43

Corpus linguistics and health communication: using corpora to examine the representation of health and illness

Gavin Brookes, Sarah Atkins and Kevin Harvey

1 Introduction

Language is central to how we constitute our experiences of health and illness, as well as being of clear practical importance to health (care) professionals, who must communicate medical ideas on a regular basis. Experiences of health and illness are not simply based in the biological "realities" of our bodies but, crucially, in the language we use to talk about them:

> illness cannot be just illness, for the simple reason that human culture is constituted in language … and that health and illness, being things which fundamentally concern humans, and hence need to be "explained", enter into language and are constituted in language, regardless of whether or not they have some independent reality in nature.
>
> *(Fox 1993: 6)*

The field of health (care) communication addresses an impressive breadth of topics and genres, including (but not limited to) patient–practitioner interactions, communication between practitioners, illness narratives, government health education material, media representations of health and advertisements that project particular ideas about health and our bodies. Linguistic studies of health (care) language were initially (largely) preoccupied with the discourse of spoken interactions between patients and a range of health care practitioners, such as doctors (Candlin 2000), nurses (Crawford *et al.* 1998), physiotherapists (Parry 2004) and pharmacists (Pilnick 1999). However, by the late 1990s, greater attention was paid to written medical discourse in various communicative contexts, such as medical note-taking (Hobbs 2003) and case histories (Francis and Kramer-Dahl 2004), and then more recently on online texts such as online support groups (Hunt and Brookes 2020) and health care provider websites (Brookes and Harvey 2016).

As a result of the long-standing focus on spoken interaction, studies of health (care) communication have tended to rely on relatively small datasets, such as samples of

DOI: 10.4324/9780367076399-43

clinical exchanges, which were most amenable to fine-grained, qualitative analysis. As such, methodologically, much health (care) communication research has taken a sociolinguistic and discourse analytic perspective, utilising approaches like conversation analysis, text/genre analysis and (critical) discourse analysis. Many health (care) language studies have combined perspectives, for example, the tendency for conversation analysis and interactional sociolinguistic methodologies to be supplemented by a strain of critical discourse analysis, with the research impetus being as much to criticise and change practices in institutional health (care) settings as to describe and understand them (e.g. McHoul and Rapley 2001).

Although small-scale datasets – and the qualitative approaches used to analyse them – have produced rich insights into the linguistic dynamics of health (care), research employing such approaches has been criticised for basing findings on limited datasets that are not necessarily representative of wider communication within the particular domains under study. Thus, since the late 1990s health (care) communication scholars have increasingly harnessed the opportunities afforded by corpus linguistic approaches to obtain a more widely representative, and so generalisable, picture of how language is used in various clinical contexts.

This chapter demonstrates how some of the corpus linguistic techniques introduced so far in this book can be usefully applied to study health (care) communication. To do this, we provide two case studies, which each explore a different health-related topic, are based on corpora representing distinct textual genres and modes and utilise a different corpus technique as their point of entry. The first case study utilises keywords to examine characteristic themes in a corpus of adolescents' health advice–seeking emails. The second uses collocation to identify metaphors used to represent dementia in a corpus of newspaper articles. In each of these case studies we supplement each of the quantitative approaches with qualitative, theory-informed analysis of patterns of interest. Before addressing these case studies, we first provide a brief overview of existing applications of corpus linguistic methods to the study of health (care) communication.

2 Health (care) communication and corpus linguistics

Initial corpus studies of health (care) communication focused primarily on spoken interactions between patients and clinical staff. For example, Thomas and Wilson (1996) utilised a 1.25-million-word corpus of practitioner–patient exchanges and set out to demonstrate that computer content analysis can overcome the 'shortcomings of straight quantitative analysis' and has 'the potential to provide results which are in some respects comparable to manual discourse analysis' (1996: 92). Although corpus techniques enabled the researchers to quickly and accurately identify significant aspects of the health practitioners' language use, this study gave little emphasis to extended stretches of language in use and to how linguistic components actually functioned in the dialogic context of the practitioner–patient exchanges under investigation.

Later research by Skelton and colleagues (e.g. Skelton and Hobbs 1999) demonstrated the methodological advantages of integrating quantitative with qualitative approaches. As a starting point, they used frequency counts of words and phrases, complementing such quantitative insights with qualitative assessments of how such phrases operate in context through the analysis of concordance outputs. Rather than providing a broad linguistic characterisation of the particular communicative practices in question, the authors confine their analyses to investigating specific linguistic phenomena (e.g. metaphor and

imprecision). For this reason, their studies did not harness comparative data, contrasting, for example, their datasets with general reference corpora.

Such a comparative approach was adopted by Adolphs *et al.* (2004) to analyse NHS Direct exchanges between professionals and patient callers. Adolphs *et al*. compared a corpus of health professionals' language with a corpus of general spoken English, identifying keywords that appeared with greater frequency in the NHS Direct consultations. The researchers examined these key items in their original discourse environments using concordance lines and techniques from conversation analysis to provide close descriptions of interactional processes. These methodological stages afforded the authors a means of understanding the uniqueness of the professional–patient exchanges, enabling them to characterise the nature of NHS Direct consultations where they identified an overarching tendency for professionals to use politeness strategies and the language of convergence in their interactions with callers.

