users to abandon politeness concerns and the potential for inflammatory or abusive content in such spaces has been documented (Hardaker, 2010).

When we take this alongside the highly sensitive personal disclosures that arguably constitute the ‘core’ content of a forum concerned with a health issue such as anxiety disorders, there are important ethical concerns to consider when we want to explore this kind of data. Townsend and Wallace (2016) outline four key areas of concern when conducting research online, which we will briefly discuss here: private versus public; anonymity; informed consent; and risk of harm.

Online spaces often make it difficult to determine whether content is public or private, and the implication for research is that with private content there is a greater need to inform participants of your work and to get their approval, by way of consent, before you begin. In order to access HealthUnlocked, users sign up and create a profile, which then allows them to access any one of the 300+ communities hosted on the site. Non-members have a very restricted view of content, that is, there is arguably enough of a preview for them to get a sense of how the site operates and what members discuss in order to make the determination whether they will become members. Nevertheless, they cannot view full posts or threads. In this sense, the data are not public. However, as a private company, HealthUnlocked stipulates in its terms of service that it works – and shares its content – with research partners. Our access to the data is the result of such a partnership. The data are anonymised, in order to protect the identities of those who use the site (which they do with a self-assigned username), and members have the capacity to manage their own settings and opt out of data-sharing practice. This means that the data as discussed in this work are anonymised and that there are some contributions that are absent, on the basis that some users elected not to consent to their contributions being used for research.

One of the concerns expressed by members of an online forum in relation to privacy and research is that the ‘meaning’ of a post is lost when taken out of context (Golder et al., 2017) and this has raised questions about researchers’ capacity to understand the communicative practices and norms of the forum (or other situation) they are observing. In response, researchers have often adopted an observer-as-participant role (Mackenzie, 2017), taking an ethnographic approach that involves familiarising themselves with the communicative environment and participating as a member. This supports a more informed, qualitative approach and the types of ‘thick’ description associated with ethnographic research (Geertz, 1973). We do not make any such claims in our analysis of the Anxiety Support forum: our observations come principally from our investigation of this particular dataset as a corpus and we do not purport to have expertise regarding anxiety disorders. Nevertheless, the corpus approach does

offer an evidence-base for our observations, in that we report patterns of language use that are determined by their frequency in these data, which does come from individuals with personal experience and lay expertise.

With respect to anonymity, the corpus approach also minimises the potential for ‘recoverability’. Since our observations largely come from aggregated data, they are not tied to any particular individual’s words. While we stress the importance of analysing occurrences in context and, subsequently, do reproduce specific examples, we have endeavoured to redact any identifying information that was not already anonymised in the data transfer from HealthUnlocked. Even for those with access to HealthUnlocked (and the Anxiety Support forum, specifically), the search function for posts is not calibrated in a way that it will find exact matches for search phrases, rather it will recover messages that broadly relate to the search terms. This is an advantage for protecting the identities of members, in that any extended quotes from the data in this work are not directly recoverable through the site itself and so contributors can remain anonymous.

The scale of the data does present a challenge for the issue of informed consent. HealthUnlocked, for instance, is not in a position to delineate all or any potential research partners in their Terms of Service, so members are faced with an all-or-nothing approach to agreeing to be included in research. Without the specific research project to refer to, participants are also left unaware of what exactly is involved in such research. For corpus linguists, there are practical challenges in providing information to (often) thousands of participants and subsequently collecting informed consent. When it comes to online support forums, decisions regarding informed consent should also be balanced with concerns for the risk of harm (Elgesem, 2015). Posed with a similar challenge, Hunt and Brookes (2020, p. 75) reason that ‘imposing on the fora to request informed consent from a large number of contributors would risk disrupting the fora and undermining their primary role as supportive, recovery-oriented communities, thereby affecting the research participants and wider forum communities in a negative way’. With the measures in place that define membership to the forum, we believe that participants have sufficient control over information about them, which is managed via their own disclosures on the site and the anonymisation measures imposed by HealthUnlocked in making the data available.