Following the early studies of spoken health (care) interactions, corpus methods then began to be applied to the study of written health (care)–related texts, for example, Crawford *et al.*'s (1998) study of nurses' written reports. An increasingly popular area of focus for corpus studies of written communication is the print media, with researchers examining press representations of (usually contemporary) health topics, for example, Brookes and Baker's (2021) corpus study of representations of obesity in the British press. More recently, corpus techniques have also been employed in conjunction with stylistic approaches to analyse literary representations of illness and health, such as Demjén's (2015) investigation of the language of affective states in Sylvia Plath's writing.

In recent years, corpus studies of health (care) communication have focused increasingly on language in digital contexts, with research exploring online platforms for advice seeking and giving being particularly prominent. For example, Harvey and colleagues examined variations of a corpus containing emails sent by young people to a health website aimed at adolescents (e.g. Harvey 2012). Another popular area of focus in corpus studies of online health (care) interaction is online, peer-to-peer support groups. A recent example includes Hunt and Brookes's (2020) study of the language around mental health in online support groups for people with anorexia, depression and diabulimia. An emerging, but less studied, area of focus for corpus studies of digital health (care) communication is patient feedback. For example, Baker *et al.* (2019) analysed the language of patient feedback in a 29-million-word corpus of online comments, along with an 11-million-word corpus of providers' responses. The recent and growing focus on digital discourse in corpus studies of health (care) communication partly reflects the growing influence of digital (communicative) technologies over the ways that people communicate – and indeed act – in relation to their health (Lupton 2017). Meanwhile, on a practical level, this penchant might also reflect the relative ease with which ready-digitised texts can be collected and prepared for corpus analysis (compared to written and spoken language).

The number of corpus studies of health (care) communication, while still relatively small, continues to grow. A notable feature of this work is its tendency towards methodological plurality – particularly the combination of quantitative corpus techniques with more qualitative approaches to linguistic analysis. Like other approaches to the study of health (care) language, the corpus approach also tends towards interdisciplinarity – incorporating perspectives from fields as diverse as medicine, psychology and sociology into linguistic analyses. This methodological and theoretical diversity also

mirrors the diversity of corpus linguistic approaches, whose flexibility provides fertile ground for such multiplicity of methods and theoretical approaches.

In the sections that follow, we provide two case studies which showcase the ways in which corpus linguistic methods can be used to interrogate health (care) language. Each case study utilises a different, established corpus technique as its starting point to investigate discourse produced in two different contexts with relevance to health. The first, which addresses the context of e-health, makes use of keywords to explore the thematic make-up of a corpus of emails submitted to an adolescent health website and gives a more in-depth analysis of the language used to disclose concerns relating to sexual health. The second, which focuses on the media, utilises collocation as a means of identifying the metaphors used by sections of the British press in their coverage of the topic of dementia. Although their respective approaches are based primarily on different corpus techniques (keywords and collocation), both of these case studies supplement their initial use of these techniques with more qualitative, context-sensitive analyses facilitated by concordance output, as well as recourse to the wider contexts of the original texts themselves.

3 Keywords: adolescent emails and sexual health

The first case study utilises the corpus technique of keywords to analyse online health communication. As a reminder, the keywords approach compares the frequency information for one corpus against another (i.e. the reference corpus) to indicate which words occur with a markedly high frequency in a target corpus, with such words constituting characteristic lexis in that data and so providing useful entry points for analysis (see Chapters 9 and 10, this volume). This case study focuses on the *Adolescent Health Email Corpus* (AHEC), a corpus of health advice-seeking emails sent to the *Teenage Health Freak* (www.teenagehealthfreak.org), a UK-based adolescent health advice website. These online requests for advice do not constitute emails in the traditional sense; they are not sent via the contributors' individual, personal email accounts, but communicated anonymously via a universal posting platform on the website. Our corpus analysis is based on the adolescent communiqués rather than the professionals' comparatively infrequent returns. Permission was given to collect and analyse emails sent to the *Teenage Health Freak* website between January 2004 and December 2005. The total dataset comprises 62,794 messages (approximately 1 million words).

In order to ascertain salient themes in the AHEC, *WordSmith Tools* (Scott 2016) was used to generate keywords by comparing this corpus against the original *British National Corpus* (BNC1994). Keywords were generated using the log-likelihood test (Dunning 1993) and then grouped thematically based on analysis of each keyword in its original contexts of use. The resultant keyword categories reflect the range of health themes raised in the adolescents' emails and includes (example keywords in brackets): mental health (e.g. *depression*, *depressed*, *suicide*), body weight and image (e.g. *anorexia*, *anorexic*, *weight*), drugs and alcohol (e.g. *drugs*, *cannabis*, *cocaine*), serious conditions (e.g. *cancer*, *epilepsy*, *diabetes*), minor conditions (*acne*, *zits*, *blackhead*) and medication (e.g. *medicine*, *medication*, *prescribed*). In this chapter, we focus on the theme of sexual health, which exhibited not only the largest number of keywords but also comprised some of the strongest keywords in the corpus. The keywords assigned to this theme are displayed in Table 43.1.

A full analysis of this theme would attend to most, if not all, of the keywords shown in this table. However, for the purpose of this case study, we focus on the topic of

Table 43.1 Sexual health keywords in AHEC, ranked by log-likelihood score

sex, sexual, penis, pregnant, period, orgasm, AIDS, infertile, STD, STI, sperm, contraception, HIV, clitoris, vagina, vulva, PMS, erection, condom, masturbate, gay, abortion, foreplay, intercourse, virgin, unprotected, lesbian, oral, pill, ovulation, herpes, thrush, chlamydia, pregnancy, tampon, testicles, genitalia, viagra, scrotum, labia, glans, ovaries, foreskin, balls, fanny, bisexual, miscarriage

sexually transmitted infections (STIs), specifically the keywords *AIDS* (n = 209) and *HIV* (114), and the adolescents' knowledge and representation of these concepts. Our choice of this topic for closer scrutiny is motivated not only by its prevalence in this corpus but also by its significance to the teenage population more generally.

Examining emails mentioning the keywords *HIV* and *AIDS* (314 in total) through the prism of concordance, we observed that these texts covered a range of themes, central among which are questions relating to HIV/AIDS terminology and conceptual definitions of the terms (of which there are 78 occurrences: 24 per cent), concerns regarding transmission and causation (72: 23 per cent) and questions about symptoms and the likelihood of having HIV/AIDS (60: 19 per cent). The lexical item *AIDS* appears nearly twice as often (209) as *HIV* (114). Of the 209 occurrences of *AIDS*, 17 co-occur with *HIV* (L5 to R5), which suggests a relationship between the two. These emails ask such questions as: "*how does **HIV/AIDS** get passed on*" and "*what happens when a man or women is **hiv** and has **aids***" (note: these and other extracts are presented with their original spelling, punctuation and capitalisation). In these instances, the adolescents clearly, and crucially, distinguish the concepts of HIV and AIDS, conceiving of them as separate entities, while also being connected in some way, given the oblique (/) or conjuncts ("and", "or") that coordinate them.

However, most of the occurrences of *AIDS* in the corpus (192: 92 per cent) appear without mention of HIV, the virus that can cause AIDS (UNESCO 2006). The following examples are typical of the range of contexts in which AIDS in isolation is used:

1. How do i know if iv got **aids**
2. Can you get **AIDS** by being fingered?
3. I had sex without using a condom and i am really scared i might be pregnant or might have **aids**
4. I have had sex with my boyfriend for the first time and the condom kept snapping so we decided to not use one what are the chances of me bein pregnant or having **aids**? please answer this im really really wottied

The absence of any reference to HIV and the foregrounding of AIDS indicate a terminological conflation of the two concepts, a misconception that is liable to have profound consequences in terms of how the adolescents conceive of and understand HIV and AIDS. For instance, in a number of the examples, there is the underlying belief that AIDS is a communicable infection, not a syndrome or range of conditions (UNESCO 2006), with its being constructed as, and confused with, a virus or disease, something that can be readily transmitted via sexual activity.

Collapsing the distinction between HIV and AIDS in this way inevitably results in confusion and reinforces 'unrealistic and unfounded fears' (Watney 1989: 184) on the part of the adolescents who might mistakenly believe themselves to be at risk of AIDS but not HIV. Such extreme worst-case scenarios conceive of AIDS as something that

sets in immediately after infection, a unitary phenomenon rather than a collection of different medical conditions – beliefs which obscure, if not efface altogether, the existence of HIV, the virus, which is indeed infectious. Such erroneous conflation of HIV infection with AIDS (by definition, the stage of HIV infection 'when a person's immune system can no longer cope'; Terrence Higgins Trust 2007: 1) repeats some of the early and fundamental misconceptions and negative attitudes about AIDS that were widespread during the 1980s and 1990s (Sikand *et al.* 1996; Helman 2007). For example, Warwick *et al.*'s (1988) in-depth study into youth beliefs about AIDS revealed that a significant number of young people, as with many adults, were unable to distinguish between HIV infection and AIDS, a finding which they attributed to the media's consistent failure to provide the public with accurate information. This fundamental misunderstanding (identifying AIDS as a transmissible disease) was related to the 'public terror about "catching" AIDS from people in public places or during casual contact' (Grover 1990: 145). Such beliefs (and the emotive linguistic choices encoding them) prevalent during that period are still apparent in the adolescent health emails communicated over 20 years later. For instance, one of the central ways in which adolescents describe becoming infected with HIV or developing AIDS is through use of the lemma CATCH, the second most common verb (13 times) used to signify Contraction after GET (n = 52). The use of CATCH as a verb encoding transmission of HIV/AIDS is telling, implying a more active role for subjects:

5. how do you prevent **catching** h. i. v
6. is the aids virus difficult to **catch**?
7. if i have sex with someone with aids without protection can i **cath** it
8. can you **catch** aids if someone masterbates you.