Participants also have control over what personal information they provide about themselves through their user profile. When they register, members are given the opportunity to pick one of the pre-defined categories for gender (Female, Male, Other) and ethnicity (White/

Caucasian; Latino/Hispanic; Black/African/Caribbean; South Asian; East Asian; Middle Eastern; Mixed/Multiple ethnic groups; Other ethnic group), select the country where they live from a pre-defined list of options, provide their month and year of birth (from which their age is calculated), as well as provide a short bio. In each case, the member also has the option of leaving the category unselected/incomplete, and in 46.70% of cases, users did not indicate their gender (accounting for 31.54% of posts to the forum). The 26.80% of users who did not list a country of residence provided 17.96% of posts, 55.31% of users did not disclose their age (contributing 44.42% of posts) and 86.78% of users did not disclose their ethnicity (providing 83.64% of posts). Given how this information overlaps, it was only in the case of 25471 (8.66%) posts that the information about the poster was missing in all four categories. Based on the information that was available, we subsequently conducted analyses relating to age, gender and country of residence, but not ethnicity.

When members register, they are also assigned a unique, alphanumeric User ID. This allows us to take into account the contributions of individual user profiles and to evaluate the distribution of a linguistic feature of interest, that is, in terms of its use across a number of different participants. However, one thing we must acknowledge is the possibility that an individual can create multiple profiles and this is even mentioned in the content of the forum in a case where one member claims that another user is the person behind a number of profiles that they have been involved in unfriendly interactions with. We, like other members of the forum, do not have the information to determine either the uniqueness of the profile or the veracity of the information relating to age, gender, ethnicity and age. As such, our observations are offered with the caveat that this is information that the user has selected about themselves, rather than making any claims based on any objective reality. One further meta-data category that does not require user selection and input is the timestamp for the post, which enables us to investigate patterns in text features comparing year-to-year, or month-to-month for example and this is one of the dimensions we discuss in Chapter 7.

The actions taken to facilitate a corpus analysis of the data and to ensure that participants are treated ethically means that we cannot discuss a 'complete' experience of the Anxiety Support forum. There are graphical features of posts that have not been recorded in our corpus, some users opted out of inclusion in research, members potentially reported posts which may have subsequently been removed and there is also a direct message function in the forum whereby members can send private messages to each other, which was not included in the

data. Nevertheless, we have investigated approximately 21 million words of forum data from over 17,000 participants, with a view to identifying patterns in how members of the Anxiety Support forum describe their experiences of anxiety and how they cultivate a community of peers based on shared experiences. We now offer an overview of the remaining content of this book, demonstrating the different aspects of the forum we have explored in our work.

Overview of the Book

In the following chapter, our analysis begins by asking ‘What are the key ways that posters use language to conceptualise anxiety?’ We use the corpus analysis tool Sketch Engine to consider how the term *anxiety* compares to other conditions by examining how forum posters construct related concepts; for example, depression, fear, stress, panic and worry. We use the same software to provide a detailed Word Sketch of *anxiety* and its use in the forum, looking at its occurrence in different grammatical patterns. This analysis identified four clines in terms of how anxiety is discursively constructed.

In Chapter 3 we delve deeper into forum posters’ descriptions of anxiety by focusing on how they describe their lived experience of anxiety, what they understand causes their anxiety and how they best believe that their anxiety can be resolved. We also consider patient narratives around their anxiety. This chapter provides an overview of the linguistic content of the forum by carrying out a keyness analysis of the online corpus to identify words and phrases that are statistically salient in forum posts. These terms are placed into thematic categories and a representative set is analysed to illustrate different aspects of discussion around anxiety. The analysis in this chapter has a comparative aspect as we compare the language used by people when creating a new topic to post about, as opposed to language which occurs when people reply to an existing topic. New topics often contain introductions, narratives and descriptions of problems and feelings whereas the replies tend to be focused more on giving advice and support. In addition to considering keywords and phrases, the analysis looks at longer recurring phrases, as well as emoji and emoticons.

Chapter 4 continues the focus on the interactive and online affordances of the forum by looking at the ways that posters respond to each other’s posts. First, we ask, what kind of language characterises an initial post which receives numerous responses and how does this differ from posts which receive no responses at all. We then introduce a recently conceived framework for coding interactions in terms of

discourse units and apply it to a sample of our forum data, asking what language is typical of each discourse unit type. We then look at how different discourse units are typically combined within a single post and how discourse units work in sequences of responses. So, for example, if someone posts a message which has the function of writing about their feelings, what kinds of messages are they likely to get in response?