Biber *et al.* (1999) describe 'catch' as an 'activity verb', a verb denoting actions and events 'that could be associated with choice' (1999: 361). As these emails illustrate, 'catch' implies notions of agency on the part of subjects in the sense that it is within their power to prevent infection, with responsibility framed in terms of both general or universal agency, encoded via the second person: "Can *you* catch …?", or individual control via the first-person singular pronoun: "Can *I* catch …?". As Johnson and Murray (1985: 152) put it, "catching" an ailment (as in catching a cold) semantically implies a degree of co-operation: 'We catch things … in ways which are our own fault; we blame ourselves – we should have worn galoshes, and should not have sat in a draught'. This notion of personal agency and responsibility, communicated through the verbal concept of "catching", is also apparent in questions referring explicitly to prevention and avoidance strategies (e.g. "how do you prevent catching h. i. v?", "is the aids virus difficult to catch?"). Similarly, it is evident in messages in which the adolescents seek clarification as to whether specific activities are liable to result in contracting HIV/AIDS and which should therefore be avoided (e.g. "can you catch aids if someone masterbates you?", "if I have sex with someone with aids without protection can i cath it?"). Here, then, both HIV and AIDS are constructed as preventable through individual agency and requisite care. Infection with HIV is not an inevitable outcome, as, alarmingly, some young people have perceived it to be (Warwick *et al.* 1988).

However, commentators and public health bodies stress that neither HIV nor AIDS can be "caught" (Watney 1989: 184). Contemporary health promotion literature produced by organisations such as UNESCO continually warns against the use of this verb

to signify the way that people might become HIV positive, since it only helps to reproduce myths about HIV and AIDS (UNESCO 2006). In the health emails noted earlier, for example, the various realisations of the lemma CATCH unavoidably and infelicitously conjure notions of the common cold and influenza, as evidence from the general language *British National Corpus* (BNC) attests. Consulting the 100-million-word BNC1994 reveals that, as a transitive verb, "catch" co-occurs with the direct objects "cold" (113), "chill" (21), "bug" (20) and "colds" (7). As these collocates indicate, one typical use of the verb "catch" in general English is to describe the acquisition of relatively minor infections, in the sense of their being widespread and generally innocuous (though "bug", of course, potentially relates to more serious infections such as methicillin-resistant *Staphylococcus aureus* [MRSA], the so-called "super bug"; Knifton 2005). With regard to more serious viruses and illnesses, other less euphemistic constructions are used in the BNC1994 to describe the process of becoming infected and the onset of morbidity: for example, HIV is typically "contracted", "got", "acquired", while AIDS is "got", "developed", "contracted". Given this association, a corollary of using "catch" to describe infection with HIV/ AIDS is to encode the assumption that the virus can be acquired via casual contact, possessing a transmission efficacy similar to both colds and influenza. As such, talk of "catching" HIV/AIDS figuratively transforms the virus from something which is, in reality, difficult to transmit and is only communicable via specific routes (Terrence Higgins Trust 2007: 2) to something highly contagious, and liable to spread rapidly and extensively.

From this short case study, it is evident that the adolescents contributing to the AHEC possessed a range of misconceptions about HIV and AIDS, among which perhaps the most alarming was the tendency for some to conflate the two, reinforcing the idea that HIV and AIDS are identical. Such a conflation is liable to obscure awareness of the ways the virus is transmitted, potentially impeding assessments of risk in relation to sexual behaviour. The corpus analysis further highlighted some of the folk conceptualisations of sexual health that the adolescents operated with, beliefs that may need to be addressed by educators. Responding to lay beliefs like these is crucial, since people are liable to filter official health education messages through popular beliefs about health (Helman 2007), reinterpreting them to suit their own needs (Aggleton and Homans 1987: 25). However, not all of the adolescents' emails, in the shape of distorting metaphorical transformations and alarmist folk beliefs, display irrational responses to HIV and AIDS. The most commonly occurring emails about HIV and AIDS in the adolescent health corpus are fundamental questions concerning definitions and terminology, specifically: "What is HIV/AIDS?" and "What does HIV/AIDS stand for?" One way of interpreting these open and elementary types of enquiry is, of course, to regard them as emblematic of knowledge deficits about sexual health. Yet equally, such questions might be considered vital responses to a contemporary and potentially life-threatening condition that is still commonly misunderstood (Helman 2007).

4 Collocation: dementia metaphors in the press[1]

Our second case study uses collocation analysis to explore the metaphors used to represent dementia in articles published by the British tabloid, the *Daily Mail*. *Dementia* is an umbrella term used to refer to a range of diseases which cause a series of cognitive impairment symptoms, such as problems with memory, reasoning, perception and communication (World Health Organization [WHO] 2017). The most prevalent type of

dementia is Alzheimer's disease, but other types include vascular dementia, dementia with Lewy bodies and frontotemporal dementia. Dementia is a progressive syndrome, meaning that its symptoms worsen over time. It is widely claimed that dementia represents a major global public health challenge, as it is currently understood to affect as many as 47.5 million people worldwide, with an estimated 7.7 million new cases each year (WHO 2017). The number of people living with dementia is projected to rise to 75.6 million by 2030, trebling to 135.5 million by 2050.

At present, there is no pharmacological cure for dementia, meaning that people living with the syndrome today will also die with it. However, it is important to note that while people living with dementia will also die with it, they will not die *of* it, for dementia itself does not cause death. Indeed, although dementia is associated with reduced life expectancy, when people with dementia die, their death is more likely to be caused by co-occurring conditions such as pneumonia, rather than by dementia directly.

With no cure in immediate sight, there is a need for research focusing on people's lived experiences and understandings of dementia in the here and now. This includes how experiences and perceptions of the syndrome might be shaped by cultural constructions of it, for instance in the media, which constitutes a primary source through which people routinely access information about, and form their impressions of, health concerns (Seale 2003). As mentioned, this case study focuses on the metaphors used to represent dementia in the *Daily Mail*. Put simply, *metaphor* refers to 'the phenomenon whereby we talk and, potentially, think about something in terms of something else', where the two things involved are different but are perceived by the speaker to share some similarity (Semino 2008: 1). Because they are often used to express abstract, subjective, sensitive and taboo subjects, it stands to reason that talk about health and illness is often rife with metaphorical language (ibid: 175–6).

For this case study, we draw upon a purpose-built, specialised corpus containing articles from the *Daily Mail* mentioning *dementia* either once in their headline or three or more times throughout the rest of the article, published between 2007 and 2017. The resulting corpus contains a total of 1,674 articles and 1,155,675 words. We decided to focus on just this newspaper, as it published more articles about dementia (i.e. mentioning *dementia* once in the headline or three or more times in the article body) during this period than any other UK national newspaper.

To identify the metaphors used to represent dementia, we utilised collocation analysis. Collocation is a linguistic device whereby words, in associating strongly with one another, become bearers of meaning by virtue of co-occurrence (see Chapters 9, 14 and 15, this volume). Collocation is typically judged to exist using a word-association measure that tells us how often two or more words occur alongside one another and whether this association is notable as a sizeable effect in our data (i.e. the words have a measurably strong preference to occur together as opposed to being randomly associated). For this analysis, we used *WordSmith Tools* to generate a list of collocates of the word *dementia* (n = 12,095) (L5 to R5, frequency ≥5). We then ranked our list of collocates using the Mutual Information (MI) statistic and removed collocates which were not assigned an MI score of at least 3 (Hunston 2002). The resulting collocates were then analysed in context to ascertain whether or not they tended to be used metaphorically. This analysis indicated the use of numerous source domains (Lakoff and Johnson 1980). However, for this case study we will focus on the source domain which exhibited not only the most but also the most frequent collocates, namely VIOLENCE. Table 43.2 shows the collocates assigned to this domain.

Table 43.2 VIOLENCE metaphorical collocates of *dementia* (MI ≥ 3), ranked by collocation frequency (in brackets)

against (121), *battle* (81), *fight* (65), *ward* (62), *victims* (48), *stricken* (40), *battling* (36), *hit* (35), *beat* (30), *killed* (20), *losing* (20), *victim* (20), *cruel* (17), *killer* (16), *struck* (15), *struggle* (15), *killing* (14), *struggling* (13), *tackle* (13), *wards* (13), *kill* (12), *robbed* (11), *tackling* (11), *war* (9), *beating* (8), *toll* (8), *hits* (7), *succumbed* (7), *weapon* (7), *enemies* (6), *combat* (5), *cosh* (5), *fighting* (5), *onslaught* (5), *striking* (5), *struggled* (5)

Although the metaphorical collocates in Table 43.2 all relate to dementia, more specifically these violence tropes target dementia itself, the person living with dementia (and their experience of the syndrome) and responses to the condition, including by individuals but also larger institutions like the government and medico-scientific researchers. Beginning with the representation of dementia itself, several of the collocates in the table contribute to the anthropomorphisation of the syndrome. This agent is described as *cruel* and is framed as inflicting acts of violence on people, foremost the act of murder. For example, dementia is construed as a killer (*killed*, *killer*, *killing*, *kill*) (collocates underlined):

9. Cancer accounts for almost a third of cases while heart disease is the biggest killer of men and **dementia** in women. Dementia now claims the lives of more than 51,000 women and men a year. **Dementia** and Alzheimer's kill three times more women than breast cancer. It's now the second biggest killer of men, with most dying of heart disease.Giana Hennigan,

(2018)

Similarly, it is also described as inflicting a *toll* and *onslaught* on those affected by it:

10. But now scientists claim the hot drink is more than just an enjoyable treat, it can actually help to prevent the onslaught of **dementia**.