In Chapter 5 we make use of the demographically tagged nature of the forum posts by comparing and contrasting posts made by participants identifying as female and male, respectively. While 31.20% of posts were made by posters who did not specify their sex, 52.02% were made by posters who identified as female and 16.71% were made by male posters. These figures are congruent with data on prevalence of anxiety by sex, which tends to indicate that women are more likely than men to be diagnosed with anxiety, although we note that these figures probably do not give accurate cases due to the fact that many people do not report anxiety, for various reasons. We examine male and female keywords in the corpus, finding that men are more likely to use problem-solving language that focuses on explanations for anxiety and strategies for resolving it. On the other hand, women are more likely to use affiliative language to express empathy, sympathy and encouragement to others.

The chapter also examines gendered discourses relating to anxiety by considering representations around words such as *man*, *woman*, *macho* and *feminine*. Our analysis finds that men are viewed as having additional problems relating to anxiety due to societal pressures to ‘man up’ and not talk about mental health issues or emotions. Women’s experiences of anxiety are more often linked to relationship problems or experiences of abuse, as well as the burden of caring for families.

Chapter 6 continues the theme of demographic variation by comparing posts from the two countries in which the highest proportion of posters identified as residing. Within the forum 72.78% of posts were made by either people based in the UK or the USA. An analysis of keyword differences indicates numerous differences which point not only to spelling (*counseling*) and lexical choices (*brilliant*) but also to ways that anxiety is understood. For example, American posters are far more likely to refer to a range of anti-anxiety medications (*Lexapro*, *Klonopin*, *Xanax*), positioning anxiety resolution as a form of consumer choice that that can be managed via the purchase of the right medication. On the other hand, UK key terms are more likely to reference how their anxiety is managed by the National Health Service that advocates various therapies, although this can also result in delays to treatment.

Chapter 7 has time as its theme, taking advantage of the fact that our corpus contains posts made over a nine-year period, along with the demographic information we have about patient age. We consider diachronic aspects of the forum in three ways. First, we look at how language has changed over time, focusing on changes in anxiety discourses. For example, there appears to be increasing focus over time on describing the experience of anxiety in terms of how it feels and its effect on the body, along with more references to tests, medication and side effects. On the other hand, posters refer less to other people over time in their posts and use affiliative language less often. Such trends may be linked to the presence or absence of different types of posters at different points, and the analysis in this chapter links back to the findings in Chapters 5 and 6. We then consider how the age of posters impacts on the ways they write about language, dividing the corpus into age-graded sub-sections (e.g., 20s, 30s). We compare linguistic differences across these age groups, seeking to identify how language use reflects different understandings or representations of anxiety in relationship to age. Finally, we consider the ‘journey’ that posters go on, comparing posters at different stages of their relationship to the forum. By collecting people’s first, twentieth, fortieth, sixtieth and final posts we can trace how posting behaviour ‘evolves’ over time – with initial posts often taking the form of narratives and seeking help, while more seasoned posters tend to offer different forms of guidance, taking on the role of ‘expert’. The analysis of the final posts considers the extent to which people provide explanations or offer a sense of narrative ‘closure’ when they leave the group.

The concluding chapter of the book summarises the main findings from the preceding chapters, bringing them together to establish overall patterns and trends in online discourses of anxiety. These representations are then related to the contexts in which they are situated as well as their implications for our understanding of mental health in wider society. The chapter also critically reflects on the approach we took, the questions that emerged as a result of engaging with the corpus of forum posts and potential extensions to our study.

Although the subject of this book is anxiety, the analysis also acts as an exemplum of how the study of language can be applied to a very practical need, namely the analysis of a large amount of text from an online forum which focuses on a particular topic. Many of the methods that we outline in the book could be used to study language use around other health conditions and even for forum data which do not relate to health, we believe that the techniques we describe, such as concordancing, Word Sketches and keywords, could still be gainfully employed.