(2016)

It is worth remembering at this point, as noted earlier, that although people with dementia have a reduced life expectancy and will die *with* the condition, they do not die *of* it. Thus, we can question the logic and accuracy of these fatalistic tropes. Yet in addition to murder, dementia commits other acts of violence, such as *hit*[ting] and *striking*. Another form of violence that dementia is presented as inflicting upon those diagnosed with it is theft. Specifically, the lexical verb collocate *robbed* is used to characterise dementia as stealing individuals' ability to perform certain actions such as walking and speaking:

11. Poor old Terry Jones, as he calls him, has since been revealed to be suffering from a form of **dementia** that has robbed him of the power of speech. There will definitely be no more reunions now. Palin's eyes fill with tears as we talk about him.

(2017)

Consistent with the types of metaphorical framings seen so far, the experience of dementia could equally be framed as a violent conflict, with life with dementia characterised as an ongoing battle (*battle*, *battling*), struggle (*struggle*, *struggling*, *struggled*) or fight *(fighting*):

12. Her heartbreaking story of a woman struggling with **dementia** was inspired by her grandmother's battle with the cruel disease.

(2016)

Other metaphors used to represent the lived experience of dementia did so in ways that were similarly violent but implied a more limited prospect for agency on the part of the person living with the syndrome. This could manifest in words implying a fatalistic outcome of life with dementia, such as characterising people with dementia as victims (*victims*, *victim*), *stricken* and as having lost (*losing*) or succumbed to the syndrome.

13. Frail WWII veteran, 89, faces five years in jail after admitting keeping loaded revolver by his bed 'so he could kill himself if he succumbed to **dementia**'

(2015)

This reduced agency was also evident in lexical verb collocates denoting violent processes (*hit*, *struck*) of which people with dementia are the passive objects:

14. TINY MEMORY LAPSES? **DEMENTIA** COULD HIT YOU IN JUST 12 YEARS

(2014)

The final aspect of the dementia experience that was framed metaphorically in violent terms was responses to the syndrome. This includes advising readers on measures they as individuals can take to reduce their risk of developing the syndrome. These actions could be lexicalised as helping to "ward off" dementia, with the implication being that the condition is an unwanted invader that individuals need to defend themselves against. More commonly, violence metaphors were used to characterise the actions and responses of larger institutions, such as the government and medico-scientific researchers. In these scenarios, dementia is cast as an enemy (*enemies*), with whom these actors are engaged in a violent conflict (*against*, *fight*, *beat*, *tackle*, *tackling*, *war*, *beating*). As the use of the term *weapon* (also a collocate of *dementia*) in this example attests, in this scenario the (prospective) pharmacological solutions to obesity were also lexicalised as violent weapons (*weapon*, *cosh*):

15. NEW WEAPON IN THE WAR ON **DEMENTIA**: THE TWO RONNIES!

(2016)

Violence metaphors are often considered problematic when used in relation to health and illness. For example, Sontag (1978) famously cautioned against the use of war [sic] metaphors to talk about illness, arguing that they can be stigmatising for those affected by the illness in question, who could be rendered as the enemy in such scenarios, and their

bodies the metaphorical "battlefields". Moreover, it is argued that if the disease prevails and the person fails to recover from it, they could be viewed as having "lost" the battle.

While these criticisms can likewise be directed at the use of violence metaphors in relation to dementia, such arguments have been countered by more recent claims that people can find such violence metaphors empowering, as they grant them a greater sense of agency in their illness experience, helping them to feel like they have greater influence over their prognosis. However, given that there is not currently a cure for or prospect of recovery from dementia, this is a battle that is inevitably lost in this case. Indeed, the focus on pharmacological interventions in metaphors describing responses to dementia not only places emphasis on prospective treatments that are not currently available but also serves to simultaneously shift focus *away from* other, non-pharmacological measures which help people with dementia to live with and manage the syndrome in the here-and-now. Moreover, such emphasis could also influence public thinking and policy decisions relating to appropriate ways of responding to dementia in the future – a relevant argument when we consider the current disparity between funding for medico-scientific research and social care.

In conclusion, then, we can question the extent to which such tropes are likely to produce contented attitudes for people with dementia and their relatives and carers and actually raise genuine awareness of the syndrome. The effect of recycling such fear-inducing tropes, which seem just as much calculated to sensationalise as they do to inform, cannot be underestimated, for as Chivers (2011: 60) observes, the public are liable to associate 'dementia with the most horrifying possible loss of self', an assumption that is deeply embedded and continues to persist in western culture. Of course, readers' responses to dementia are not always uniform, and readers can resist dominant media narratives about topics like dementia. Most analyses of corpora of media texts, such as the one presented here, do not (and cannot) provide empirical insight into the ways in which audiences respond to media representations of health topics. Recent corpus studies of media texts have begun to address this limitation by engaging with readers' comments on online articles (so-called "below-the-line" comments), in which readers offer their own views and perspectives on stories and in the process offer their own representations which can be consistent with but also challenge those offered in the corresponding articles. An example relevant to media representations of health is Brookes and Baker's (2021) corpus study of British press representations of obesity, in which they compare discourses in a sample of reader comments against the articles on which they are based. Yet it remains a fact that many people have a poor understanding of and continue to fear dementia more than they do other serious diseases, such as cancer, a view shared not only by older adults but also young people (Alzheimer's Research Trust 2011).

Although we have focused on the *Daily Mail* in this study, the metaphorical patterns we have identified are certainly not particular to this newspaper but reflect wider trends in UK press discourse. In defence of the press, we could argue that metaphor provides a useful set of linguistic and conceptual apparatus with which to communicate about complex conditions like dementia, conferring a sense of symbolic order over them for the purposes of communicating about them with the general public. However, for this purpose we would prefer to see the press utilising other metaphors that are not only less stigmatising for people living with dementia but which also help to raise more accurate and useful public awareness of it. Possible alternatives include metaphors of JOURNEY and COMPANIONSHIP, which, respectively, articulate the progressive and chronic nature of the syndrome. Although there is some evidence of such tropes in our *Daily Mail* corpus, they are comparatively infrequent relative to the VIOLENCE metaphors explored here. It

is perhaps the case, then, that the pressure on (particularly tabloid) newspapers to print negative and sensationalistic headlines in order to sell more copies (Conboy 2006), or to generate more clicks, outweighs their sense of duty to combat stigma and raise accurate and genuinely useful awareness of health concerns like dementia.

5 Conclusion

The case studies reported in this chapter demonstrate the capacity of corpus linguistic methods to provide both quantitative *and* qualitative perspectives on data that represents a wider range of communicative styles, routines, lived experiences and understandings of health (care) and illness than would be feasible using purely qualitative approaches, producing – by and large – more generalisable findings in the process. Although this chapter has focused on language in two specific health (care) contexts, the analytical techniques afforded by corpus linguistics are flexible and can, in theory, be applied to the analysis of language produced in relation to any health issue or within any health (care) context. Whatever the topic under study, by affording the opportunity to examine large quantities of authentic language data, corpus linguistic methods could be said to go some way towards appeasing the commitment to more objective approaches to large datasets that is commonplace in the domain of empirical health research (Brown *et al.* 2006), potentially helping researchers to bridge the gap between social scientific and biomedical perspectives on health and illness.

In terms of future applications, with increasing emphasis on evidence-based practice in health care training (Brown *et al.* 2006), corpus linguistic methods could provide practitioners and instructors with substantial evidence bases on which to both teach and learn about the communicative routines of particular (clinical) groups. Indeed, given the significant contribution that corpus linguistics has already made to practices in language teaching, including increasingly in the teaching of languages for health care purposes, it is not difficult to envisage a role for corpora – as vast repositories of authentic health (care) language – in more general clinical communication training. The main challenge, as we see it, resides in convincing health practitioners and policymakers, who are familiar with and committed to their own already established research methods and standards of data representativeness, of the virtues and values of corpus linguistic approaches in the long-term.

Note

1 This research is supported by a UKRI Future Leaders Fellowship, grant reference MR/V022954/1.

Further reading

Gwyn, R. (2002) *Communicating Health and Illness*, London: Sage. (This is an accessible discourse-based survey to health communication.)

Harvey, K. and Koteyko, N. (2012) *Exploring Health Communication*, London: Routledge. (This text provides an accessible overview of key concepts in language and health but also has more focus on corpus linguistic applications in this domain.)

Jones, R. (2013) *Health and Risk Communication: An Applied Linguistic Perspective*, London: Routledge. (This book introduces key concepts in linguistic studies of health discourse.)

References

Adolphs, S., Brown, B., Carter, R., Crawford, P. and Sahota, O. (2004) 'Applied Clinical Linguistics: Corpus Linguistics in Health Care Settings', *Journal of Applied Linguistics* 1: 9–28.

Aggleton, P. and Homans, H. (1987) 'Teaching about AIDS', *Social Science Teacher* 17: 24–8.

Alzheimer's Research Trust (2011) *Alzheimer's Research UK Launches as Public Dementia Fears Spiral*, Retrieved from: http://www.alzheimersresearchuk.org/AlzheimersResearchUKlaunch/.

Baker, P., Brookes, G. and Evans, C. (2019) *The Language of Patient Feedback: A Corpus Linguistic Study of Online Health Communication*, London: Routledge.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*, London: Longman.

Brookes, G. and Baker, P. (2021) *Obesity in the News: Language and Representation in the Press*, Cambridge: Cambridge University Press.

Brookes, G. and Harvey, K. (2016) 'Opening up the NHS to Market: Using Multimodal Critical Discourse Analysis to Examine the Ongoing Corporatisation of Health Care Communication', *Journal of Language and Politics* 15: 288–302.

Brown, B., Crawford, P. and Carter, R. (2006) *Evidence-Based Health Communication*, Maidenhead: Open University Press.

Candlin, S. (2000) 'New Dynamics in the Nurse–Patient Relationship?', in S. Sarangi and M. Coulthard (eds) *Discourse and Social Life*, London: Longman, pp. 230–45.

Chivers, S. (2011) *The Silvering Screen: Old Age and Disability in Cinema*, Toronto: University of Toronto Press.

Conboy, M. (2006) *Tabloid Britain*, London: Routledge.

Crawford, P., Brown, B. and Nolan, P. (1998) *Communicating Care: The Language of Nursing*, Gloucester: Stanley Thornes.

Demjén, Z. (2015) *Sylvia Plath and the Language of Affective States Written Discourse and the Experience of Depression*, London: Bloomsbury.

Dunning, T. (1993). 'Accurate Methods for the Statistics of Surprise and Coincidence'. *Computational Linguistics* 19(1): 61–74.

Fox, N. (1993) *Postmodernism, Sociology and Health*, Buckingham: Open University Press.

Francis, G. and Kramer-Dahl, A. (2004) 'Grammar in the Construction of Medical Case Histories', in C. Coffin, A. Hewings and K. O'Halloran (eds) *Applying English Grammar: Functional and Corpus Approaches*, London: Arnold, pp. 172–90.

Grover, J. (1990) 'AIDS: Keywords', in C. Ricks and L. Michaels (eds) *The State of the Language*, 1990s edition, London: Faber & Faber, pp. 142–62.

Harvey, K. (2012) 'Disclosures of Depression: Using Corpus Linguistics Methods to Interrogate Young People's Online Health Concerns', *International Journal of Corpus Linguistics* 17: 349–79.

Helman, C. (2007) *Culture, Health and Illness*, 5th edn, London: Hodder Arnold.

Hobbs, P. (2003) 'The Use of Evidentiality in Physician's Progress Notes', *Discourse Studies* 5: 451–78.

Hunston, S. (2002) *Corpora in Applied Linguistics*, Cambridge: Cambridge University Press.

Hunt, D. and Brookes, G. (2020) *Corpus, Discourse and Mental Health*, London: Bloomsbury.

Johnson, D. and Murray, J. (1985) 'Do Doctors Mean What They Say?', in D. Enright (ed.) *Fair of Speech: The Uses of Euphemism*, Oxford: Oxford University Press, pp. 151–58.

Knifton, C. (2005) 'Social Work and the Rise of the MRSA "Super Bug"', *Practice* 17: 39–42.

Lakoff, G. and Johnson, M. (1980) *Metaphors We Live By*, Chicago: University of Chicago Press.

Lupton, D. (2017) *Digital Health: Critical and Cross-Disciplinary Perspectives*, London: Routledge.

McHoul, A. and Rapley, M. (2001) 'Preface: With a Little Help from Our Friends', in A. McHoul and M. Rapley (eds) *How to Analyse Talk in Institutional Settings*, London: Continuum, pp. 25–38.

Parry, R. (2004) 'Communication During Goal-Setting in Physiotherapy Treatment Settings', *Clinical Rehabilitation* 18: 668–82.

Pilnick, A. (1999) '"Patient Counselling" by Pharmacists: Advice, Information or Instruction?', *Sociological Quarterly* 40: 613–22.

Scott, M. (2016) *WordSmith Tools, version 7*, Stroud: Lexical Analysis Software.

Seale, C. (2003) *Media and Health*, London: Sage.
Semino, E. (2008) *Metaphor and Discourse*, Cambridge: Cambridge University Press.
Sikand, A., Fisher, M. and Friedman, S. (1996) 'AIDS Knowledge, Concerns, and Behavioural Changes among Inner-City High School Students', *Journal of Adolescent Health* 18: 325–28.
Skelton, J. and Hobbs, F. (1999) 'Concordancing: Use of Language-Based Research in Medical Communication', *Lancet* 353: 108–11.
Sontag, S. (1978) *Illness as Metaphor*, New York: Farrar, Straus and Giroux.
Terrence Higgins Trust (2007) *Understanding HIV Infection: HIV? AIDS?*, 5th edn, London: Terrence Higgins Trust.
Thomas, J. and Wilson, A. (1996) 'Methodologies for Studying a Corpus of Doctor-Patient Interaction', in J. Thomas and M. Short (eds) *Using Corpora for Language Research*, London: Longman, pp. 92–109.
UNESCO (United Nations Educational, Scientific and Cultural Organisation) (2006) UNESCO Guidelines on Language and Content in HIV- and AIDS-Related Materials, Paris: UNESCO.
Warwick, I., Aggleton, P. and Homans, H. (1988) 'Young People's Health Beliefs and AIDS', in P. Aggleton and H. Homans (eds) *Social Aspects of AIDS*, Sussex: Falmer Press, pp. 106–25.
Watney, S. (1989) 'AIDS, Language and the Third World', in E. Carter and S. Watney (eds.) *Taking Liberties: AIDS and Cultural Politics*, London: Serpent's Tail, pp. 183–92.
World Health Organization (2017) *Dementia: Fact Sheet*, Retrieved from http://www.who.int/mediacentre/factsheets/fs362/en/.